Massively Parallel Sequencing of Gene Fusion-Associated Sarcomas

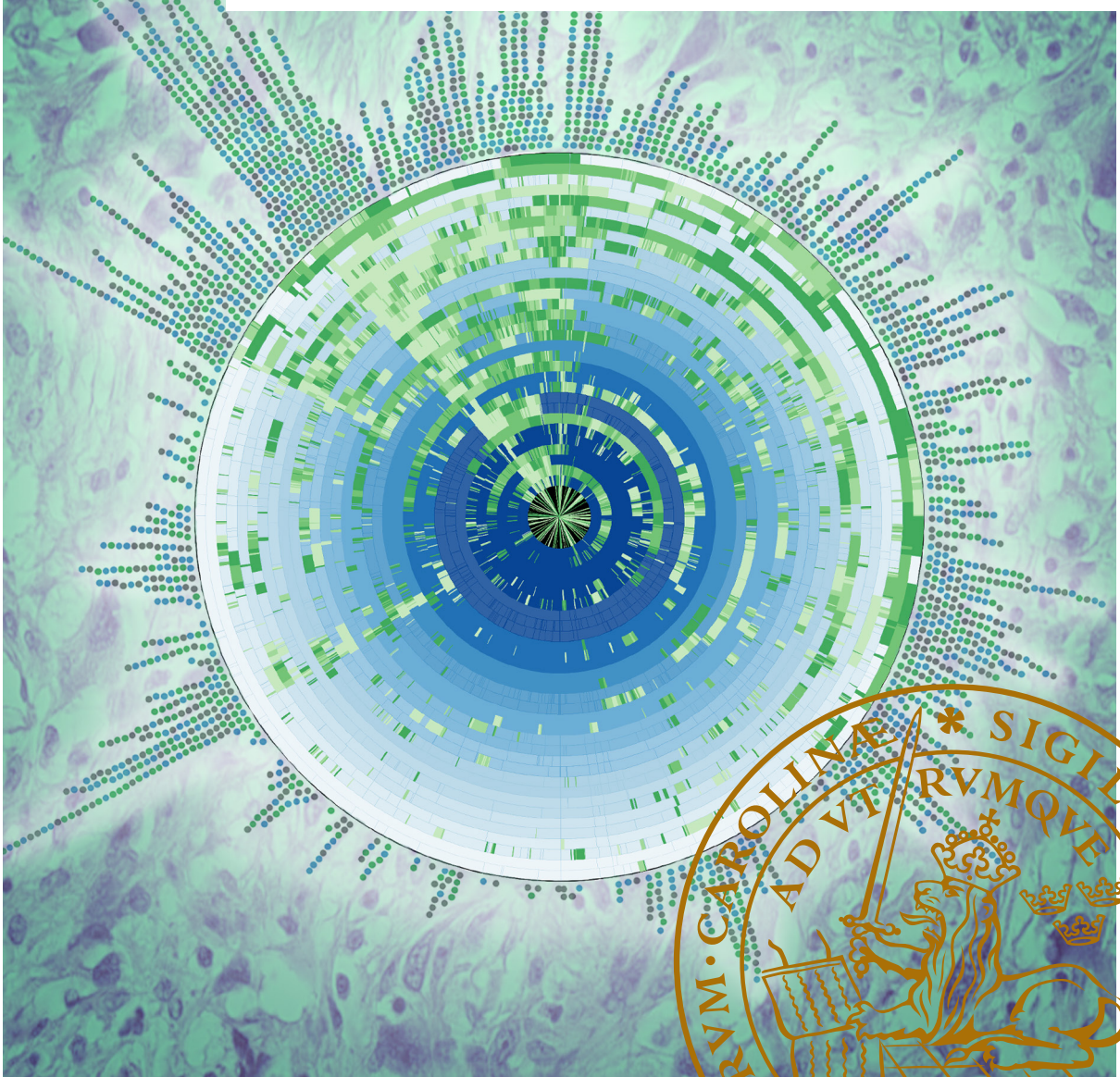
Hofvander, Jakob

2019

Document Version:
Publisher's PDF, also known as Version of record

Massively Parallel Sequencing of Gene Fusion-Associated Sarcomas

JAKOB HOFVANDER
FACULTY OF MEDICINE | LUND UNIVERSITY



Massively Parallel Sequencing of Gene Fusion-Associated Sarcomas

Jakob Hofvander

LUND
UNIVERSITY

DOCTORAL DISSERTATION
by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Belfragesalen, BMC, Lund.
January 18, 2019, at 13:00.

Faculty opponent
Professor Ola Myklebost
University of Bergen, Norway

Organization LUND UNIVERSITY Division of clinical genetics	Document name Doctoral Dissertation
	Date of issue January 18, 2019
Author: Jakob Hofvander	Sponsoring organization: The Swedish Childhood Cancer Foundation, The Swedish Cancer Society, Region Skåne and the Royal Physiographic Society of Lund

Title and subtitle: Massively Parallel Sequencing of Gene Fusion-Associated Sarcomas

Abstract

This thesis concerns genomic and transcriptomic analysis of sarcomas i.e., malignant tumors arising in tissues of mesenchymal origin. There are more than 50 sarcoma subtypes and differentiating between them can be difficult due to their rarity and morphologic overlap. Additionally, the genetic mechanisms underlying sarcoma development are only partly characterized, making accurate diagnosis challenging and treatment options limited. The aim of this thesis was to, with the use of massively parallel sequencing, further investigate and characterize the genetic changes underlying the development of primarily gene fusion-associated sarcomas.

In article I, we wanted to evaluate to what extent seemingly unique structural aberrations result in a functional fusion transcript. By using a combination of RNA-sequencing (RNA-seq) and cytogenetic data, we found that there is an increased likelihood of finding novel gene fusions in sarcomas displaying simple structural rearrangements. Additionally, we found it advantageous to run multiple gene fusion-detecting algorithms to obtain accurate results when analysing RNA-seq data from sarcomas.

In article II, we searched for gene fusions in undifferentiated pleomorphic sarcoma (UPS) and identified two novel fusions involving the transcription factor PRDM10, either as a *MED12-PRDM10* or a *CITED2-PRDM10* fusion. In article III, we identified a larger series of PRDM10-positive tumors (PPT) and characterized their genomic and transcriptomic features. PPT appeared to be genetically distinct from high-grade UPS as they displayed a unique gene expression profile and few genomic alterations. Additionally, they seem less aggressive than classical high-grade UPS and might therefor be clinically important to identify. We also investigated the effects of the *CITED2-PRDM10* fusion in cell lines and identified promising diagnostic markers for immunohistochemistry.

In article IV, we wanted to compare clonal evolution in tumors arising through different mechanisms by investigating the dynamics of copy number changes and nucleotide level mutations in three types of sarcoma; amplicon-driven well-differentiated liposarcoma, gene fusion-driven myxoid liposarcoma, and sarcomas with complex genomes (CXS). We found that the type and rate of clonal evolution differed considerably among sarcomas caused by different pathogenetic mechanisms. While both types of liposarcoma displayed a remarkable paucity of clonal evolution at the DNA level, suggesting that they obtain a genetic fitness maximum early in tumor development, the development of new mutations in many CXS fitted well with data on carcinomas.

Key words: Sarcomas, gene fusions, tumor evolution, clonal heterogeneity, RNA-seq, WES, SNP array

Classification system and/or index terms (if any)

Supplementary bibliographical information		Language: English
ISSN and key title: 1652-8220		ISBN: 978-91-7619-736-3
Recipient's notes	Number of pages 49	Price
	Security classification	

Massively Parallel Sequencing of Gene Fusion-Associated Sarcomas

Jakob Hofvander

LUND
UNIVERSITY

Cover photo by Jakob Hofvander

Lund University
Faculty of Medicine Doctoral Dissertation Series 2019:7

Department of Laboratory Medicine
Division of Clinical Genetics

ISBN 978-91-7619-736-3
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2018

Media-Tryck is an environmentally certified and ISO 14001:2015 certified provider of printed material.
Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Table of Contents

Original Articles

This thesis is based on the following articles:

I. Hofvander J, Tayebwa J, Nilsson J, Magnusson L, Brosjö O, Larsson O, Vult von Steyern F, Domanski HA, Mandahl N, Mertens F. **RNA sequencing of sarcomas with simple karyotypes: identification and enrichment of fusion transcripts**. *Lab Invest*. 2015;95:603-609.

II. Hofvander J, Tayebwa J, Nilsson J, Magnusson L, Brosjö O, Larsson O, Vult von Steyern F, Mandahl N, Fletcher CDM, Mertens F. **Recurrent *PRDM10* gene fusions in undifferentiated pleomorphic sarcoma.** *Clin Cancer Res*. 2015;21:864-869.

III. Hofvander J, Puls F, Pillay N, Steele CD, Flanagan A, Magnusson L, Nilsson J, Mertens F. **Undifferentiated pleomorphic sarcomas with *PRDM10* fusions have a distinct gene expression profile**. *Manuscript*.

IV. Hofvander J, Viklund B, Isaksson A, Brosjö O, Vult von Steyern F, Rissler P, Mandahl N, Mertens F. **Different patterns of clonal evolution among different sarcoma subtypes followed for up to 25 years.** *Nat Commun.* 2018;9:3662.

Articles not included in the thesis:

Walther C, Hofvander J, Nilsson J, Magnusson L, Domanski HA, Gisselsson D, Tayebwa J, Doyle LA, Fletcher CDM, Mertens F. **Gene fusion detection in formalin-fixed paraffin-embedded benign fibrous histiocytomas using fluorescence in situ hybridization and RNA sequencing**. *Lab Invest.* 2015;95:1071-1076.

Hofvander J, Jo VY, Ghanei I, Gisselsson D, Mårtensson E, Mertens, F. **Comprehensive genetic analysis of a paediatric pleomorphic myxoid liposarcoma reveals near-haploidization and loss of the RB1 gene**. *Histopathology*. 2016;69:141-147.

Puls F, Hofvander J, Magnusson L, Nilsson J, Haywood E, Sumathi VP, Mangham DC, Kindblom LG, Mertens F. ***FN1-EGF*** **gene fusions are recurrent in calcifying aponeurotic fibroma**. *J Pathol.* 2016;238:502-507.

Walther C, Mayrhofer M, Nilsson J, Hofvander J, Jonson T, Mandahl N, Öra I, Gisselsson D, Mertens F. **Genetic heterogeneity in rhabdomyosarcoma revealed by SNP array analysis**. *Genes Chromosomes Cancer. 2016*;55:3-15.

Arbajian E, Puls F, Antonescu CR, Amary F, Sciot R, Debiec-Rychter M, Sumathi VP, Järås M, Magnusson L, Nilsson J, Hofvander J, Mertens F. **In-depth genetic analysis of sclerosing epithelioid fibrosarcoma reveals recurrent genomic alterations and potential treatment targets**. *Clin Cancer Res.* 2017;23:7426-7434.

Hofvander J*, Arbajian E*, Stenkula KG, Lindkvist-Petersson K, Larsson M, Nilsson J, Magnusson L, Vult von Steyern F, Rissler P, Hornick JL, Mertens F. **Frequent low-level mutations of protein kinase D2 in angiolipoma**. *J Pathol* 2017;241:578-582.

Al-Ibraheemi A, Folpe AL, Perez-Atayde AR, Perry K, Hofvander J, Arbajian E, Magnusson L, Nilsson J, Mertens F. **Aberrant receptor tyrosine kinase signaling in lipofibromatosis: a clinicopathological and molecular genetic study of 20 cases**. *Mod Pathol*. 2018; in press.

Puls F, Pillay N, Fagman H, Palin A, Rissler P, McCulloch T, Kindblom LG, Sumathi VP, Hansson M, Hofvander J, Magnusson L, Flanagan A, Mertens F. **Soft tissue tumors with** ***PRDM10*** **gene fusions: a clinicopathologic study of nine cases.** *Am J Surg Pathol*. 2018; in press.

Abbreviations

CXS	Sarcomas with complex genomes
DFSP	Dermatofibrosarcoma protuberans
ERMS	Embryonal rhabdomyosarcoma
ESV	Exonic structural variants
FFPE	Formalin-fixed paraffin-embedded
FISH	Fluorescence in situ hybridization
GCS	Genomic changes at SNP array
GIST	Gastrointestinal stromal tumor
IHC	Immunohistochemistry
Indel	Short insertions/deletions
LGFMS	Low-grade fibromyxoid sarcoma
LR	Local recurrence
Met	Metastasis
MFS	Myxofibrosarcoma
MLS	Myxoid liposarcoma
MPS	Massively parallel sequencing
PK	Protein kinase
PPT	PRDM10 positive tumors
PT	Primary tumor
RNA-seq	RNA sequencing
RT-PCR	Reverse transcriptase polymerase chain reaction
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
STT	Soft tissue tumors
TF	Transcription factor
UPS	Undifferentiated pleomorphic sarcoma
WDLS	Well-differentiated liposarcoma
WES	Whole exome sequencing
WGS	Whole genome sequencing

Introduction

Tumorigenesis

The human body is made up of trillions of cells that grow, divide and die in an orderly fashion, following a tightly regulated process called the cell cycle. As all new cells are generated from pre-existing cells via cell division, a dividing cell must first duplicate its genome before it can split and form a daughter cell. The very complex DNA replication system that performs this task includes multiple proofreading and repair mechanisms to ensure high fidelity. Nonetheless, the system is imperfect and a small number of errors are introduced into the genome of the daughter cell during each division. Such mutations in the DNA of a single cell can disrupt the cell cycle and generate a proliferative advantage for that cell, allowing it to outcompete the surrounding normal cells. Cells displaying such uncontrolled growth are referred to as neoplastic, and their expansion results in the formation of abnormal tissues commonly known as neoplasms or tumors.

This theory, that the transformation of a normal cell to a neoplastic cell is caused by genetic changes, was first presented over a century ago by the German zoologist Theodor Boveri (Boveri, 1914). Strong support for the theory was not observed until the 1960s with the discovery of the first recurrent somatic aberration in cancer, the so called Philadelphia chromosome in chronic myeloid leukemia. Since then, the validity of the hypothesis has been confirmed by numerous studies and it is now commonly accepted that all neoplasms arise as a result of genomic changes. Much work has gone into identifying the mutations underlying neoplastic transformation and in understanding the alterations in cell physiology that they convey.

In 2000 Hanahan & Weinberg published the influential paper “The hallmarks of cancer” in which they propose a conceptual model of tumorigenesis, describing six different physiological features essential for malignant transformation to occur. The neoplastic cells must become self-sufficient in growth signals, develop reduced sensitivity to growth-inhibitory signals, be able to avoid programmed cell death (apoptosis), and acquire unlimited replicative potential. Furthermore, as a tumor expands in size, it becomes essential to be able to induce vascular supply (angiogenesis) and malignant lesions need to obtain the ability to invade surrounding tissues and spread to distal sites via the blood or lymph system. Their proposed model for understanding cancer biology was further updated a decade later

(Hanahan and Weinberg, 2011) with two new emerging hallmarks, reprogramming of energy metabolism and evading immune destruction. Additionally, it has been suggested that an increased mutational rate, contributing to a rapid development of sub-populations with increased fitness, is a necessity, at least for malignant tumors (Loeb, 2016). It has thus become apparent that tumors are more than a lump of rapidly dividing cells, but rather a complex tissue where cancer cells interact and recruit normal cells to form a favorable microenviroment that contributes to the development of some of the proposed features (Hanahan and Weinberg, 2011; Wang *et al*., 2017).

It is unlikely that all these features can be gained through a single mutation. Instead, a combination of several mutations, accumulating over time, are generally believed to be needed in order for neoplastic transformation to occur. Indeed, most tumors, especially malignant ones, display numerous mutations ranging from large structural aberrations to single base pair substitutions and epigenetic changes.

The number and type of mutations can vary greatly between tumor entities, ranging from a handful in some pediatric tumors to thousands in tumors with impaired DNA repair systems (Vogelstein *et al*., 2013). Still, the majority of mutations found in tumors has little or no impact on tumorigenesis, they are so-called passenger mutations. The opposite, so called "driver-gene mutations" confer growth advantage (Vogelstein *et al*., 2013). When larger parts of chromosomes become deleted or duplicated it becomes difficult to identify which of the genes that are gained or lost that confer the growth advantage. Thus, the more complex the genome of a neoplasm, the more difficult it is to distinguish between driver and passenger mutations. By focusing on those that are frequently reported, that are present in early stages of tumor development, or that are accompanied by a limited number of additional mutations it has been possible to classify a large set of mutations with significant roles as drivers of tumorigenesis (Futreal *et al*., 2004; Bailey *et al.,* 2018). Genes harbouring driver mutations are further subdivided into those that positively or negatively regulate cell growth and survival, often referred to as oncogenes and tumor suppressor genes, respectively. The definitions though, are somewhat context-dependent since genes like *TP53* and *RET* can act as both tumor suppressor genes and oncogenes depending on type of mutation.

Even though there are countless ways by which oncogenes may be activated and tumor suppressor genes inactivated, three major mechanisms have emerged as particularly important for tumorigenesis; small genetic variants, chromosomal imbalances and gene fusions. All three mechanisms are readily observed in so-called soft tissue tumors.

Soft Tissue Tumors

Tumors that arise from cells of mesenchymal origin are called soft tissue tumors (STT). The mesenchymal soft tissue supports and connects the body organs and include adipose tissue, muscle, nerve sheaths, blood vessels, connective tissue and tendons (Goldblum *et al*., 2014). STT are a very heterogeneous group that display both varied levels of differentiation and a wide range of morphological appearances. There are currently more than 100 different histological subtypes that are diagnosed based on their resemblance to normal tissues (Fletcher *et al*., 2013a). The malignant STT, with the ability to invade both surrounding and distal tissues, are called sarcomas. Sarcomas are a very rare type of cancer, only accounting for about 1 % of all malignant neoplasms. The benign STT, however, are much more common, outcompeting their malignant counterparts by a factor of 200 (Fletcher *et al*., 2013a).

STT can arise anywhere in the body but they are primarily situated in the limbs, trunk wall, and intra-abdominally. The benign tumors tend to be smaller in size, generally less than 5 cm in diameter, and superficially located while sarcomas tend to be larger, median diameter of 9 cm, and two thirds are deeply situated.

Though sarcomas can affect patients at any point in life, the incidence increases with age and the median age of diagnosis is 65 years (Rydholm, 1983; Gustafson, 1994). Notably, the age distribution can vary greatly between tumor types and some, for instance embryonal rhabdomyosarcoma (ERMS), almost exclusively occur in children.

The most common form of treatment for STT is surgical removal. For aggressive sarcomas and/or when surgery with wide margins of normal tissue is not possible, radiotherapy and chemotherapy may be added. Targeted therapies are still uncommon in the treatment of STT, but gastrointestinal stromal tumors (GIST) with activating mutations in *KIT* or *PDGFRA* respond well to tyrosine kinase inhibitors, and many other drugs targeting specific cellular processes are currently underway (Dufresne *et al*., 2018).

The rarity of the tumors, in combination with overlapping morphology between subtypes, make accurate diagnosis challenging. Notably, it can sometimes be difficult to distinguish not only between sarcoma types but also between benign and malignant lesions. It is therefore of great importance to identify robust genetic markers that can improve accurate diagnosis.

Genetic Variants in STT

Small genetic variants

The most common type of genetic variation in cancer, as well as in constitutional DNA, is single nucleotide variants (SNVs) and short insertions/deletions (indels) of up to 10,000 nucleotides (Alexandrov and Stratton, 2014). Every individual is believed to deviate from the reference genome at 4-5 million sites (The 1000 Genomes Project Consortium, 2015) and neoplastic cells can display thousands of small genetic variants that are not observed in the corresponding constitutional DNA. The majority of this variation is seen outside of coding regions and their impact on tumorigenesis is thus difficult to predict. These alterations should not be completely discarded as noise since some have been described as important drivers of tumorigenesis. One example are SNVs affecting the *TERT* promoter region which have been reported in 80% of myxoid liposarcoma (MLS) and in a substantial fraction of solitary fibrous tumor (Killela *et al*., 2013). Still, mutations affecting coding regions, resulting in amino acid alteration, have attracted the most attention. Partly because most large scale sequencing studies have been performed by whole exome sequencing (WES), and partly because their pathogenetic consequences are easier to predict. SNVs can result in the alteration of a single amino acid (non-synonymous SNVs), introduction of a premature stop codon or changes in splice recognition sites. Indels can cause shifts in the open reading frame, resulting in novel amino acid sequences or truncated proteins.

Though most non-synonymous SNVs are regarded as noise, occurring prior to neoplastic transformation or simply being passenger events as a result of increased genetic instability, some non-synonymous SNVs seem to have a major impact on tumor development (driver mutations) resulting in activation of oncogenes or inactivation of tumor suppressor genes. Only a handful of such driver mutations have been identified as early events in sarcomas, including *KIT* and *PDGFRA* mutations in GIST, *RAS* signaling pathway mutations in ERMS, and *MYOD1* mutations in spindle cell rhabdomyosarcoma (Heinrich *et al*., 2003; Wardelmann *et al*., 2004; Agaram *et al*., 2014; Szuhai *et al*., 2014).

In a recent large scale sequencing study (TCGA, 2017) including six of the major classes of adult sarcomas, the number of reported non-synonymous somatic mutations was much lower than in the more common solid tumors of epithelial origin. It should be kept in mind, though, that the number of WES and WGS studies of sarcomas are still relatively few in comparison to the more common carcinomas for which many more driver mutations have been identified. The relatively low mutational burden in sarcomas may, at least in part, be explained by the presence of

other strong driver mutations in the form of structural and numerical rearrangements.

Structural and numerical rearrangements

Numerical chromosomal aberrations, ranging from gain or loss of individual chromosomes (aneusomy) to gain or loss of one or more copies of almost the entire genome (aneuploidy) is commonly observed in STT. This is particularly so in sarcomas, more than 80% of which display numerical chromosome aberrations when subjected to chromosome banding analysis (Mitelman *et al*., 2018).

Since these types of alterations can affect thousands of different genes, pinpointing their specific targets or estimating their pathogenetic consequences is extremely difficult. As a result, most of the characteristic numerical aberrations in sarcomas, such as gain of chromosome 8 in ERMS, remain poorly understood. In addition, copy number changes of parts of chromosomes are commonly observed in sarcomas. For instance, the *COL1A1-PDGFB* fusion in dermatofibrosarcoma protuberans (DFSP) is typically accompanied by gain of parts of chromosome arms 17q and 22q (Pedeutour *et al*., 1993, 1995).

Of particular interest are homozygous deletions, which often result in loss of tumor suppressor genes, and gene amplification; i.e., $\geq$3-5 fold gain of a DNA sequence in relation to surrounding sequences on the same chromosome (Santarius *et al*., 2010), resulting in increased expression of oncogenes. The most extensively studied example is probably the widespread amplicons in chromosome arm 12q in well-differentiated liposarcomas (WDLS) targeting the *CDK4*, *HMGA2*, and *MDM2* genes (Italiano *et al*., 2009; Kanojia *et al.,* 2015). Other recurrent high-level amplifications affect, e.g., distal 17q in malignant peripheral nerve sheath tumors and the *VGLL3* gene in 3p in various sarcomas (Hallor *et al*., 2009; Mantripragada *et al*., 2009; Helias-Rodzewicz *et al*., 2010; TCGA, 2017).

Gene fusions

Structural chromosome rearrangements such as translocations, inversions or deletions can cause a juxtapositioning of two previously independent genes, resulting in the formation of a fusion gene. This may lead to the translation of a deregulated and/or chimeric protein. Such gene fusions have been described in all types of neoplasia and can be found in about one third of sarcomas (Mertens *et al*., 2015, 2016; Yoshihara *et al*., 2015; Mitelman *et al*., 2018). Though some gene fusions appear to be passenger events, being a by-product of the extensive genomic rearrangements that are observed in many malignant tumors, others are likely to have a strong impact on tumor development. Gene fusions that are recurrent or are

associated with relatively few other mutations are often suggested to have a significant pathogenetic role. For some cases, this has been further supported by *in vitro* studies and experimental animal models, showing that a gene fusion is sometimes sufficient for malignant transformation (Haldar *et al*., 2007; Riggi *et al*., 2007; Straessler *et al*., 2013).

The detection of fusion genes in sarcomas is of high clinical significance as many fusions are strongly associated with one or a few morphologic subtypes, rendering them ideal as diagnostic and prognostic markers (Mitelman *et al*., 2007). In addition, some chimeric proteins constitute promising therapeutic targets (Højfeldt *et al*., 2013; Feng *et al*., 2014; Parker *et al*., 2014) and pharmacological treatment of sarcomas displaying fusions that activate growth factors, such as PDGFB, or protein kinases, such as ALK, is already in clinical use. In 1992, the first gene fusion was described in sarcomas, *EWSR1-FLI1* in Ewing sarcoma (Delattre *et al*., 1992). Since then close to 200 fusions have been reported in STT and more than half of them are recurrent in a specific subtype (Mitelman *et al*., 2018).

The most common type of genes involved in fusions in STT are so called transcriptional regulators, including transcription factors (TFs) and co-activators/co-repressors (Mertens *et al*., 2016). Around two-thirds of gene fusions in STT include such genes, typically as the 3'-partner (Mertens *et al*., 2016). TFs are generally subdivided into classes and families on the basis of their DNA-binding domains (Wingender *et al*., 2015) and a specific type of TF is typically seen in only one tumor type; for example, it is only Ewing sarcoma that displays recurrent fusions involving an Ets-related factor as the carboxyterminal partner. Some explanation for this specificity has been demonstrated in experimental systems where studies of chimeric transcripts involving TFs, such as *EWSR1-FLI1* or *EWSR1-ATF1*, show that only certain cell types can be transformed and that the affected genetic programs and phenotypic effects vary dependent on in which cell it is expressed (Haldar *et al*., 2007; Straessler *et al*., 2013).

Protein kinases (PKs), primarily receptor tyrosine kinases, are also commonly observed in STT fusions. The PK-encoding gene is always the 3'-partner and, in contrast to the gene fusions involving TFs, a large variety of different 5'-partners can be observed. The fusions result in activation of the kinase domain and the main role of the 5'-partner is to ensure a high expression of the chimeric transcript by providing a more active promoter. Gene fusions involving PKs seem less tissue-specific than those affecting TFs, for instance the *ETV6-NTRK3* and *EML4-ALK* fusions occur in STT as well as in a variety of other neoplasms (Mitelman *et al*., 2018). In addition, fusions involving receptor tyrosine kinases constitute excellent therapeutic targets, in both sarcomas and other malignancies, due to their high susceptibility to kinase inhibitors (Shaw *et al*., 2013; Lovly *et al*., 2014; Stransky *et al*., 2014).

Proteins involved in epigenetic regulation in the form of chromatin modification and remodelling have emerged as significant actors in tumorigenesis (Chen and Dent, 2014; McBride and Kadoch, 2018). Where fusions involving TFs are thought to confer target specificity by binding a specific DNA motif, fusions causing chromatin deregulation may result in genome wide alterations of gene expression. This could to some extent explain why sarcomas with such fusions are either undifferentiated, like undifferentiated round cell sarcomas with the *BCOR*-*CCNB3* fusion, or display disparate lines of differentiation, such as ossifying fibromyxoid tumor with *PHF1* fusions or synovial sarcoma with *SS18*-*SSX* fusions. Possibly, the successful introduction of DNA methylation and histone deacetylase inhibitors for treatment of other cancers might also become useful for epigenetic treatment of some sarcomas (Højfeldt *et al*., 2013).

The present study

Aims

The overall aims of my thesis were to

Article I - Investigate to what extent sarcomas with unique structural rearrangements display gene fusions when subjected to RNA-sequencing.

Article II - To search for novel gene fusions in undifferentiated pleomorphic sarcoma.

Article III - To study in more detail the cellular effects of *PRDM10* fusions.

Article IV - To compare clonal evolution in tumors arising through different mechanisms. We assessed the dynamics of chromosome and nucleotide level mutations by cytogenetics, SNP array and WES in three types of sarcoma; amplicon-driven WDLS, gene fusion-driven MLS, and sarcomas with complex genomes (CXS).

Materials and Methods

Below follows a brief desciption of the main methods used in the articles included in this thesis.

Patients and tumor samples

The tumor material used for our studies were either fresh frozen tumor biopsies or formalin-fixed paraffin-embedded (FFPE) tumor blocks. Samples were retrieved from the sarcoma centres at Lund University Hospital and the Karolinska Hospital, Stockholm. Additional samples were obtained from collaborators at Sahlgrenska University Hospital or Royal Orthopaedic Hospital NHS Foundation Trust, Birmingham, UK.

The studies were approved by the local ethical committees.

SNP array

Copy number changes, i.e., gains and losses of genomic material, as well as the allele frequency (copy neutral LOH) can be detected using single nucleotide polymorphism (SNP) arrays. SNPs are defined as inter-individual single nucleotide variations between homologous chromosomes naturally occurring in at least 1 % of the population. The SNP arrays consist of millions of oligonucleotide probes, homologous to known SNPs, attached to a surface. A fragmented DNA sample is allowed to hybridize to the probes and the amount of DNA bound to each probe is quantified by fluorescent or light-absorbing tags and yields two types of values. The signal intensity is normalized against the average signal and log2-transformed to give the log ratio value, indicating the copy number. The second value is the B-allele frequency, indicating allelic distribution at that locus.

SNP array analyses were performed in articles III and IV using a combination of Illumina and Affymetrix arrays, and were performed on both fresh frozen, and FFPE material. Though the general principles are the same for the different arrays, their resolution varies due to the number and distribution of probes.

SNP arrays only measure the amount of DNA and thus give no information of how DNA fragments are connected. This limitation leads to the inability to detect balanced chromosomal rearrangements. Thus, many gene fusions associated with soft tissue tumors cannot be detected since they do not result in copy number shifts. Also, co-amplified sequences, such as the material on ring chromosomes in WDLS, are not correctly visualized and are instead seen as separate amplicons in one or more chromosomes. In addition, the results are highly dependent on the admixture

of normal and tumor cells. Tumor-associated imbalances will not be detected if the tumor cells constitute less than 15-20% of the sample.

Targeted approaches to gene fusion discovery

Traditionally, gene fusions have been identified through a multistep procedure, starting with mapping of chromosomal rearrangements to specific chromosome bands with G-banding analysis. Then, breakpoint regions could be more specifically pinpointed by fluorescence in situ hybridization (FISH) and potential fusion transcripts directly tested with reverse transcriptase polymerase chain reaction (RT-PCR; Mertens and Tayebwa, 2013). Though the use of these techniques for novel gene fusion discovery has largely been replaced by massively parallel sequencing (MPS) approaches, they are still widely used for validation of MPS results.

FISH analysis is used to visualize genomic loci in single cells. By using fluorescently labelled probes, complementary to DNA sequences of interest, both numerical and structural rearrangements can be detected by fluorescence microscopy (Trask, 1991). FISH can be particularly useful for detecting gene rearrangements by so-called break-apart probes, as exemplified in article II where the status of the *PRDM10* locus was investigated by using two probes in different colours flanking the gene. If the gene locus is intact, the two signals remain close to each other, but if the gene is affected by a structural rearrangement, in this case a translocation, the probe is split and seen as separate signals. Interphase FISH, performed in article II, can be used on both FFPE sections or fixed cells while metaphase FISH can be used on fresh tumor material if cells are cultured prior to fixation. Drawbacks with the technique include both false positive and false negative results due to unspecific probe binding or imperfect hybridization, respectively. In addition, FISH is unable to detect smaller copy number changes and low-frequency aberrations.

RT-PCR was used to verify chimeric transcripts in all of the articles included in this thesis. Briefly, RT-PCR is used to detect gene expression through the creation of cDNA from RNA by reverse transcription. The cDNA from the fusion product can then be amplified by traditional PCR using sequence-specific primers. PCR products of expected size can be identified and separated by gel electrophoresis and Sanger sequencing used to confirm the identity of the involved genes. Technically, both RNA from fresh frozen and FFPE samples can be used for this target approach; however, the usually highly fragmented RNA obtained from FFPE samples makes the technique more prone to failure.

Massively parallel sequencing

Though the above mentioned methods are useful for either copy number analysis or gene fusion detection, they suffer from having insufficient resolution (SNP array) or from being directed (FISH, RT-PCR). MPS, also known as next generation sequencing or high throughput sequencing, overcomes these obstacles by providing both width and depth in a single analysis via simultaneous sequencing of millions of DNA or RNA fragments. This rapid and large scale generation of data has revolutionized genetics and its widespread success has resulted in the development of multiple applications. It is now possible to analyse whole genomes (WGS), whole exomes (WES), transcriptomes and targeted regions providing detailed information on genomic structures, SNVs, fusion transcripts, gene expression profiles, accessible chromatin regions and much more. The main applications of MPS used in this study were WES and RNA-seq.

It is not surprising that RNA-seq has become immensely popular when searching for gene fusions, as RNA-seq will, in theory, detect both known as well as previously unknown fusions. Its power can be illustrated by the fact that the number of known STT-associated gene fusions has increased from 44 in 2008 to 191 today, ten years later (Figure 1; Mitelman *et al*., 2018).

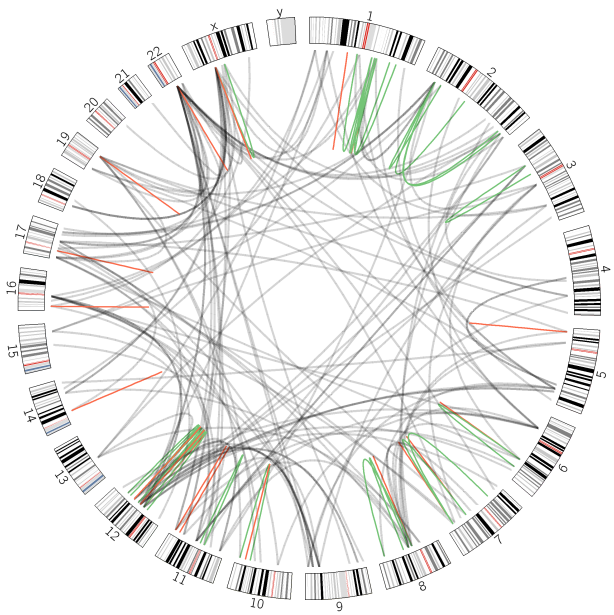


Figure 1. The 191 gene fusions reported in SST. Each fusion is indicated by a line; black lines indicate fusions between genes located on different chromosomes, green lines indicated intra-chromosomal fusions and red lines indicate fusions between genes located in the same cytogenetic band (Data from Mitelman *et al*., 2018).

The technique is somewhat limited by the quality of the RNA, the depth of the analysis and the inability to detect fusion events resulting in transcriptional silencing or promotor swapping. Another drawback of RNA-seq is the often high number of false positive discoveries generated as a result of incorrect mapping, trans-splicing or template switching (Ozsolak and Milos, 2011; Mertens and Tayebwa, 2013). RNA-seq was used in articles I, II and III with the purpose of identifying chimeric transcripts and in article III to study global gene expression patterns.

WES was performed in article IV with the aim of identifying SNVs and indels. WES focuses on sequencing only the exons, the protein coding regions of the genome. As exons only constitute roughly 1 % of the human DNA, the costs and time needed for analysis is less than for WGS. Despite recent efforts in analysing high grade sarcomas (Shern *et al*., 2014; TCGA, 2017) the spectrum of SNVs and indels in many sarcoma subtypes still remains poorly explored.

One obvious drawback of this approach is the possibility to miss important mutations occurring outside of exons. Additionally, as individuals differ from the reference genome at thousands of nucleotide positions, identification of relevant somatic mutations are difficult without data from the corresponding normal sample. This increases the cost of the analysis, but it should be pointed out that WES can also be used to obtain copy number information and thus, potentially, could replace genomic arrays (Hehir-Kwa *et al*., 2015) to lower the overall costs of genetic analysis.

Library preparations and sequencing

For samples from which high quality RNA could be obtained, libraries were prepared using the Truseq RNA sample preparation kit v2 (Illumina, San Diego, USA), while libraries from samples with low quality RNA were prepared using the Truseq RNA Access library Prep kit (Illumina).

The main difference between the two kits is their approach for enrichment of protein-coding RNA. For high quality samples, RNA is enriched for by poly-A tail binding magnetic beads. Enriched RNA is then fragmented before cDNA is generated by reverse transcription. Adaptor sequences are then ligated to cDNA ends before the fragments are PCR amplified prior to sequencing. For low quality RNA, cDNA generation and adaptor ligation is performed prior to the enrichment, which is based on hybridization to capturing probes.

Exome libraries were generated using the Nextera Rapid capture Exomes kit (Illumina). Briefly, genomic DNA is tagmented, i.e., concurrently fragmented and tagged with adaptor sequences, using the Tn5 transposome. Tagged DNA fragments are amplified via PCR, simultaneously introducing sample specific indexes.

Fragments corresponding to coding sequences (exons) are then enriched for by hybridization to a pool of capturing probes.

The Illumina sequencing technology was used through this thesis, and works in the same way independent on how the libraries were generated. Briefly, the prepared libraries are loaded onto a flow cell which is coated with two types of oligonucleotides that are attached to the surface. The sample fragments hybridize to the oligonucleotides on the flow cell and a complement of the bound fragment is generated by a polymerase. The template strands are then clonally amplified by bridge-amplification to generate millions of clusters. Cluster generation is followed by the actual sequencing process, called sequencing by synthesis. This is a three-step cycle starting with the addition of modified nucleotides containing fluorescently tagged reversible terminators which compete for binding to the template strand. Once a nucleotide is bound it blocks additional elongation to ensure that only one nucleotide is incorporated per cycle. The clusters are excited by a light source and a nucleotide- specific fluorescent signal is emitted and registered. The fluorophore is then cleaved off and washed away. The cycle can then be restarted as the cleaving allows for additional incorporation of new nucleotides (Goodwin *et al*., 2016).

We have primarily used the NextSeq500 (Illumina) sequencing machine, which, if successfully run, generates around 400 million paired end reads.

Bioinformatics

The advances of MPS technologies have radically increased the amount of genomic information and the speed at which it can be obtained, resulting in the generation of very large datasets. For instance, The Cancer Genome Atlas (TCGA) dataset includes genomic information from 11 000 patients from 33 types of cancer and currently hosts more than 2.5 petabytes of publicly available sequencing data (https://cancergenome.nih.gov). It is thus not surprising that the ability to handle such data has become more or less essential for modern genomic research. As a result, the field of bioinformatics is also rapidly evolving and a continuously growing plethora of tools and algorithms is available to aid in data analysis. Below follow the main steps and tools used to analyze MPS data presented in this thesis.

First, the raw output from the sequencing machine is converted to a platform-independent file format called fastq. The fastq file is a text file containing information for each individual read including a unique read name, the nucleotide sequence for the read and a per base quality score representing the probability that the called base is a sequencing error.

In article IV, WES data were used for identification of SNVs and indels, and processing largely followed published best practise guidelines. Briefly, after initial

quality control and trimming of remaining synthetic adaptor sequences, the reads were aligned to the reference genome using the bwa mem algorithm (Li, 2013). Post-mapping processing was performed using the Picard (https://broadinstitute.github.io/picard/) and GATK (McKenna *et al*., 2010; DePristo *et al*., 2011) toolkits. Somatic variant calling was performed using MuTect (Cibulskis *et al*., 2013) and Strelka (Saunders *et al*., 2012). Called variants were annotated using VEP (McLaren *et al*., 2016).

In article III, the RNA-seq data were used to study global gene expression patterns and identify differentially expressed genes among different tumor types. Raw reads were aligned to the reference genome using the STAR aligner (Dobin *et al*., 2013). Though accurate alignment of reads generated from DNA sequencing experiments is a difficult task (Reinert *et al*., 2015), RNA-seq data poses the additional challenge of aligning the reads to non-contiguous genomic regions as they often span different exons.

Using the cufflinks suite (including cuffquant and cuffnorm) the number of reads mapping to specific genes is counted and the gene expression values are calculated as fragments per kilobase of transcript per million reads (FPKM). This normalization takes into account both the gene length and number of mapped reads for a sample, allowing for both between gene and between sample comparisons.

In articles I-III, multiple different gene fusion finding algorithms, including ChimeraScan (Iyer *et al*., 2011), TopHat-Fusion (Kim *et al*., 2013), SOAPfuse (Jia *et al*., 2013), FusionCatcher (Nicorici *et al*., 2014) and STAR-Fusion (Haas *et al*., 2017), were used. Comparisons of these software have reported large variation in false discovery rates, specificity and computational requirements (Carrara *et al*., 2013; Kumar *et al*., 2016) and we often find it necessary to run multiple algorithms, particularly if the sample stems from FFPE material, to achieve reliable results.

Results and Discussion

Article I

In this study, we wanted to evaluate to what extent seemingly unique structural aberrations result in a functional fusion transcript and assess the advantages of using a combination of RNA-seq and cytogenetic data to identify them. We therefore selected nine samples from eight sarcoma patients displaying “simple” karyotypes, harbouring only one or a few structural rearrangements that did not correspond to any known fusion genes, and performed RNA-seq with the aim of identifying chimeric transcripts.

The RNA-seq data were investigated by three different gene fusion-detecting algorithms: TopHat-Fusion (Kim *et al*., 2013), SOAPfuse (Jia *et al.*, 2013), and ChimeraScan (Iyer *et al.*, 2011). The results varied greatly between the different algorithms. TopHat reported a total of 26 potential fusion transcripts whereas SOAPfuse reported 81 and ChimeraScan 1,329. The list of potential fusions was filtered based on several criteria. Fusion events were correlated with the cytogenetic data for the individual samples, retaining genes that were located close to the breakpoints in the karyotype. Transcripts involving genes previously reported in fusions were considered. Reported transcripts that did not match the above criteria were discarded if they had no reads spanning the fusion junction, had less than five reads surrounding the fusion junction, were regarded as read-through transcripts, or involved pseudogenes. After the filtering, six chimeric transcripts reported by TopHat, two by SOAPfuse and six by ChimeraScan remained, and RT-PCR was performed to verify these events. We were able to confirm five different fusions, three of them being novel.

Case 3 was a myxofibrosarcoma (MFS) having a translocation involving chromosomes 2 and 6 as well as ring chromosomes involving chromosomes 9 and 12. TopHat reported the fusions, *AFF3-PHF1*, which correlated well with the translocation and could be confirmed by RT-PCR. Both *PHF1* and *AFF3* have been described in other fusion events before but have never been reported together (Mitelman *et al*., 2018). This novel fusion results in an out-of-frame transcript, making it difficult to speculate on its pathogenetic impact. However, it should be noted that rearrangements involving *PHF1* have been suggested to be important for tumor development in other sarcomas (Gebre-Medhin *et al.*, 2012; Antonescu *et al*., 2014).

Unsurprisingly, several fusions correlating with the material present in the ring chromosome were reported. Out of these, the in frame *KIAA2026-NUDT11* and out of frame *CCBL1-ARL1* could be verified by RT-PCR. As ring chromosomes undergo a series of breakage-fusion-bridge events, causing the DNA to continuously break and re-join during each cell division (Gisselsson *et al*., 1999; Gebhart, 2008), these fusions might simply be chance events. Additionally, the *CCBL1- ARL1* fusion resulted in an out-of-frame transcript and *KIAA2026* is an uncharacterized gene, thus, speculation on the importance, if any, for tumor development is premature.

Case 4 had been initially diagnosed as a benign fibroblastic-myofibroblastic lesion based on the preoperative fine- and core-needle aspirates. After surgical excision of the tumor, histopathologic analysis suggested a malignant tumor, but a precise diagnosis could not be reached. The karyotype displayed a balanced three-way translocation t(7;13;11)(q32;q34;q23), which did not provide any diagnostic clues. However, the RNA-seq reported a *FUS-CREB3L2* fusion, which was confirmed by RT-PCR. This fusion has been reported in 76–96% of low-grade fibromyxoid

sarcoma (LGFMS), but is usually seen as either a t(7;16)(q33;p11) or as a ring chromosome at the cytogenetic level (Folpe *et al*., 2013). An external review of the morphology indeed suggested LGFMS and a correct diagnosis was eventually reached. This finding serves as an example that translocations giving rise to characteristic gene fusions sometimes are masked as more complex rearrangements (Mitelman *et al*., 2007).

The karyotype of case 8 displayed a complex exchange of material between chromosomes 5 and 8 as the sole aberration. The tumor, from a 26-year-old woman, was initially diagnosed as MLS despite the lack of cytogenetic support for this diagnosis; MLS displays a *FUS-DDIT3* or *EWSR1-DDIT3* fusion in close to 100% of the cases. As MLS is a malignant tumor, the patient was checked regularly for local recurrences and lung metastases. However, RNA-seq and RT-PCR could confirm a *HAS2-PLAG1* fusion which is specific for a benign tumor type, lipoblastoma (Hibbard *et al*., 2000). The two tumor types are usually not mistaken for one another, despite being morphologically very similar, as lipoblastomas are extremely rare in adult patients; 90% of lipoblastomas occur in children below 3 years of age (Weiss and Goldblum, 2008). As a result of the *HAS2-PLAG1* finding, the morphology was re-reviewed and the diagnosis altered to lipoblastoma. The finding in case 8 illustrates that RNA-seq can provide vital information for differential diagnosis, and also highlights the importance of robust genetic markers as some benign soft tissue tumors are morphologically similar to sarcomas.

No fusions were confirmed in the remaining five cases. Whether this is because their rearrangements did not result in the generation of a fusion gene or because we were unable to detect them is unknown.

Some chimeric genes are technically impossible to detect with the used RNA-seq approach, which is based on mRNA enrichment by poly-A tail selection. Hence, chimeric transcripts lacking poly-A tails are not sequenced and therefore missed. An alternative way to identify these types of fusion would be by total RNA-seq. This technique, however, would still be unable to identify gene fusions arising through promotor swapping as is does not result in a chimeric transcript (Ozsolak and Milos, 2011).

To be able to say with greater confidence that the cytogenetic aberrations observed in the negative cases did not results in a fusion gene, WGS could have been performed. This would give exact information of the genomic breakpoints that gave rise to the aberrations, and potential fusion transcripts could be verified with RT-PCR.

The article also highlights some of the bioinformatic limitations of gene fusion discovery, namely the high generation of false positive results and the discrepancies between algorithms. From the results of the present study it was clear that the use

of only one of any of the three programs was insufficient for accurate analysis as none of the algorithms was able to identify all of the verified fusions. Additionally, only two of the fusions were independently detected by more than one algorithm.

In conclusion, we showed that there is an increased likelihood of finding novel gene fusions by RNA-seq of tumors displaying simple structural rearrangements and that karyotypes, in the absence of WGS data, are valuable when evaluating the significance of identified fusion transcripts.

Articles II and III

UPS is one of the most common sarcoma subtypes, accounting for up to 20% of cases in adults (Fletcher *et al*., 2013b). It is generally an aggressive tumor type, associated with a high metastatic rate and poor prognosis. There is currently no specific treatment available for UPS patients (Goldblum, 2014). UPS lacks any defined line of differentiation and tends to be morphologically heterogeneous, partly overlapping with other sarcomas; however, all cases share a marked cellular pleomorphism. This diagnosis of exclusion is likely to encompass multiple sarcoma subtypes, thus representing a common morphological state rather than a distinct tumor entity (Fletcher *et al*., 2001). Most UPS cases display highly complex karyotypes and copy-number profiles (Gibault *et al*., 2011; Fletcher *et al*., 2013b; Guled *et al*., 2014; Mitelman *et al*., 2018), however, the genomic complexity varies considerably, with a small subset harboring only a few structural and/or numerical aberrations.

In Article II, two such cases of UPS with simple karyotypes were subjected to RNA-seq, identifying two novel gene fusions: *CITED2-PRDM10* and *MED12-PRDM10*. The chimeric transcripts could be verified by RT-PCR, and FISH showed a break also at the genomic level in the *PRDM10* gene for one of the cases. By using qPCR with probes for both the 5' and 3' parts of *PRDM10*, higher expression of the latter part of the gene was observed in both cases. Since no specific recurrent gene fusion had previously been identified in UPS, we wanted to evaluate the frequency of *PRDM10* fusions in a larger series. We collected an extended cohort of 82 sarcomas which were in part selected based on their karyotypes, having structural rearrangements correlating with the location of the *MED12*, *CITED2*, or *PRDM10* genes in chromosome arms Xq, 6q, and 11q, respectively. In addition, 16 tumors were selected based on their diagnosis as low-grade malignant UPS, MFS, or leiomyosarcoma. Using the same probes for *PRDM10* as in the two index cases, 78 tumors from the extended cohort were screened with qPCR with the aim of identifying differential expression between the 5' and 3' parts of the gene. None of the tumors showed the same increased expression of the 3' part of *PRDM10* that had been observed in the two fusion positive cases. However, six cases displayed a decreased expression and RT-PCR could verify a *MED12-PRDM10* fusion in one

of these cases. The remaining samples, four MFSs, were analyzed by RNA-seq but no fusion transcript involving *PRDM10* was identified.

Among the three fusion-positive cases, no distinct morphologic features, setting them apart from other UPS, could be identified. They were, however, classified as low-grade malignant tumors, on the basis of lower mitotic counts, when re-reviewed and none of the three patients developed metastases. The findings thus suggested that a subset of UPS that is less aggressive than classical high-grade UPS could be identified, and that these patients might benefit from less extensive treatment.

In article III, we had collected an extended series of eight PRDM10-positive tumors (PPT), and their genomic and transcriptomic features were characterized by RNA-seq, SNP array and WGS. The gene expression of PPT was compared with that of regular, high-grade UPS and other morphologically similar tumor types including MFS, myxoinflammatory fibroblastic sarcomas, DFSP, and benign fibrous histiocytoma. The PPT tumors formed a distinct cluster, easily distinguishable from the other tumor entities, by unsupervised hierarchical clustering of the gene expression data. *PRDM10* was not among the most differentially expressed genes, and this, in combination with the fact that fusion events were generally supported by few chimeric reads, suggests that the *PRDM10* fusions are expressed at relatively low levels in the tumors and hence might be suboptimal as molecular markers. Instead, we identified the surface receptor-encoding *CADM3* gene as one of the most differentially expressed genes and verified its potential as a differential marker at the protein level by immunohistochemistry (IHC).

A clear difference between PPT and UPS was also observed at the genomic level. G-banding, SNP array and WGS analyses identified few structural variants in PPT. In contrast, high grade UPS has been reported to have one of the highest numbers of structural variants among all cancers analysed so far (TCGA, 2017). For the two cases with WGS data, the number of reported SNVs and indels was substantially lower than what has been described for UPS and MFS. This lack of secondary aberrations in PPT is a strong indication that the reported *PRDM10* fusions are the main drivers of tumorigenesis for these cases.

The impact of the *CITED2-PRDM10* fusion was further evaluated *in vitro* by comparing cell lines expressing either the fusion or an empty vector. By RNA-seq, a significant part of the gene expression profile observed in the tumors was recapitulated in the cell lines, including high expression of *CADM3*. Staining for CADM3 by IHC confirmed its expression also on the protein level. An assay for transposase-accessible chromatin (ATAC-seq) was performed to study genome-wide changes in chromatin accessibility (Buenrostro *et al*., 2015). Expression of the fusion transcript seemed to have major effects on histone regulation as numerous regions had differential accessibility when comparing the cell lines. Interestingly, *de novo* motif discovery identified the PRDM10 motif as the most enriched

transcription factor-binding motif in the regions differentially open in cell lines expressing *CITED2-PRDM10*. In summary, the data indicate that the fusion gene accounts for much of the variation in gene expression observed in PPT and that many of the upregulated genes are direct targets of the fusion protein as they harbour open regions containing the PRDM10 binding motif.

Article III also strengthened the hypothesis that PPT is clinically important to recognize; none of the tumors in the extended cohort, compared to around 30% of high-grade UPS (Fletcher *et al*., 2013b), metastasized. Indeed, a recent investigation of the morphological features of PPT showed that they are consistently associated with a low mitotic count and a good prognosis (Puls *et al*., 2018).

In conclusion, the marked differences in both clinical outcome and genomic complexity provide compelling evidence that PPT is a distinct tumor type, separate from classical UPS.

Article IV

Genetic instability is thought to be an essential feature of cancer cells (Cahill *et al*., 1999; Hanahan and Weinberg, 2011; Loeb, 2016). As a result, many tumors display extensive intratumoral heterogeneity with regard to genetic alterations and clonal evolution is often observed in tumors that are repeatedly sampled during disease progression (Nowell, 1976; Loeb, 1991; Shah *et al.,* 2009; Gerlinger *et al*., 2012; Burrell *et al*., 2013; Heim and Mitelman, 2015). However, most of the conclusions have been drawn from data on highly malignant epithelial neoplasms in adults which might not be representative for other solid tumors or hematopoietic malignancies that arise through different pathogenetic mechanisms. While it is known that different tumor types show different mutational profiles and that SNVs predominate over chromosomal rearrangements in some tumors and oppositely in others (Nowak *et al.,* 2002; Vogelstein *et al.,* 2013), it is poorly explored to what extent these elements affect clonal evolution. Additionally, data on tumors that have been followed for long periods of time are scarce.

To compare clonal evolution in tumors arising through different mechanisms, we selected three types of sarcoma; amplicon-driven WDLS, gene fusion-driven MLS, and sarcomas characterized by complex genomic rearrangements (CXS) that had been followed for one to 25 years. We investigated the dynamics of chromosome and nucleotide level mutations by cytogenetics, SNP array analysis and WES. In addition to the longitudinal aspect of clonal heterogeneity, we could study intratumoral heterogeneity in four WDLS and two CXS, as well as inter-cellular variation at the chromosome level in all WDLS and 15 MLS lesions.

WDLS displays supernumerary ring chromosomes including amplified material from multiple genomic regions. The amplified material always contains a

substantial portion of chromosome arm 12q, with the genes *MDM2*, *CDK4*, and *HMGA2* being the most important targets (Italiano *et al*., 2009; Kanojia *et al*., 2015). The mitotic instability of the ring chromosomes induces extensive inter-cellular genetic variation (Gisselsson *et al*., 2000). From five patients with WDLS both WES, SNP array and G-banding data were obtained from 20 samples from 12 lesions. Time interval between first and last sampling ranged from 57–306 months.

At G-banding analysis, substantial inter-cellular variation was observed including considerable difference in both the size and number of ring chromosomes. Additionally, there were both numerical and structural non-clonal changes in 42% of the cells. This extensive variation was not reflected in the SNP array or WES results. Three spatially separated samples from the same primary tumor (PT) could be analyzed in four cases but no differences were found. The 12 tumors had a median of 35 GCS at SNP array and almost all were gains. When any two lesions from the same patient was compared, the median number of shared breakpoints was 49% and the median overlap of the total extension of GCS was 0.57. Greater overlap was observed for the amplified sequences in chromosome 12 for which the corresponding values were 65% and 0.71. There was no indication that the samples became less similar with time and the amount of GCS did not increase at relapse. In fact, the two patients from which three samples could be obtained showed higher similarity between the first and last sample (0.97 and 0.99) than between the second and third or first and second samples (0.69–0.72) when the overlap of GCS on chromosome 12 was compared. The amplified material varied greatly between patients, the only region amplified in all 12 samples was a discontinuous 856 Kb sequence in 12q14–15. In line with previous data, these core amplicons included *MDM2* and the first three exons of *HMGA2*, suggested to be essential for tumorigenesis (Italiano *et al*., 2009).

The number of reported SNVs and indels (exonic structural variants; ESV) was low for WDLS samples, median 7, and usually presented at low allele frequencies, median 21%. The intra-lesional heterogeneity was also low, with 82-100% of the ESV present in all samples from the same PT. However, over time, the majority of the mutations seemed unique for each lesion and the number of EVSs only moderately increased at relapse. Of the 72 detected mutations only three were shared with another lesion. None of the mutations was shared by different patients and none has been reported in soft tissue tumors previously. The WES data strongly imply that ESV have little or no significance in WDLS development. Additionally, the scarcity or absence of ESV that were shared by all lesions from the same patient implies that their progenitor cell has undergone far fewer cell divisions prior to neoplastic transformation than a typical precursor cell in a carcinoma.

These results are in line with the suggestion that tumors might reach a genetic fitness maximum relatively early given a stable microenvironment (Loyed *et al.,* 2016).

Indeed, all WDLS follow-up samples were LR and none of the patients had received any chemotherapy that could have shifted the selection pressure.

For fusion-driven sarcomas, we selected MLS displaying the *FUS-DDIT3* fusion gene, which is considered to be a strong driver mutation (Riggi *et al*., 2006). MLS display few recurrent chromosomal imbalances, notably trisomy 8 and idic(7)(p11) (Mandahl *et al*., 1994) and exonic SNVs are scarce and few are seen in more than 15% of the cases (Barretina *et al*., 2010). However, a frequent mutation affecting the promoter region of the TERT gene is seen in around 70–90% of the tumors (Killela *et al*., 2013; Koelsche *et al*., 2014). The clinical behavior of MLS varies substantially, around 15-35% of the patients develop metastases, and it has been suggested that specific mutations, e.g., in *PIK3CA* and *TP53*, are associated with aggressive behavior (Kilpatrick *et al*., 1996; Oda *et al*., 2005; Joseph *et al*., 2014).

We included nine PT from *FUS-DDIT3*-positive MLS and 1–4 LR and/or metastases (Met), occurring 12–104 months after diagnosis. The inter-cellular variation at G-banding was very low among the 15 samples that could be assessed; only 1.3% of cells showed non-clonal structural aberrations and karyotypes were consistently identical when comparing 2–3 samples from the same PT.

When combining cytogenetic and SNP array data, few chromosome level aberrations were found per PT (1-6) and there were few differences between a PT and its LR or Met (0-8). Two LR, cases 1 and 6, had fewer chromosome aberrations than their PT and 6/13 Met had the same number as their corresponding PT.

WES was performed on 11 samples from four patients and reported 7–165 (median 15.5) ESV per PT. In cases 1-3, the majority (61–100%) of the ESV detected in the PT was also present at relapse. In case 4, a dramatic decrease was seen, from 165 ESV in the PT to only 11–24 ESV in the four Met. Still, the clonal relationship between the PT and the Met was unquestionable as six ESV and six chromosome level aberrations were shared by all samples. The 165 ESV in PT included well-known cancer-associated genes such as *BCOR*, *CHEK2*, and *TP53* which have been implicated in MLS progression before (Oda *et al*., 2005; Barretina *et al*., 2010; Joseph *et al*., 2014). Notably, the allele frequencies of *CHEK2* (54–68%) and *TP53* (36–43%) mutations suggest that they occurred early, potentially triggering the massive accumulation of ESV. In contrast, the six ESV shared by all samples occurred at low allele frequencies (5–10%) in the PT and then increased in all the Met. This suggests that a small subclone in PT, with lower nucleotide level instability, gave rise to all the Met.

The results show that MLS cells are genetically relatively stable with a slow clonal evolution at the chromosome level, displaying few deviations from the stemline, even in metastatic lesions. There was a more pronounced accumulation of ESV as the relapse samples had more ESV than the PT, case 4 being an extreme exception.

Our results are in agreement with data on pancreatic carcinomas, for which it has been suggested that cells which eventually form metastases may arise relatively early in the PT as metastases share most if not all important driver mutations with their PT (Makohon-Moore *et al* 2017; Reiter *et al*., 2017).

However, we cannot exclude that much of the morphological and clinical variation in MLS, such as the transition from a low-grade to a high-grade tumor in cases 3, 6, and 9, could be caused by epigenetic factors or mutations in non-coding sequences. Furthermore, the findings in case 4 demonstrate that analysis of the PT might suggest therapeutic targets that are not present in the metastatic lesions which is in contrast to the more common notion that mutations in small subclones of a PT might be overlooked when only a single sample is analysed.

For comparison with gene fusion- and amplicon-driven liposarcomas, we analyzed 6 CXS with 2–3 lesions per case occurring 77–294 months after diagnosis. Though the pathogenetic mechanisms in CXS sarcomas remains sparsely investigated, it is well known that there exists an extensive genetic and clinical variation both between and within morphologic subgroups (Chibon *et al*., 2010; Fletcher *et al*., 2013a; Heim and Mitelman, 2015).

Intra-lesional heterogeneity was studied in two cases, where 2 or 3 samples from the PT could be analyzed using both SNP array and WES. In case 20, one of the three samples had 7 additional imbalances at SNP array analysis while no differences were seen between the samples in case 18.

SNP array analysis identified 22–151 (median 87.5) GCS per sample. The fraction of shared breakpoints and median overlap of GCS in samples from the same patient varied greatly, ranging from 6–83% (median 42%) and 0.24–0.93 (median 0.58), respectively.

Also the number of ESV (5-68) varied greatly between samples and the median number (26) was higher than for liposarcomas (7 in WDLS and 16 in MLS). The number of ESV steadily increased with time in all but one patient.

The findings in CXS were in good agreement with recent large scale sequencing data on adult sarcomas (TCGA, 2017). That study showed that the most common CXS subtype studied here, MFS, displays complex copy number profiles but few significant SNVs.

The CXS samples were highly heterogeneous, both with regard to rate and type of clonal evolution. For instance, the PT in case 20 shared no ESV with its LR, obtained 24.5 years later, but the GCS overlap was high (0.79). In contrast, LR1 of case 16 shared 12 of its 29 ESV with the LR6 (occurring 8 years later), but at the same time there were massive changes at the chromosome level resulting in a low GCS overlap (0.24). It is thus difficult to draw any firm conclusions on the

longitudinal clonal dynamics in these tumors without additional CXS cases, including other morphologic subtypes.

An obvious limitation of the present study is that the patients were selected on the basis of having late relapses, and we therefore cannot rule out the possibility that sarcomas displaying rapidly occurring relapses would have generated different results.

Conclusions

Article I

- There is an increased likelihood of finding novel gene fusions by RNA-seq in tumors displaying simple structural rearrangements.
- Karyotypes, in the absence of WGS data, are valuable when evaluating the significance of identified fusion transcripts.
- It is often necessary to run multiple gene fusion-detecting algorithms to obtain accurate results when analysing RNA-seq data from sarcomas.
- RNA-seq can provide vital information for differential diagnosis.

Articles II and III

- A small subset of UPS harbours gene fusions involving the transcription factor PRDM10, either as a *MED12-PRDM10* or a *CITED2-PRDM10* fusion.
- This subset of UPS appears to be less aggressive than classical high-grade UPS.
- PRDM10 positive tumors are genetically distinct from high-grade UPS as they exhibit a unique gene expression profile and few genomic alterations.
- Expression of the *CITED2-PRDM10* fusion in cell lines mimics the gene expression profile observed in tumors and causes significant change in chromatin accessibility.
- CADM3 constitutes a promising diagnostic marker at IHC.

Article IV

- The type and rate of clonal evolution vary considerably among sarcomas caused by different pathogenetic mechanisms.

- The data on WDLS demonstrate that high genetic variation at the single cell level does not necessarily translate into major changes in the predominant tumor clone.
- ESV have little or no significance in WDLS development and these tumors might relatively early reach a genetic fitness maximum.
- MLS are genetically relatively stable with a slow clonal evolution at the chromosome level, even in metastatic lesions.
- CXS are highly heterogeneous, both with regard to rate and type of clonal evolution.

Svensk sammanfattning

Den mänskliga kroppen består av miljarder av celler som växer, delar sig och dör enligt en ordnad mall. Nya celler skapas genom celldelning då en redan existerande cell kopierar sin arvsmassa och delar sig i två. Kopieringen av arvsmassan är dock in helt felfri och förändringar i DNA, så kallade mutationer, introduceras vid varje celldelning. Då det är arvsmassan som styr cellens funktioner kan förändringar i en cell resultera i att den upphör att följa den ordnade mallen. En sådan cell, med okontrollerad celldelning, kan konkurrera ut de normala cellerna och expandera i antal. Detta resulterar i att det formas en onormal vävnad som kallas tumör.

Tumörer som uppstår i mesenkymal vävnad kallas för mjukdelstumörer (MDT) och de maligna varianterna, med förmågan att invadera och sprida sig till andra delar av kroppen, kallas för sarkom. Då det finns fler än 100 olika subtyper av MDT, som delvis överlappar morfologiskt, kan diagnostiken vara problematisk. Information om de genetiska avvikelserna i de olika tumörerna kan underlätta vid diagnostisering och i vissa fall även utnyttjas som terapeutiska måltavlor. De genetiska förändringar som ligger till grund för tumörutvecklingen i sarkom är dock relativt dåligt utvärderade.

Målet med denna avhandling var att, med hjälp av djupsekvensering, i större utsträckning kunna studera och identifiera några av dessa genetiska förändringar, framförallt så kallade genfusioner.

I artikel I undersökte vi i vilken utsträckning unika strukturella rearrangemang ger upphov till genfusioner i sarkom med enkla cytogenetiska avvikelser. Med hjälp av RNA-sekvensering och cytogenetisk analys av 9 tumörer från 8 patienter lyckades vi identifiera 5 olika fusioner, varav 3 aldrig tidigare beskrivits.

I artikel II och III studerades en av de vanligaste undergrupperna av sarkom, odifferentierade pleomorfa sarkom (OPS). Med hjälp av RNA-sekvensering hittades två tidigare obeskrivna genfusioner, *MED12-PRDM10* och *CITED2-PRDM10*. OPS är vanligtvis en aggressiv tumörtyp men inga av fallen med fusion visade tecken på metastas och fusionerna tycks således indikera god prognos. Analys av både DNA och RNA visade att de PRDM10 positiva tumörerna (PPT) även genetiskt skiljer sig kraftigt åt från OPS då de uppvisar ett unikt genuttrycksmönster och få förändringar på DNA nivå. Genom att jämföra genuttrycket för PPT med flera morfologiskt lika tumörtyper identifierade vi genen *CADM3* som en potentiell

markör för förbättrad diagnostik. Vi visade även att markören var användbar på proteinnivå med hjälp av immunhistokemi.

I artikel IV undersöktes klonal evolution i sarkom som uppstår genom olika genetiska mekanismer. Vi analyserade dynamiken mellan kromosom- och nukleotidförändringar i tre typer av sarkom; amplikondrivna WDLS, genfusionsdrivna MLS och sarkom med komplexa genom. Vi visade att typen och hastigheten av klonal utveckling skiljer sig kraftigt åt bland sarkom som uppstår genom olika genetiska mekanismer.

Acknowledgements

Well, I obviously have not done all of this myself. This work would not have been possible without the help and support I have received throughout the years. Here I would like to, or at least try to, acknowledge everyone who has, in one way or another, been involved in this process.

First and foremost, my supervisor **Figge.** Not only Sweden's best cancer researcher but also its best supervisor. The latter award will not result in a televised gala, but something tells me this is preferable. I am very thankful for all the freedom you have given me, allowing me to work independently while always being just a few loud steps away. For all your support and knowledge and for always having one or a few new projects up your sleeve.

My co-supervisor **Karolin** for your energetic support and optimism. Where you get this energy from remains a mystery to me. It cannot possibly come from that brown water you insist on calling coffee?

To my fantastic group. **Elsa** for our effortless teamwork, stalking sessions and all the fun we have had during our exotic conference trips. The all-knowing entity "**Jenny** and **Linda"** for all the tremendous help and for teaching me everything I know about lab work. Though you more than willingly have offered me your help, it has often been accompanied with a side of verbal abuse and even if I most certainly blame you for my rapidly increasing number of grey hairs, I have to admit that messing with you have been the absolute highlight of most of my work days. **Nisse** for your constant good mood and for, after quite a few years of research, still being so enthusiastic and curious. I would also like to thank and wish viel glück to the newcomers, **Jan** and **Saskia.**

Felix and the PIs, **Anna**, **Bertil**, **David**, **Kajsa** (COYS), **Marcus** and **Thoas** for creating and maintaining a successful department with a friendly, relaxed and supportive work environment.

Anette for your incredible administrative powers. I just keep handing you random papers and things tend to work out. I like it.

All the past and present PhDs, **Anders, Charles, Pablo, Larissa, Ram, Naveen, Maria** and **Hiroaki** for a creative and warm atmosphere filled with fun social activities, both in and outside the lab. **Niklas** for being the only one cool enough to play golf, **Axel** for introducing me to the borderline lethal combination of grappa and svartklubb, my personal trainer **Palfy** for helping me defeat my nemesis, **Mia** for simply being out standing in the field of joyful conversations, **Mattias** for all the passionate discussions about bioinformatics and technical discussion about football. Also the great roommates I have had throughout the years, **Sette** for the comical amount of conversation, **Lindsay** for hooking me up with my sweet crib, **Karim** for always being down for a beer at inferno and **Giuseppe** for bringing the inferno to the room.

My lab partner **Marianne** for all your great advice (including how to clean a wound in the most painful way possible) and for inspiring me to learn Photoshop. I usually only head to the lab for two reasons, to strike up another conversation with you or to heckle Linda. Both are equally rewarding.

To all my other amazing **co-workers** for happily enjoying all the cakes I brought, for the unforgettable fika and lunch conversations, and for making C13 the friendliest of work places. The technical staff for keeping this place up and running, and for the occasional cheerful email. **Henrik** for always taking the time to answer any questions. **Maria** (thought you are no longer with us do to legal reasons) for your snappy comments. Our friendly neighbours, the protein people, that I after initial scepticism have learned to like and also **Ton** that I liked from the beginning.

Jag skulle även vilja tacka hela tjocka släkten**,** framförallt **morsan** och **farsan** för att ni alltid har ställt upp i alla lägen. Ert konstanta stöd har betytt mycket för mig och har format mig till den bortskämda människa jag är idag. Även ett stort tack till mina **bröder** för allt skjutsande och för att ni gjort platsen som favoritson så lättillgänglig.

Alla mina härliga vänner för allt kul som händer utanför labbet. Framförallt, **Markus**, **David** och **Jonas** för all shiit vi hittar på, utan er hade jag inte fått uppleva historiska ögonblick såsom gången Jonas bar väst eller sommarens minst sagt problematiska pool-party. **Wieg** för alla kvällar vi, utan större framgång, har stormat längs kanten. **Juffe** och **Emilia** för alla episka resor, Sean Bean kvällar, cinco de mayos och för att ni äntligen har tagit ert förnuft till fånga och flyttat hem till Skåne.

I would also like to sincerely thank my opponent **Ola Myklebost** and the examination board **Magnus Hansson, Göran Jönsson** and **Mikael Eriksson** for reviewing my thesis and **Catarina Lundin** for agreeing to chair my defense.

Finally, I would like to thank **everyone** who is attending my defense. If you feel like you have completely lost track of what is going on, feel free to enjoy this sudoku.

1					5			4
	2	3	4		6		5	
							6	
9		4					7	
			5	6	7			
	5					8		3
	4							
	3		7		1	9	8	
2			8					7

References

Agaram, NP, Chen, CL, Zhang, L, LaQuaglia, MP, Wexler, L, Antonescu, CR. (2014). Recurrent MYOD1 mutations in pediatric and adult sclerosing and spindle cell rhabdomyosarcomas: evidence for a common pathogenesis. *Genes Chromosomes Cancer* 53:779-787.

Alexandrov, LB, Stratton, MR. (2014). Mutational signatures: the patterns of somatic mutations hidden in cancer genomes. *Curr Opin Genet Dev* 24:52-60.

Antonescu, CR, Sung, YS, Chen, CL, Zhang, L, Chen, HW, Singer, S, et al. (2014). 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* 53:183-193.

Bailey, MH, Tokheim, C, Porta-Pardo, E, Sengupta, S, Bertrand, D, Weerasinghe, A, et al. (2018). Comprehensive characterization of cancer driver genes and mutations. *Cell* 173:371-385.

Barretina, J, Taylor, BS, Banerji, S, Ramos, AH, Lagos-Quintana, M, DeCarolis, PL, et al. (2010). Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nat Genet* 42:715-721.

Boveri, T. (1914). Zur Frage der Entstehung maligner Tumoren. (Gustav Fisher Verlag, Jena).

Buenrostro, JD, Wu, B, Chang, HY, Greenleaf, WJ. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* 109:21.29.21-29.

Burrell, RA, McGranahan, N, Bartek, J, Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501:338-345.

Cahill, DP, Kinzler, KW, Vogelstein, B, Lengauer, C. (1999). Genetic instability and darwinian selection in tumours. *Trends in Cell Biol* 9:M57-M60.

Carrara, M, Beccuti, M, Lazzarato, F, Cavallo, F, Cordero, F, Donatelli, S, et al. (2013). State-of-the-art fusion-finder algorithms sensitivity and specificity. *Biomed Res Int* 2013:340620.

Chen, T, Dent, SY. (2014). Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet* 15:93-106.

Chibon, F, Lagarde, P, Salas, S, Perot, G, Brouste, V, Tirode, F, et al. (2010). Validated prediction of clinical outcome in sarcomas and multiple types of

cancer on the basis of a gene expression signature related to genome complexity. *Nat Med* 16:781-781.

Cibulskis, K, Lawrence, MS, Carter, SL, Sivachenko, A, Jaffe, D, Sougnez, C, et al. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31:213-219.

Delattre, O, Zucman, J, Plougastel, B, Desmaze, C, Melot, T, Peter, M, et al. (1992). Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* 359:162-165.

DePristo, MA, Banks, E, Poplin, R, Garimella, KV, Maguire, JR, Hartl, C, et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491-498.

Dobin, A, Davis, CA, Schlesinger, F, Drenkow, J, Zaleski, C, Jha, S, et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15-21.

Dufresne, A, Brahmi, M, Karanian, M, Blay, JY. (2018). Using biology to guide the treatment of sarcomas and aggressive connective-tissue tumours. *Nat Rev Clin Oncol* 15:443-458.

Feng, FY, Brenner, JC, Hussain, M, Chinnaiyan, AM. (2014). Molecular pathways: targeting ETS gene fusions in cancer. *Clin Cancer Res* 20:4442-4448.

Fletcher, CDM, Gustafson, P, Rydholm, A, Willen, H, Åkerman, M. (2001). Clinicopathologic re-evaluation of 100 malignant fibrous histiocytomas: prognostic relevance of subclassification. *J Clin Oncol* 19:3045-3050.

Fletcher, CDM, Bridge, JA, Hogendoorn, PCW, Mertens, F. (2013a). WHO Classification of Tumours of Soft Tissue and Bone 4th ed. (IARC, Lyon).

Fletcher, CDM, Chibon, F, Mertens, F. (2013b). Undifferentiated/unclassified sarcomas. In: Fletcher, CDM, Bridge, JA, Hogendoorn, PCW, Mertens, F (eds). WHO Classification of Tumours of Soft Tissue and Bone 4th ed. (IARC, Lyon). p 236-238.

Folpe, AL, Hornick, JL, Mertens, F. (2013). 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, Lyon). p 95-96.

Futreal, PA, Coin, L, Marshall, M, Down, T, Hubbard, T, Wooster, R, et al. (2004). A census of human cancer genes. *Nat Rev Cancer* 4:177-183.

Gebhart, E. (2008). Ring chromosomes in human neoplasias. *Cytogenet Genome Res* 121:149-173.

Gebre-Medhin, S, Nord, KH, Möller, E, Mandahl, N, Magnusson, L, Nilsson, J, et al. (2012). Recurrent rearrangement of the PHF1 gene in ossifying fibromyxoid tumors. *Am J Pathol* 181:1069-1077.

Gerlinger, M, Rowan, AJ, Horswell, S, Math, M, Larkin, J, Endesfelder, D, et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* 366:883-892.

Gibault, L, Perot, G, Chibon, F, Bonnin, S, Lagarde, P, Terrier, P, et al. (2011). New insights in sarcoma oncogenesis: a comprehensive analysis of a large series

of 160 soft tissue sarcomas with complex genomics. *J Pathol* 223:64-71.
Gisselsson, D, Höglund, M, Mertens, F, Johansson, B, Dal Cin, P, Van den Berghe, H, et al. (1999). The structure and dynamics of ring chromosomes in human neoplastic and non-neoplastic cells. *Hum Genet* 104:315-325.
Gisselsson, D, Pettersson, L, Höglund, M, Heidenblad, M, Gorunova, L, Wiegant, J, et al. (2000). Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc Natl Acad Sci USA* 97:5357-5362.
Goldblum, JR. (2014). An approach to pleomorphic sarcomas: can we subclassify, and does it matter? *Mod Pathol* 27:39-46.
Goldblum, JR, Folpe, AL, Weiss, SW. (2014). General considerations. In: Goldblum, JR, Folpe, AL, Weiss, SW (eds). Ezinger and Weiss's Soft Tissue Tumors 6th ed. (Mosby, New York). p 1-9.
Goodwin, S, McPherson, JD, McCombie, WR. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet* 17:333-351.
Guled, M, Pazzaglia, L, Borze, I, Mosakhani, N, Novello, C, Benassi, MS, et al. (2014). Differentiating soft tissue leiomyosarcoma and undifferentiated pleomorphic sarcoma: A miRNA analysis. *Genes Chromosomes Cancer* 53:693-702.
Gustafson, P. (1994). Soft tissue sarcoma. Epidemiology and prognosis in 508 patients. *Acta Orthop Scand Suppl* 259:1-31.
Haas, B, Dobin, A, Stransky, N, Li, B, Yang, X, Tickle, T, et al. (2017). STAR-Fusion: fast and accurate fusion transcript detection from RNA-seq. *bioRxiv*.
Haldar, M, Hancock, JD, Coffin, CM, Lessnick, SL, Capecchi, MR. (2007). A conditional mouse model of synovial sarcoma: insights into a myogenic origin. *Cancer Cell* 11:375-388.
Hallor, KH, Sciot, R, Staaf, J, Heidenblad, M, Rydholm, A, Bauer, HC, et al. (2009). Two genetic pathways, t(1;10) and amplification of 3p11-12, in myxoinflammatory fibroblastic sarcoma, haemosiderotic fibrolipomatous tumour, and morphologically similar lesions. *J Pathol* 217:716-727.
Hanahan, D, Weinberg, RA. (2000). The hallmarks of cancer. *Cell* 100:57-70.
Hanahan, D, Weinberg, RA. (2011). Hallmarks of cancer: the next generation. *Cell* 144:646-674.
Hehir-Kwa, JY, Pfundt, R, Veltman, JA. (2015). Exome sequencing and whole genome sequencing for the detection of copy number variation. *Expert Rev Mol Diagn* 15:1023-1032.
Heim, S, Mitelman, F. (2015). Cancer Cytogenetics 4th ed. (Wiley Blackwell, Oxford).
Heinrich, MC, Corless, CL, Duensing, A, McGreevey, L, Chen, CJ, Joseph, N, et al. (2003). PDGFRA activating mutations in gastrointestinal stromal tumors. *Science* 299:708-710.
Helias-Rodzewicz, Z, Perot, G, Chibon, F, Ferreira, C, Lagarde, P, Terrier, P, et al. (2010). YAP1 and VGLL3, encoding two cofactors of TEAD transcription

factors, are amplified and overexpressed in a subset of soft tissue sarcomas. *Genes Chromosomes Cancer* 49:1161-1171.

Hibbard, MK, Kozakewich, HP, Dal Cin, P, Sciot, R, Tan, XL, Xiao, S, et al. (2000). PLAG1 fusion oncogenes in lipoblastoma. *Cancer Res* 60:4869-4872.

Højfeldt, JW, Agger, K, Helin, K. (2013). Histone lysine demethylases as targets for anticancer therapy. *Nat Rev Drug Discov* 12:917-930.

Italiano, A, Bianchini, L, Gjernes, E, Keslair, F, Ranchere-Vince, D, Dumollard, J-M, et al. (2009). Clinical and biological significance of CDK4 amplification in well-differentiated and dedifferentiated liposarcomas. *Clin Cancer Res* 15:5696-5703.

Iyer, MK, Chinnaiyan, AM, Maher, CA. (2011). ChimeraScan: a tool for identifying chimeric transcription in sequencing data. *Bioinformatics* 27:2903-2904.

Jia, W, Qiu, K, He, M, Song, P, Zhou, Q, Zhou, F, et al. (2013). SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol* 14:R12.

Joseph, CG, Hwang, H, Jiao, Y, Wood, LD, Kinde, I, Wu, J, et al. (2014). Exomic analysis of myxoid liposarcomas, synovial sarcomas, and osteosarcomas. *Genes Chromosomes Cancer* 53:15-24.

Kanojia, D, Nagata, Y, Garg, M, Lee, DH, Sato, A, Yoshida, K, et al. (2015). Genomic landscape of liposarcoma. *Oncotarget* 6:42429-42444.

Killela, PJ, Reitman, ZJ, Jiao, Y, Bettegowda, C, Agrawal, N, Diaz, LA, Jr., et al. (2013). TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc Natl Acad Sci USA* 110:6021-6026.

Kilpatrick, SE, Doyon, J, Choong, PF, Sim, FH, Nascimento, AG. (1996). The clinicopathologic spectrum of myxoid and round cell liposarcoma. A study of 95 cases. *Cancer* 77:1450-1458.

Kim, D, Pertea, G, Trapnell, C, Pimentel, H, Kelley, R, Salzberg, SL. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 14:R36.

Koelsche, C, Renner, M, Hartmann, W, Brandt, R, Lehner, B, Waldburger, N, et al. (2014). TERT promoter hotspot mutations are recurrent in myxoid liposarcomas but rare in other soft tissue sarcoma entities. *J Exp Clin Cancer Res* 33:33.

Kumar, S, Vo, AD, Qin, F, Li, H. (2016). Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Sci Rep* 6:21597.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* Prepr. arXiv 0,3.

Lloyd, MC, Cunningham, JJ, Bui, MM, Gillies, RJ, Brown, JS, Gatenby, RA. (2016). Darwinian dynamics of intratumoral heterogeneity: not solely random mutations but also variable environmental selection forces. *Cancer Res* 76:3136-3144.

Loeb, LA. (1991). Mutator phenotype may be required for multistage

carcinogenesis. *Cancer Res* 51:3075-3079.
Loeb, LA. (2016). Human cancers express a mutator phenotype: hypothesis, origin, and consequences. *Cancer Res* 76:2057-2059.
Lovly, CM, Gupta, A, Lipson, D, Otto, G, Brennan, T, Chung, CT, et al. (2014). Inflammatory myofibroblastic tumors harbor multiple potentially actionable kinase fusions. *Cancer Discov* 4:889-895.
Makohon-Moore, AP, Zhang, M, Reiter, JG, Bozic, I, Allen, B, Kundu, D, et al. (2017). Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat Genet* 49:358-366.
Mandahl, N, Mertens, F, Aman, P, Rydholm, A, Brosjo, O, Willen, H, et al. (1994). Nonrandom secondary chromosome-aberrations in liposarcomas with t(12, 16). *Int J Oncol* 4:307-310.
Mantripragada, KK, Diaz de Stahl, T, Patridge, C, Menzel, U, Andersson, R, Chuzhanova, N, et al. (2009). Genome-wide high-resolution analysis of DNA copy number alterations in NF1-associated malignant peripheral nerve sheath tumors using 32K BAC array. *Genes Chromosomes Cancer* 48:897-907.
McBride, MJ, Kadoch, C. (2018). Disruption of mammalian SWI/SNF and polycomb complexes in human sarcomas: mechanisms and therapeutic opportunities. *J Pathol* 244:638-649.
McKenna, A, Hanna, M, Banks, E, Sivachenko, A, Cibulskis, K, Kernytsky, A, et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297-1303.
McLaren, W, Gil, L, Hunt, SE, Riat, HS, Ritchie, GRS, Thormann, A, et al. (2016). The Ensembl variant effect predictor. *Genome Biol* 17:122.
Mertens, F, Antonescu, CR, Mitelman, F. (2016). Gene fusions in soft tissue tumors: Recurrent and overlapping pathogenetic themes. *Genes Chromosomes Cancer* 55:291-310.
Mertens, F, Johansson, B, Fioretos, T, Mitelman, F. (2015). The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* 15:371-381.
Mertens, F, Tayebwa, J. (2013). Evolving techniques for gene fusion detection in soft tissue tumours. *Histopathology* 64:151-162.
Mitelman, F, Johansson, B, Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 7:233-245.
Mitelman, F, Johansson, B, Mertens, F, eds. (2018). *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer*. http://cgap.nci.nih.gov/Chromosomes/Mitelman.
Nicorici, D, Satalan, M, Edgren, H, Kangaspeska, S, Murumagi, A, Kallioniemi, O, et al. (2014). FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. *bioRxiv*.
Nowak, MA, Komarova, NL, Sengupta, A, Jallepalli, PV, Shih, IM, Vogelstein, B,

et al. (2002). The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci USA* 99:16226-16231.

Nowell, PC. (1976). Clonal evolution of tumor-cell populations. *Science* 194:23-28.

Oda, Y, Yamamoto, H, Takahira, T, Kobayashi, C, Kawaguchi, K, Tateishi, N, et al. (2005). Frequent alteration of p16(INK4a)/p14(ARF) and p53 pathways in the round cell component of myxoid/round cell liposarcoma: p53 gene alterations and reduced p14(ARF) expression both correlate with poor prognosis. *J Pathol* 207:410-421.

Ozsolak, F, Milos, PM. (2011). RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* 12:87-98.

Parker, BC, Engels, M, Annala, M, Zhang, W. (2014). Emergence of FGFR family gene fusions as therapeutic targets in a wide spectrum of solid tumours. *J Pathol* 232:4-15.

Pedeutour, F, Coindre, JM, Nicolo, G, Bouchot, C, Ayraud, N, Carel, CT. (1993). Ring chromosomes in dermatofibrosarcoma protuberans contain chromosome 17 sequences: fluorescence in situ hybridization. *Cancer Genet Cytogenet* 67:149.

Pedeutour, F, Simon, MP, Minoletti, F, Sozzi, G, Pierotti, MA, Hecht, F, et al. (1995). Ring 22 chromosomes in dermatofibrosarcoma protuberans are low-level amplifiers of chromosome 17 and 22 sequences. *Cancer Res* 55:2400-2403.

Puls, F, Pillay, N, Fagman, H, Palin, A, Rissler, P, McCulloch, T, et al. (2018). Soft tissue tumors with PRDM10 gene fusions: a clinicopathologic study of nine cases. *Am J Surg Pathol*, in press.

Reinert, K, Langmead, B, Weese, D, Evers, DJ. (2015). Alignment of next-generation sequencing reads. *Annu Rev Genomics Hum Genet* 16:133-151.

Reiter, JG, Makohon-Moore, AP, Gerold, JM, Bozic, I, Chatterjee, K, Iacobuzio-Donahue, CA, et al. (2017). Reconstructing metastatic seeding patterns of human cancers. *Nat Commun* 8:14114.

Riggi, N, Cironi, L, Provero, P, Suva, ML, Stehle, JC, Baumer, K, et al. (2006). Expression of the FUS-CHOP fusion protein in primary mesenchymal progenitor cells gives rise to a model of myxoid liposarcoma. *Cancer Res* 66:7016-7023.

Riggi, N, Cironi, L, Suva, ML, Stamenkovic, I. (2007). Sarcomas: genetics, signalling, and cellular origins. Part I: The fellowship of TET. *J Pathol* 213:4-20.

Rydholm, A. (1983). Management of patients with soft-tissue tumors. Strategy developed at a regional oncology center. *Acta Orthop Scand Suppl* 203:13-77.

Santarius, T, Shipley, J, Brewer, D, Stratton, MR, Cooper, CS. (2010). A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* 10:59-64.

Saunders, CT, Wong, WS, Swamy, S, Becq, J, Murray, LJ, Cheetham, RK. (2012).

Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28:1811-1817.

Shah, SP, Morin, RD, Khattra, J, Prentice, L, Pugh, T, Burleigh, A, et al. (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* 461:809-867.

Shaw, AT, Hsu, PP, Awad, MM, Engelman, JA. (2013). Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat Rev Cancer* 13:772-787.

Shern, JF, Chen, L, Chmielecki, J, Wei, JS, Patidar, R, Rosenberg, M, et al. (2014). Comprehensive genomic analysis of rhabdomyosarcoma reveals a landscape of alterations affecting a common genetic axis in fusion-positive and fusion-negative tumors. *Cancer Discov* 4:216-231.

Straessler, KM, Jones, KB, Hu, H, Jin, H, van de Rijn, M, Capecchi, MR. (2013). Modeling clear cell sarcomagenesis in the mouse: cell of origin differentiation state impacts tumor characteristics. *Cancer Cell* 23:215-227.

Stransky, N, Cerami, E, Schalm, S, Kim, JL, Lengauer, C. (2014). The landscape of kinase fusions in cancer. *Nat Commun* 5:4846.

Szuhai, K, de Jong, D, Leung, WY, Fletcher, CD, Hogendoorn, PC. (2014). Transactivating mutation of the MYOD1 gene is a frequent event in adult spindle cell rhabdomyosarcoma. *J Pathol* 232:300-307.

TCGA. (2017). Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* 171:950-965.

The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature* 526:68-74.

Trask, BJ. (1991). Fluorescence in situ hybridization: applications in cytogenetics and gene mapping. *Trends Genet* 7:149-154.

Vogelstein, B, Papadopoulos, N, Velculescu, VE, Zhou, S, Diaz, LA, Jr., Kinzler, KW. (2013). Cancer genome landscapes. *Science* 339:1546-1558.

Wang, M, Zhao, J, Zhang, L, Wei, F, Lian, Y, Wu, Y, et al. (2017). Role of tumor microenvironment in tumorigenesis. *J Cancer* 8:761-773.

Wardelmann, E, Hrychyk, A, Merkelbach-Bruse, S, Pauls, K, Goldstein, J, Hohenberger, P, et al. (2004). Association of platelet-derived growth factor receptor alpha mutations with gastric primary site and epithelioid or mixed cell morphology in gastrointestinal stromal tumors. *J Mol Diagn* 6:197-204.

Weiss, SW, Goldblum, JR. (2008). Benign lipomatous tumors. In: Weiss, SW, Goldblum, JR (eds). Ezinger and Weiss's Soft Tissue Tumors 5th ed. (Mosby, New York). p 452-455.

Wingender, E, Schoeps, T, Haubrock, M, Donitz, J. (2015). TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res* 43:97-102.

Yoshihara, K, Wang, Q, Torres-Garcia, W, Zheng, S, Vegesna, R, Kim, H, et al. (2015). The landscape and therapeutic relevance of cancer-associated transcript fusions. *Oncogene* 34:4845-4854.

Article I

Laboratory Investigation (2015) 95, 603–609

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

Jakob Hofvander[1], Johnbosco Tayebwa[1], Jenny Nilsson[1], Linda Magnusson[1], Otte Brosjö[2], Olle Larsson[3], Fredrik Vult von Steyern[4], Henryk A Domanski[5], Nils Mandahl[1] and Fredrik Mertens[1]

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

[1]Department of Clinical Genetics, University and Regional Laboratories, Lund University, Lund, Sweden; [2]Department of Orthopedics, Karolinska University Hospital, Solna, Sweden; [3]Department of Pathology, Karolinska University Hospital, Solna, Sweden; [4]Department of Orthopedics, Skåne University Hospital, Lund University, Lund, Sweden and [5]Department 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 described[14] 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 Superscript 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 ddH_2O 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 *C1QTNF6-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	*AFF3-PHF1*	T	Out of frame	*AFF3* exon 11 fused with *PHF1* exon 13
3	*CCBL1-ARL*	T, C, S	Out of frame	*CCBL1* exon 1 fused with *ARL1* exon 5
3	*KIAA2026-NUDT11*	T, C, S	In frame	*KIAA2026* exon 1 fused with *NUDT11* exon 2
4	*FUS-CREB3L2*	C	In frame	*FUS* exon 6 fused with *CREB3L2* exon 6
7	*NOTCH3-?*	C		*NOTCH3* exon 27 fused with inverted *NOTCH3* exon 33 and 7 unidentified bp
8	*HAS2-PLAG1*	C	In frame	*HAS2* exon 1 fused with *PLAG1* exon 3

[a]T = TopHat; C = ChimeraScan; S=SOAPfuse.

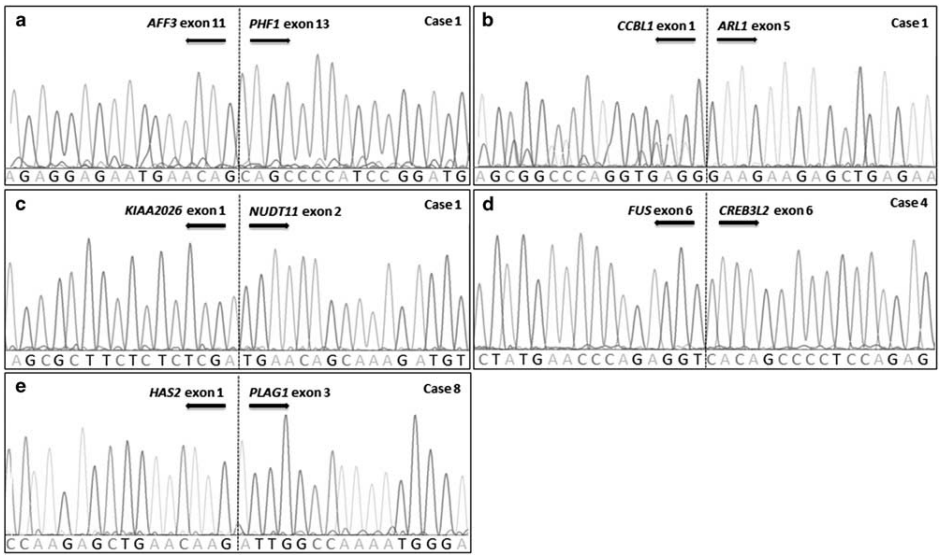


Figure 1 (**a**–**e**): Partial chromatograms of RT-PCR-amplified fusion transcripts from chimeric genes identified by RNA-Seq.

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 $\geq$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 $\geq$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 *EWSR1-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 bioinformatical 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.

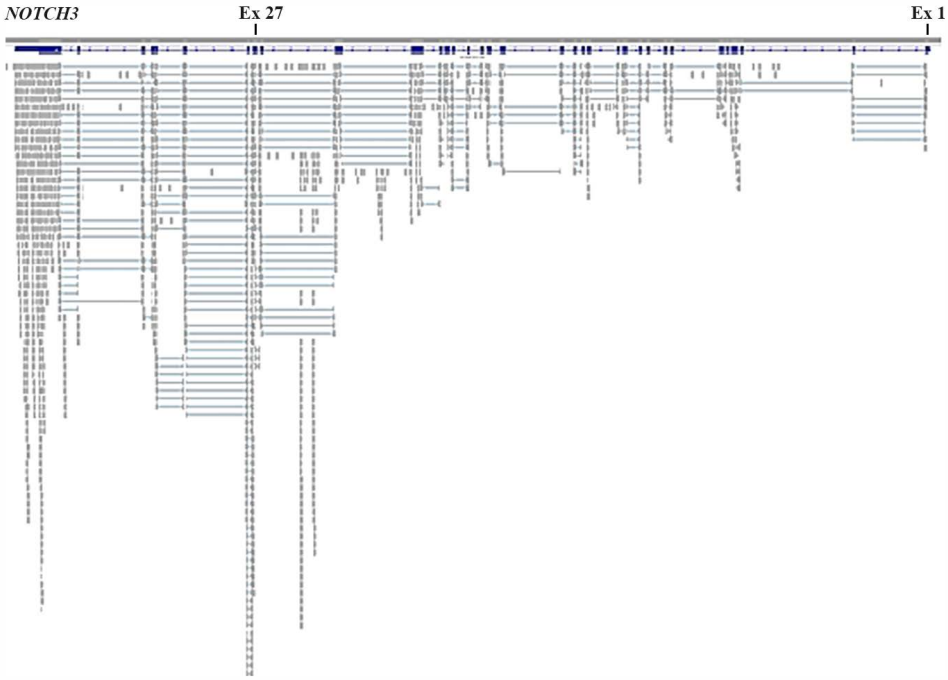
1. Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F. WHO Classification of Tumours of Soft Tissue and Bone. IARC Press: Lyon, 2013.
2. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 2007;7:233–245.
3. Welch JS, Ley TJ, Link DC *et 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, Pertea 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 haemangioendothelioma. 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.

Supplementary information for:

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

Hofvander J, Tayebwa J, Nilsson J, Magnusson L, Brosjö O, Larsson O, Vult von Steyern F, Domanski HA, Mandahl N, Mertens F. *Lab Invest*. 2015;95:603-609.



Supplementary Figure 1. Paired-end reads where one mate maps to the NOTCH3 gene and the other mate is unmapped. More reads with unmapped mates are present in the 3´part of the gene, exons 25-33, indicating that a rearrangement resulting in a split of the gene has occurred.

Supplementary Table S1

Unfiltered TopHat-Fusion output file . TopHat version 2.0.7 was run with the following parameters, fusion minimal distance of 100,000 bp (fusion-min-dist), mate inner distance (-R) of 200, standard deviation (mate-std-dev) of 200 and a fusion anchor length of 13. The output files were further filtered by running TopHat-fusion-post demanding at least one fusion-spanning read and two fusion-spanning mate pairs. The GRCh37/hg19 build was used as the reference genome.

chr 5´partner	Breakpoint 5´gene	chr 3´partner	Breakpoint 3´gene	Gene 5´partner	Gene 3´partner	Spanning reads	Spanning mate pairs	Spanning mate pairs where one end spans a fusion
Sample 1a								
-								
Sample 1b								
chr17	33954691	chrX	92479095	AP2B1	ENSG00000234130	1	4	0
chr4	16428062	chr4	16538252	ENSG00000248138	LDB2	2	3	0
chr3	61728590	chr5	86180670	PTPRG	ENSG00000242477	1	33	0
chr11	87883122	chr11	88033697	RAB38	CTSC	37	18	30
Sample 2								
chr6	74002062	chr8	109252304	C6orf147	EIF3E	1	9	1
chr2	**23657751**	**chr17**	**8045586**	**KLHL29**	**PER1**	**8**	**5**	**7**
chr2	**23658097**	**chr17**	**8045541**	**KLHL29**	**PER1**	**7**	**5**	**4**
Sample 3								
chr2	**100286034**	**chr6**	**33383092**	**AFF3**	**PHF1**	**33**	**5**	**31**
chr9	72089178	chr12	109094899	APBA1	CORO1C	11	3	12
chr9	72089181	chr12	109094899	APBA1	CORO1C	2	3	4
chr9	100043950	chr12	52599962	BDAG1	LOC283403	10	7	10
chr9	**131644175**	**chr12**	**101790354**	**CCBL1**	**ARL1**	**17**	**3**	**21**
chr12	**91572118**	**chr12**	**111612720**	**DCN**	**CUX2**	**2**	**10**	**2**
chr12	**91572118**	**chr12**	**111620519**	**DCN**	**CUX2**	**21**	**10**	**7**
chr9	19316835	chr9	97430034	DENND4C	ENSG00000204343	2	2	3
chr12	**96617550**	**chr12**	**107262492**	**ELK3**	**RIC8B**	**5**	**3**	**3**
chr9	**35738628**	**chr9**	**131013218**	**GBA2**	**DNM1**	**13**	**3**	**16**
chr9	**6007194**	**chrX**	**51234600**	**KIAA2026**	**NUDT11**	**4**	**2**	**4**
Sample 4								
chr1	166135290	chr1	166304564	FAM78B	ENSG00000229588	6	2	4
chr11	131781541	chr11	132527213	NTM	OPCML	11	2	14
Sample 5								
chr22	22712187	chr22	23237554	ENSG00000211648	IGLL5	3	30	1
chr22	22735293	chr22	23237554	ENSG00000211651	IGLL5	1	28	0
chr22	22735711	chr22	23235959	ENSG00000211651	IGLL5	25	28	18
chr22	22786798	chr22	23235959	ENSG00000211655	IGLL5	16	9	16
chr22	23101694	chr22	23235959	ENSG00000211666	IGLL5	22	114	25
Sample 6								
-								
Sample 7								
chr1	568918	chr15	45003712	ENSG00000240409	B2M	1	6	0
Sample 8								
-								

Supplementary Table S2. Unfiltered ChimerScan output file . The default setting was used to run ChimeraScan and the GRCh37/hg19 build was used as the reference genome.

Table too large for printing

Supplementary Table S3

Unfiltered output from SOAPfuse. SOAPfuse version 1.26 was run with a fusion minimal distance of 1000 bp, fusion anchor length of 10 bp and the minimum sum number of junction reads and spanning reads was 2. The GRCh37/hg19 build was used as the reference genome.

chr 5´partner	Genomic 5´end	chr 3´partner	Genomic 3´start	Gene 5´partner	Gene 3´partner	Spanning reads	Junction reads
Sample 1a							
-							
Sample 1b							
chr17	20771214	chr15	20876597	CCDC144NL	NBEAP1	7	2
chr17	20771218	chr15	20876597	CCDC144NL	NBEAP1	6	1
chr1	85028940	chr1	84967653	CTBS	GNG5	15	23
chr11	88029301	chr11	87883123	CTSC	RAB38	41	1
chr11	88033698	chr11	87883123	CTSC	RAB38	30	53
chr11	88045556	chr11	87883123	CTSC	RAB38	2	2
chr1	150778337	chr5	151049345	CTSK	SPARC	1	1
chr14	24761405	chr14	24740517	DHRS1	RABGGTA	5	23
chr14	24761644	chr14	24740517	DHRS1	RABGGTA	2	2
chr20	18768652	chr8	66955723	LINC00652	DNAJC5B	9	10
chr11	1908806	chr11	1944087	LSP1SOAPfuse2SOAPfuse	TNNT3	1	6
chr17	47304009	chr17	47286316	PHOSPHO1	GNGT2	8	10
chr1	153934696	chr1	153927642	SLC39A1	CRTC2	1	3
chrX	46405062	chrX	46457195	ZNF674-AS1	CHST7	1	2
Sample 2							
chr4	110663647	chr7	4837597	CFI	RADIL	5	2
chr1	85028940	chr1	84967653	CTBS	GNG5	2	2
chr9	124094955	chr16	16243276	GSN	ABCC6	2	2
chr19	8495751	chr19	8520289	MARCH2	HNRNPM	2	1
chr15	65153774	chr15	65218265	PLEKHO2	ANKDD1A	1	1
Sample 3							
chr12	90049452	chr12	123915202	ATP2B1	RILPL2	2	3
chr9	**131644176**	**chr12**	**101790355**	**CCBL1**	**ARL1**	**7**	**17**
chr1	85028940	chr1	84967653	CTBS	GNG5	3	7
chr14	24761405	chr14	24740517	DHRS1	RABGGTA	4	10
chr3	15604865	chr3	15531144	HACL1	COLQ	1	2
chr20	33095713	chr20	33114073	ITCH	DYNLRB1	2	2
chr9	**6007195**	**chrX**	**51234601**	**KIAA2026**	**NUDT11**	**4**	**4**
chr10	75010573	chr10	75016174	MRPS16	TTC18	27	6
chr10	75011521	chr10	75016174	MRPS16	TTC18	6	22
chr12	119419818	chr9	99246744	SRRM4	HABP4	2	3
Sample 4							
chr19	15507961	chr19	15487825	AKAP8L	AKAP8	2	2
chr5	40764616	chr5	40747121	PRKAA1	TTC33	1	3
chr21	45548507	chr21	45553984	PWP2	C21orf33	1	2
chr10	51387763	chr10	51732772	TIMM23B	LINC00843	4	13
Sample 5							
chr17	29286025	chr17	29311638	ADAP2	RNF135	4	4
chr15	40854180	chr7	26241365	C15orf57	CBX3	5	2
chr19	51870711	chr19	51857562	CLDND2	ETFB	4	8

chr11	88029301	chr11	87883123	CTSC	RAB38	2	1
chr20	18768652	chr8	66955723	LINC00652	DNAJC5B	5	5
chr4	680321	chr4	667190	MFSD7	ATP5I	1	2
chrX	108297380	chr5	149792410	RNA28S5	CD74	2	3
chr14	31119856	chr13	47224336	SCFD1	LRCH1	2	2
chr1	143916308	chr1	121107154	SRGAP2B	SRGAP2C	4	8
chr7	75625763	chr7	75621875	STYXL1	TMEM120A	2	7
chr10	51387763	chr10	51732772	TIMM23B	LINC00843	2	5
Sample 6							
chr22	40762502	chr22	40796700	ADSL	SGSM3	3	4
chr2	130887371	chr2	130948042	CCDC74B-AS1	MZT2B	4	3
chr11	117710496	chr11	117698823	FXYD6	FXYD2	2	1
chr17	25944454	chr17	25965289	KSR1	LGALS9	2	1
chr20	18548240	chr20	18574375	LINC00493	DTD1	4	4
chr5	138629494	chr5	138699448	MATR3	PAIP2	1	2
chr19	42365281	chr19	42383060	RPS19	CD79A	3	1
chr7	45808259	chr7	56088924	SEPT7P2	PSPH	3	2
chr6	44200165	chr6	44216367	SLC29A1	HSP90AB1	1	1
chr9	138852871	chr9	138719430	UBAC1	CAMSAP1	3	2
chr12	6574056	chr12	6557923	VAMP1	CD27-AS1	6	6
Sample 7							
chrX	108297589	chr1	163116933	RNA28S5	RGS5	2	2
Sample 8							
chr1	227128100	chr1	228284779	ADCK3	ARF1	1	2
chr17	65822453	chr17	66246329	BPTF	AMZ2	8	2
chr3	10157503	chr3	10188198	BRK1	VHL	3	1
chr11	2418194	chr11	2423524	CD81	TSSC4	4	8
chr11	870121	chr11	840480	CHID1	POLR2L	2	1
chr12	69633486	chr9	94173188	CPSF6	NFIL3	2	1
chr9	139757740	chr16	2757299	EDF1	KCTD5	2	1
chr17	37886516	chr17	37898505	ERBB2	GRB7	4	2
chr7	128498525	chr7	128505431	FLNC	ATP6V1F	24	2
chr1	36863368	chr1	36826941	LSM10	STK40	1	1
chr5	138629494	chr5	138699448	MATR3	PAIP2	1	2
chr18	47799194	chr18	47788589	MBD1	CCDC11	2	1
chr22	50968333	chr22	50962853	ODF3B	SCO2	10	24
chr19	44156377	chr19	44131942	PLAUR	CADM4	1	1
chr8	128806980	chr8	128750494	PVT1	MYC	2	1
chr11	117064679	chr11	117073718	SIDT2	TAGLN	3	2
chrX	153716018	chrX	153714670	SLC10A3	UBL4A	2	2
chr17	37818597	chr17	37821969	STARD3	TCAP	1	1
chr5	68665484	chr1	155183028	TAF9	MTX1	1	1
chr5	68665484	chr1	155203655	TAF9	MTX1P1	1	1
chr10	51387763	chr10	51732772	TIMM23B	LINC00843	13	15
chr22	50968333	chr22	50962853	TYMP	SCO2	10	26
chr4	1360219	chr4	1326545	UVSSA	MAEA	1	1
chr1	68603463	chr1	68513045	WLS	DIRAS3	1	3
chr7	99096339	chr7	99057816	ZNF394	ATP5J2	2	3

Supplementary Table S4.

Filtered versions of the Chimerascan (C), TopHat (T) and SOAPfuse (S) output files. Only reported fusions that were further investigated with RT-PCR are displayed and confimed fusions are highlighted in bold.

Location 5´gene	Breakpoint 5´gene	Location 3´gene	Breakpoint 3´gene	5´gene	3´gene	Algorithm
Case 1a						
22q13	37,581,993	19q13	46,808,501	*C1QTNF6*	*HIF3A*	C
Case 1b						
22q13	37,581,993	19q13	46,808,501	*C1QTNF6*	*HIF3A*	C
Case 2						
2p24	23,657,751	17p13	8,045,586	*KLHL29*	*PER1*	T
Case 3						
6p21	100,286,034	2q11	33,383,092	***PHF1***	***AFF3***	T
9q34	131,644,175	12q23	101,790,354	***CCBL1***	***ARL1***	T, C, S
9p24	6,007,194	Xp11	51,234,600	***KIAA2026***	***NUDT11***	T, C, S
9q34	131,013,218	9p13	35,736,863	*DNM1*	*GBA2*	C
12q21	91,572,118	12q24	111,612,720	*DCN*	*CUX2*	T
12q23	96,617,550	12q23	107,262,492	*ELK3*	*RIC8B*	T
Case 4						
16p11	31,196,499	7q34	137,559,726	***FUS***	***CREB3L2***	C
Case 5						
-						
Case 6						
-						
Case 7						
19p13	15,311,791	7q33	134,800,131	***NOTCH3***	***AGBL3***	C
Case 8						
8q24	122,653,629	8q12	57,073,468	***HAS2***	***PLAG1***	C

Supplementary Table S5. Primers Used for Gene Fusion Verification. Only primers that gave results are listed.

Case	Fusion	Primer	Sequence	Reference sequence
3	*CCBL1-ARL1*	CCBL1-116-F	ACAGGGACTGCTGCAACCTA	**NM_004059.4**
		ARL1-582-R	TCGGTCCTTCAAGGCAGGTA	**NM_001177.4**
3	*PHF1-AFF3*	AFF3-1176-F	GATGCAGAGCCAGAGAGTCC	**NM_002285.2**
		PHF1-1780-R	TACTGCACAGAGCCATCAGG	**NM_002636.4**
3	*KIAA2026-NUDT11*	KIAA2026-681-F	GAGTTCGTGGCGGACTTCA	**NM_001017969.2**
		NUDT11-779-R	GAGTGTCCCAAGATGCAGGAA	**NM_018159.3**
3	*KIAA2026-NUDT11*	KIAA2026-637-F	AGATGGAAGAGAAGTTCGCCA	**NM_001017969.2**
		NUDT11-812-R	CCAGAGCAAGAGTCAGTGGTAT	**NM_018159.3**
4	*FUS-CREB3L2*	FUS-582-F	CCAGTACAACAGCAGCAGTG	**NM_004960.3**
		CREB3L2-1467-R	TGAAGCTTCTGGAGTTGCTG	**NM_194071.3**
7	*NOTCH3*	Universal primer	-(Clontech)	
		NOTCH3-EX29-5319-R	GCAACCAGATGGTGTTGAGTCCACTGAC	**NM_000435.2**
8	*HAS2-PLAG1*	HAS2- 467-F	GTCGTCTCAAATTCATCTGATCTC	**XM_005250900.1**
		PLAG1-868-R	GTTCTTGCCACATTCTTCGC	**NM_002655.2**
8	*HAS2-PLAG1*	HAS2- 467-F	GTCGTCTCAAATTCATCTGATCTC	**XM_005250900.1**
		PLAG1- 461-R	TCTTGTTGGACACTTGGGAAC	**NM_002655.2**

Article II

Biology of Human Tumors

Clinical Cancer Research

Recurrent *PRDM10* Gene Fusions in Undifferentiated Pleomorphic Sarcoma

Jakob Hofvander[1], Johnbosco Tayebwa[1], Jenny Nilsson[1], Linda Magnusson[1], Otte Brosjö[2], Olle Larsson[3], Fredrik Vult von Steyern[4], Nils Mandahl[1], Christopher D.M. Fletcher[5], and Fredrik Mertens[1]

Abstract

Purpose: Undifferentiated pleomorphic sarcoma (UPS) is defined as a sarcoma with cellular pleomorphism and no identifiable line of differentiation. It is typically a high-grade lesion with a metastatic rate of about one third. No tumor-specific rearrangement has been identified, and genetic markers that could be used for treatment stratification are lacking. We performed transcriptome sequencing (RNA-Seq) to search for novel gene fusions.

Experimental design: RNA-Seq, FISH, and/or various PCR methodologies were used to search for gene fusions and rearrangements of the *PRDM10* gene in 84 soft tissue sarcomas.

Results: Using RNA-Seq, two cases of UPS were found to display novel gene fusions, both involving the transcription factor *PRDM10* as the 3′ partner and either *MED12* or *CITED2* as the 5′ partner gene. Further screening of 82 soft tissue sarcomas for rearrangements of the PRDM10 locus revealed one more UPS with a MED12/PRDM10 fusion. None of these genes has been implicated in neoplasia-associated gene fusions before.

Conclusions: Our results suggest that *PRDM10* fusions are present in around 5% of UPS. Although the fusion-positive cases in our series showed the same nuclear pleomorphism and lack of differentiation as other UPS, it is noteworthy that all three were morphologically low grade and that none of the patients developed metastases. Thus, *PRDM10* fusion-positive sarcomas may constitute a clinically important subset of UPS. *Clin Cancer Res; 21(4); 864–9. ©2014 AACR.*

Introduction

Undifferentiated sarcoma is defined as a sarcoma with no identifiable line of differentiation, excluding dedifferentiated types of specific sarcomas (1). Undifferentiated sarcomas, accounting for approximately 20% of all soft tissue sarcomas, may be further subdivided according to cellular shape (round cell, spindle cell, epithelioid, or pleomorphic). The pleomorphic variant (undifferentiated pleomorphic sarcoma, UPS) is particularly common among adults, and most frequently arises in the lower extremities (2). It is typically a high-grade lesion with a local recurrence rate ranging between 19% and 31%, a metastatic rate of 31% to 35%, and a five-year survival of 65% to 70% (3). UPS have a highly variable morphology, all sharing a marked pleomorphism often admixed with spindle cells and bizarre multinucleated giant cells. Treatment is based on the same strategy as for most other soft tissue sarcomas, that is, surgery with wide margins. Depending on surgical margins, location, and tumor-associated risk factors, adjuvant treatment, including radiotherapy and chemotherapy, is considered.

The genetic aspects of UPS are still poorly defined, partly due to shifting diagnostic criteria; although many sarcomas now diagnosed as UPS were previously classified as malignant fibrous histiocytoma (MFH), a substantial subset of MFH tumors was shown to constitute poorly differentiated forms of other sarcomas, such as leiomyosarcoma or liposarcoma (1). The karyotypes and copy-number profiles for UPS tend to be highly complex, with extensive intercellular variation, and a complete description of all chromosomal aberrations is rare (4–7). However, the level of cytogenetic complexity varies considerably, with a subset showing only a few structural and/or numerical aberrations. Still, no specific recurrent aberration has so far been identified, and there are no good genetic markers that could be used for treatment stratification.

In an attempt to identify clinically and biologically relevant subgroups of UPS, we performed transcriptome sequencing (RNA-Seq), and we here report the finding of two novel gene fusions in UPS, both involving the transcription factor *PRDM10* as the 3′ partner and either *MED12* or *CITED2* as the 5′ partner gene.

Materials and Methods

Patients and tumors

RNA-Seq of two UPS (cases 1 and 2), selected on the basis of their simple karyotypes, showed that they harbored gene fusions involving the *PRDM10* gene. To evaluate the frequency and distribution of *PRDM10* fusions in UPS and other soft tissue sarcomas, a cohort of 82 additional soft tissue sarcomas was analyzed (26 UPS, 22 myxofibrosarcomas, 10 leiomyosarcomas,

[1]Department of Clinical Genetics, University and Regional Laboratories, Lund University, Lund, Sweden. [2]Department of Orthopedics, Karolinska University Hospital, Solna, Sweden. [3]Department of Pathology, Karolinska University Hospital, Solna, Sweden. [4]Department of Orthopedics, Skåne University Hospital, Lund University, Lund, Sweden. [5]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

Corresponding Author: Jakob Hofvander, Department of Clinical Genetics, Lund University Hospital, SE-221 85 Lund, Sweden. Phone: 464-617-3387; Fax: 464-613-1061; E-mail: jakob.hofvander@med.lu.se

doi: 10.1158/1078-0432.CCR-14-2399

Translational Relevance

Undifferentiated pleomorphic sarcoma (UPS) is one of the most common subtypes of soft tissue sarcomas. The clinical behavior is unpredictable, and metastases occur in about one third of the patients. Treatment is based on surgery with wide margins. Depending on surgical margins, location, and tumor-associated risk factors, adjuvant treatment, including radiotherapy and chemotherapy, is considered. Biomarkers that could distinguish UPS from other types of sarcoma as well as improve treatment stratification are needed. Previous genetic analyses have failed to reveal any consistent or tumor-specific aberrations. We here describe the finding of novel, and so far tumor-specific, gene fusions—*MED12*/*PRDM10* and *CITED2*/*PRDM10*—in a subset of UPS. None of the patients with these gene fusions has developed any metastases and all tumors were diagnosed as low-grade malignant at morphologic re-review, suggesting that fusion-positive tumors may represent a less aggressive subset of UPS.

5 low-grade fibromyxoid sarcomas, 5 myofibroblastic sarcomas, 3 myxoid liposarcomas, 2 malignant peripheral nerve sheath tumors, 1 solitary fibrous tumor, 4 spindle cell sarcomas, 1 fibroblastic sarcoma, and 3 unclassifiable sarcomas). The tumors in this extended cohort were partly selected on the basis of their karyotypes. Thus, tumors with structural rearrangements of chromosome arms Xq, 6q, and 11q, that is, the locations of the *MED12*, *CITED2*, and *PRDM10* genes, respectively, at G-banding analysis were retrieved from the archives of the Department of Clinical Genetics in Lund; Xq, 6q, and/or 11q rearrangements were present in 12, 15, and 29 cases, respectively. We also specifically retrieved 16 tumors that had been diagnosed as low-grade malignant UPS, myxofibrosarcoma, or leiomyosarcoma by querying the Scandinavian Sarcoma Group registry. All tumors were diagnosed according to established criteria (1, 8). Clinical, morphologic, and cytogenetic data are presented in Supplementary Table S1. All samples were obtained after written consent and all studies were approved by the institutional ethical committees.

Cytogenetic and FISH analyses

Cell culturing, harvesting, and G-banding were performed as described, and the karyotypes were written following the recommendations of the International System for Human Cytogenetic Nomenclature (9, 10).

FISH was performed on interphase nuclei from cases 2, 27, 29, 52, and 75 using bacterial artificial chromosome (BAC) clones flanking the *PRDM10* locus obtained from the BAC PAC resources. 5′ probes were RP11–664J16, RP11–237N19, and RP11–61J24 and 3′ probes were RP11–1104M18, RP11–121M22, and RP11–110K10. Clone preparation, hybridization, and analysis were performed as described previously (11). No material for FISH was available from case 1.

RNA-Seq

RNA-Seq and bioinformatic analysis to identify candidate fusion transcripts were performed on cases 1, 2, 35, 36, 44, and 49. mRNA libraries were prepared for sequencing using the Truseq RNA Sample Preparation Kit v 2 (Illumina) as previously described (12). Briefly, poly-A–tailed RNA was enriched from total RNA using magnetic oligo-dT beads. RNA was fragmented to a median size of 200 nucleotides and cDNA was synthesized from these fragments using Superscript II reverse transcriptase (Invitrogen). 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-bp reads were generated from the mRNA libraries using the HiScanSQ System (Illumina).

To identify candidate fusion transcripts from the sequence data, analyses were performed on fastq files using Chimerascan (13) version 0.4.5, SOAPfuse (14) version 1.26, and TopHat (15) version 2.0.7. The GRCh37/hg19 build was used as the human reference genome.

Quantitative real-time PCR

To evaluate differences in the expression levels of the 5′ and 3′ parts of *PRDM10*, indicative of a chromosomal breakage within the gene, TaqMan gene-expression assays were performed with: Hs00360640 (*PRDM10* 5′) covering exons 5–6 and Hs000999748 (*PRDM10* 3′) covering exons 20–21. The *TBP* gene was used as endogenous control. Quantitative real-time PCR (qPCR) was performed according to the manufacturer's instructions, and all reactions were run in triplicate (Applied Biosystems). Calculations were done using the comparative C_t method (i.e., $\Delta\Delta C_t$ method; 16) using the SDS software 1.3.1 (Applied Biosystems).

RT-PCR

Total RNA was extracted from frozen tumor samples using the RNeasy Lipid Tissue Kit (Qiagen). Reverse transcription and PCR amplifications were performed as described previously (11, 17). Primers specific for *MED12*, *CITED2*, and *PRDM10* were designed to detect possible fusion transcripts (Supplementary Table S2). Transcripts were amplified using an initial denaturation for 2 minutes at 94°C, followed by 30 cycles of 30 seconds at 94°C, 30 seconds at 58°C, and 3 minutes at 72°C, and a final extension for 3 minutes at 72°C. Amplified fragments were purified from agarose gels and directly sequenced using the Big Dye v1.1 Cycle Sequencing Kit (Applied Biosystems) on an ABI-3130 genetic analyzer (Applied Biosystems). The BLASTN software (http://www.ncbi.nlm.nih.gov/blast) was used for the analysis of *MED12*, *CITED2*, and *PRDM10* sequence data.

Results

Genetic findings in the two index cases

RNA-Seq resulted in 13,955,975 reads in case 1 and 12,758,033 reads in case 2. In case 1, Chimerascan identified a *MED12*/*PRDM10* fusion, supported by three unique flanking reads, and in case 2 SOAPfuse identified a *CITED2*/*PRDM10* fusion supported by two spanning reads and six junction reads. In both cases, the genes implicated in the fusions map to breakpoints identified at G-banding analysis: *MED12* maps to Xq13, *PRDM10* to 11q24, and *CITED2* to 6q24. Thus, both fusions were in agreement with the karyotypes, that is, a t(X;1;11)(q13;p36;q23) in case 1 and a t (6;11)(q24;q24) in case 2 (Fig. 1; Supplementary Table S1). Additional detected potential fusion transcripts were considered read-through transcripts or other artefacts.

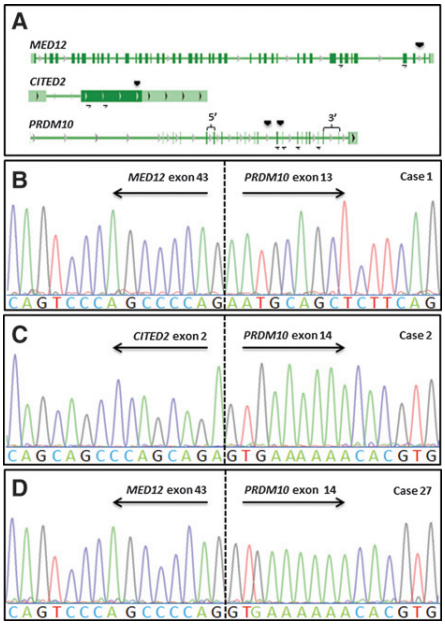


Figure 1.
A, illustration of the *MED12*, *CITED2*, and *PRDM10* genes with vertical arrow heads indicating the breakpoint locations, horizontal arrows indicating the locations of PCR primers, and braces indicating the locations of probes for quantitative real-time PCR. The coding parts of the genes are indicated in dark green color. B to D, partial chromatograms of amplified fragments corresponding to in-frame *MED12*/*PRDM10* and *CITED2*/*PRDM10* fusion transcripts.

RT-PCR and subsequent sequencing of amplified products from cases 1 and 2 identified in-frame *MED12*/*PRDM10* and *CITED2*/*PRDM10* fusions, respectively (Fig. 1). No reciprocal transcript, that is, *PRDM10*/*MED12* or *PRDM10*/*CITED2*, could be detected (data not shown). FISH with *PRDM10*-specific probes in case 2 verified the break in *PRDM10* also at the genomic level (Fig. 2).

The breakpoints in the two 5′ genes (*MED12* and *CITED2*) were located toward the ends of their coding parts. In *MED12*, the breakpoint was located in the intron between exons 43 and 44. *MED12* thus only loses two of its 45 exons in the fusion event. *CITED2* has two exons, and the breakpoint was located within exon 2, at nucleotide position 1047 (NM_006079.4), which is only 9 nucleotide from the stop codon. The shared 3′ partner, *PRDM10*, has 22 exons. In case 1, the fusion breakpoint was located between exons 12 and 13 and in case 2 between exons 13 and 14.

qPCR showed higher expression of the 3′ part of *PRDM10* in both cases. The ratios between the expression levels of the 3′ and 5′ probes were 1.82 and 4.15 in cases 1 and 2, respectively.

Genetic findings in an extended cohort of soft tissue sarcomas

Because of the possibility of multiple 5′ partners to *PRDM10* and the finding of differential expression of the 5′- and 3′-parts of *PRDM10* in the two fusion-positive index tumors (cases 1 and 2), 78 additional soft tissue sarcomas were analyzed by qPCR. Neither the 3′ nor the 5′ expression levels were consistently higher among fusion-positive tumors than among fusion-negative tumors (Supplementary Fig. S1). None of the tumors showed a 3′:5′ ratio above 1.2, whereas six had ratios below 0.7. All these six cases were analyzed by RT-PCR for *MED12*/*PRDM10* and *CITED2*/*PRDM10* fusion transcripts, using multiple primer pairs (Supplementary Table S2), revealing a *MED12*/*PRDM10* fusion in one (case 27, an UPS). Sequencing confirmed a fusion between the last nucleotide of *MED12* exon 43 and the first nucleotide of *PRDM10* exon 14 (Fig. 1). Three of the five RT-PCR negative cases could be analyzed also by interphase FISH using a break-apart probe for *PRDM10*; all were negative. Finally, four myxofibrosarcomas were subjected to RNA-Seq, but did not display any fusion transcript involving *PRDM10*. Thus, only one additional *PRDM10* gene fusion was detected among the 82 soft tissue sarcomas, including 26 UPS, in the extended cohort (Supplementary Table S1).

Morphology of *PRDM10*-positive tumors

All three cases showed features of an UPS with neither morphologic nor immunophenotypic evidence of any specific line of differentiation (Fig. 3). Each consisted of eosinophilic spindled, ovoid, or multinucleate cells with bizarre, irregular, vesicular nuclei. Each had a variably prominent collagenous stroma containing multifocally scattered lymphocytes. In contrast with most pleomorphic sarcomas, in each case, mitoses numbered less than 1 per 10 high power fields and there was no necrosis. These unusual tumors were graded subjectively as low grade based on the experience of one of the authors (C.D.M. Fletcher). Aside from this finding, there were no features that distinguished these tumors from other UPS in general. One case each had focally

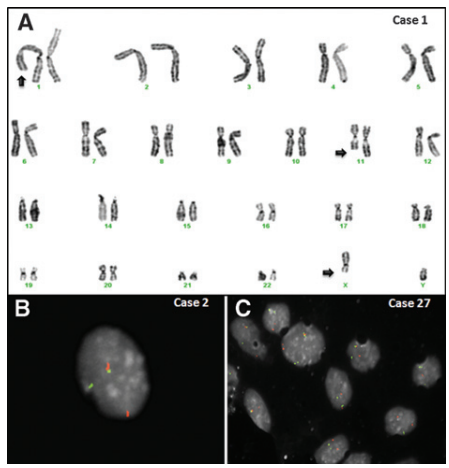


Figure 2.
A, representative karyotype of case 1 showing a translocation t(X;1;11)(q13;p36;q23) correlating with the genomic location of the *MED12* and *PRDM10* genes. B and C, interphase FISH analysis revealing split signals with BAC probes covering 5′and 3′ regions neighboring the *PRDM10* gene.

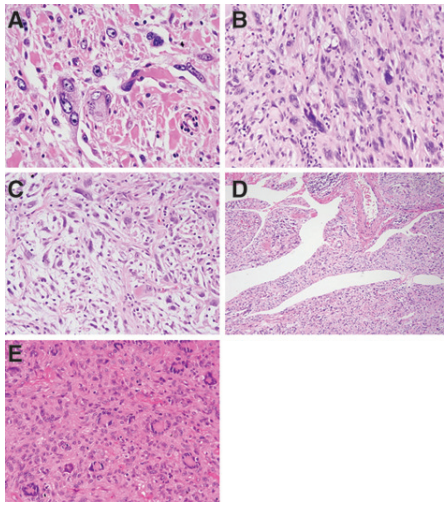


Figure 3.
Morphology of *PRDM10* fusion-positive UPS. A, case 2 showing nondistinctive pleomorphic spindle and ovoid cells. B, high power of case 1 highlights bizarre nuclear morphology. C to E, individual cases showing focally myxoid stroma (case 2), pseudovascular clefts (case 1), and prominent tumor giant cells (case 27).

myxoid matrix (case 2), prominent pseudovascular clefts (case 1) and numerous multinucleate giant cells (case 27), respectively. Tumor cells expressed only CD34, which is not lineage specific.

Discussion

Although UPS is one of the most common sarcoma subtypes, its genetic features remain poorly explored. Marked differences in clinical outcome that cannot be explained merely by differences in tumor size or location, combined with a lack of targeted treatments, provide compelling arguments for more comprehensive attempts to delineate genetic subgroups of UPS. We have in this study been able to identify a small but significant subset of UPS showing gene fusions in which either *MED12* or *CITED2* is fused with *PRDM10*. None of these gene fusions has been described in any other neoplasm, suggesting that they are specific for UPS.

The cohort studied here included a total of 84 soft tissue sarcomas, 28 of which were diagnosed as UPS. The three sarcomas that were positive for *PRDM10* fusions were all diagnosed as UPS, but we find it unlikely that *PRDM10* fusions are present in as much as 10% (3/28) of UPS. Only one of the 26 UPS cases in the extended cohort was positive, and that case had been selected because it had been classified as a low-grade malignant tumor; low-grade malignant lesions constitute a minority of all UPS (1). Also, a comparison between the features of the present cases and previous cytogenetic data on UPS indicate that *PRDM10* fusions are rare events. In the present study, two of three fusion-positive cases had simple karyotypes with a balanced translocation, either as the sole change or together with a few numerical aberrations; the cytogenetic analysis failed in the third case. Abnormal karyotypes have been described in 85 cases of UPS, the majority (57/85) showing highly complex karyotypes with 50 to 100 chromosomes and multiple structural and numerical changes (4); only 13 of the cases had a near-diploid karyotype with less than five structural rearrangements and without any sign of gene amplification (ring chromosomes or double minutes). Thus, we estimate the frequency of *PRDM10* fusion-positive tumors to be around 5% of all UPS.

Even if the *PRDM10* fusion-positive cases constitute a minority of all UPS, it may be clinically important to identify them. The high metastasis rate of UPS, approximately one third, calls for aggressive treatment. Possibly, *PRDM10* fusion-positive tumors have a lower propensity for metastasizing; none of our 3 patients has developed metastases and they were all in complete remission after 41 months to 21 years of follow-up. Furthermore, and in agreement with the favorable outcome, all three were classified as low-grade malignant tumors when re-reviewed; it should be emphasized, though, that two had initially been diagnosed as high-grade lesions. However, there were no distinct morphologic features among the *PRDM10* fusion-positive cases setting them apart from other UPS. Thus, fusions involving *PRDM10* could possibly function as a marker to identify a patient subset with favorable clinical outcome. Needless to say, however, the behavior of *PRDM10* fusion-positive tumors needs to be evaluated in a much larger series of cases, before it can be decided whether they should be treated in other ways than other UPS.

PRDM10 is a poorly studied member of the PRDM (PRDI-BF1 and RIZ homology domain containing) family of proteins. It lacks enzymatic activity and is believed to function as a transcriptional cofactor by recruiting histone-modifying enzymes to target promoters, and is suggested to have an important role during development of the central nervous system (18). The protein is characterized by multiple zinc-finger domains and an N-terminal PR domain (19) . Several other members of the PRDM family are associated with cancer and gene fusions involving *PRDM16* have been reported in cases of acute myelogenous leukemia and myelodysplastic syndrome. *PRDM16* can have several partner genes and all reported fusions lead to overexpression of parts of the gene, usually not containing the PR domain, or the complete gene by promoter swapping (20).

MED12 is part of a large multiprotein complex known as the mediator complex, which functions as a protein bridge between transcription factors and RNA polymerase II to initiate transcription (21). This complex also affects later stages of the transcription process, including elongation and termination. MED12, MED13, Cyclin C, and cyclin-dependent kinase 8 together form a dissociable part of the mediator complex known as the CDK8 module (22). The CDK8 module functions as a negative regulator of transcription by competing for the same binding site as RNA polymerase II on the core mediator complex. However, there are also reports implicating CDK8 as a transcriptional activator (23). This multifunctional module plays major roles in proliferation and differentiation and participates in various molecular pathways, including the p53 and Wnt/β pathways (24). MED12 regulates the kinase activity of the Cdk8 module and mutations in *MED12* are associated with several diseases, including neoplasia. Mutations, especially in exon 2, are found at high frequencies in uterine leiomyoma and fibroadenoma of the breast (25, 26), as well as in malignancies, such as colorectal cancer, leiomyosarcoma, and prostate cancer (21, 27).

CITED2 is a non–DNA-binding transcriptional coactivator that affects the activity of multiple genes by recruiting CBP/p300 to chromatin via the DNA-binding transcription factor AP2. CITED2 also competitively inhibits the transcription of hypoxia-activated genes by blocking the interactions between HIF-A1 and CBP/p300 (28). It is a multifunctional protein best known for its importance during development but also in cancer. It has been reported to be overexpressed in breast cancer in which it modulates the transcriptional activity of the estrogen receptor (29).

It is difficult to make predictions on the functional outcome of fusion genes without further analysis at the protein level. However, it is reasonable to assume that both *MED12*/*PRDM10* and *CITED2*/*PRDM10* act as driver mutations; all previously identified recurrent gene fusions occurring in sarcomas with simple karyotypes, that is, with few or no additional aberrations other than the translocations underlying the fusions, have been shown to be strong driver mutations (30). It is also worth noting that all genes involved in the *PRDM10* fusions play important roles in gene regulation. The breakpoint in *PRDM10* reveals that the PR domain is lost, but nine of the 10 zinc-finger domains are included in the fusion. The breakpoints in *MED12* and *CITED2* are located close to the 3′end of the genes, which might indicate that the functions of these proteins are still intact despite the fusion events. Recruiting functional transcription regulators to a new set of target genes by fusing them to the zinc-finger domains of *PRDM10* could potentially be a mode of action to promote tumor development in these cases.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: J. Hofvander, F. Mertens
Development of methodology: J. Hofvander, J. Tayebwa
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): O. Brosjö, O. Larsson, F.V. von Steyern, N. Mandahl, C.D.M. Fletcher, F. Mertens
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): J. Hofvander, J. Tayebwa, N. Mandahl, C.D.M. Fletcher, F. Mertens
Writing, review, and/or revision of the manuscript: J. Hofvander, J. Tayebwa, F.V. von Steyern, N. Mandahl, C.D.M. Fletcher, F. Mertens
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): J. Hofvander, J. Nilsson, L. Magnusson, F.V. von Steyern, F. Mertens
Study supervision: F. Mertens

Acknowledgments

The authors acknowledge the help from Elisabeth Johansson at the Scandinavian Sarcoma Group central registry.

Grant Support

The study was supported by grants (to F. Mertens) from the Swedish Cancer Society, the National Research Council of Sweden, the Gunnar Nilsson Cancer Foundation, the IngaBritt and Arne Lundberg Foundation, and the Medical Faculty of Lund University.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 22, 2014; revised November 4, 2014; accepted December 2, 2014; published OnlineFirst December 16, 2014.

References

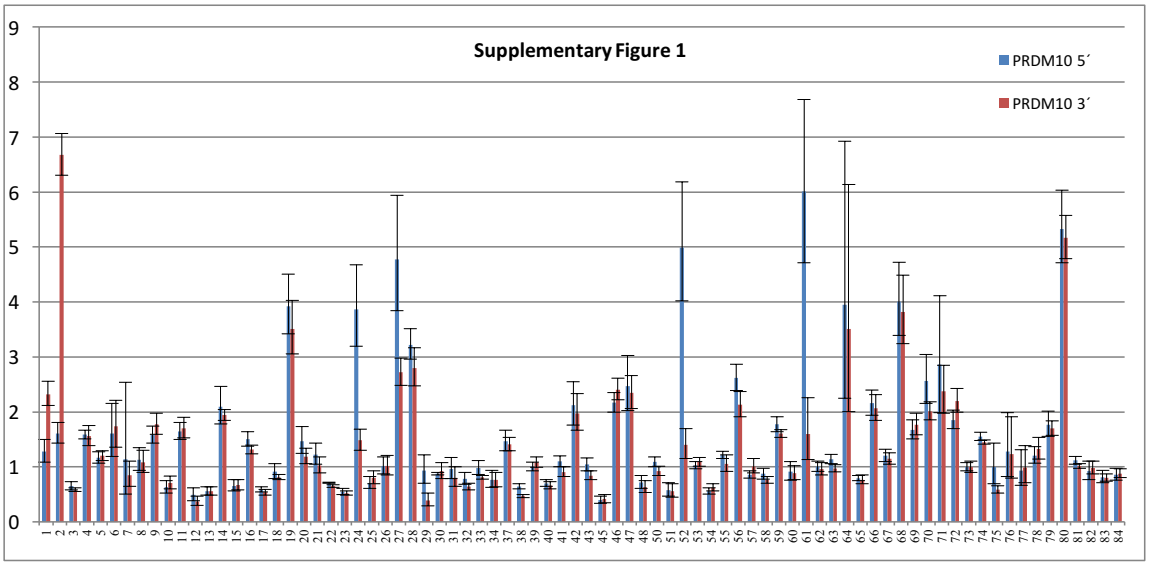
1. Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F (editors). In: WHO classification of tumours of soft tissue and bone. Lyon: IARC Press; 2013. p. 468.
2. Dei Tos AP. Classification of pleomorphic sarcomas: where are we now? Histopathology 2006;48:51–62.
3. Goldblum JR. An approach to pleomorphic sarcomas: can we subclassify, and does it matter? Mod Pathol 2014;27:S39-S46.
4. In: Mitelman F, Johansson B, Mertens F (editors). Mitelman database of chromosome aberrations and gene fusions in cancer. 2014. Available from: http://cgap.nci.nih.gov/Chromosomes/Mitelman.
5. Gibault L, Pérot G, Chibon F, Bonnin S, Lagarde P, Terrier P, et al. New insights in sarcoma oncogenesis: a comprehensive analysis of a large series of 160 soft tissue sarcomas with complex genomics. J Pathol 2011;223:64–71.
6. Fletcher CDM, Chibon F, Mertens F. Undifferentiated/unclassified sarcomas. In:Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F (editors). WHO Classification of tumours of soft tissue and bone. Lyon: IARC Press; 2013. p. 468.
7. Guled M, Pazzaglia L, Borze I, Mosakhani M, Novello C, Benassi MS, et al. Differentiating soft tissue leiomyosarcoma and undifferentiated pleomorphic sarcoma: a miRNA analysis. Genes Chromosomes Cancer 2014;53:693–702.
8. In: Fletcher CDM, Unni KK, Mertens F (editors.). In: World Health Organization classification of tumours. Pathology and genetics of tumours of soft tissue and bone. Lyon: IARC Press; 2002. p. 427.
9. Mandahl N, Mertens F, Willén H, Rydholm A, Brosjö O, Mitelman F. A new cytogenetic subgroup in lipomas: loss of chromosome 16 material in spindle and pleomorphic lipomas. J Cancer Res Oncol 1994;120:707–11.
10. Shaffer LG, McGowan-Jordan J, Schmid M. An International System for human cytogenetic nomenclature (2013). 2013. Basel: Karger. p. 140.
11. Jin Y, Möller E, Nord KH, Mandahl N, Vult Von Steyern F, Domanski HA, et al. Fusion of the *AHRR* and *NCOA2* genes through a recurrent translocation t(5;8)(p15;q13) in soft tissue angiofibroma results in upregulation of aryl hydrocarbon receptor target genes. Genes Chromosomes Cancer 2012;51:510–20.
12. Walther C, Tayebwa J, Lilljebjorn H, Magnusson L, Nilsson J, von Steyern FV, et al. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. J Pathol 2014;232:534–40.
13. Iyer MK, Chinnaiyan AM, Maher CA. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics 2011;27:2903–4.
14. Jia W, Qiu K, He M, Song P, Zhou Q, Zhou F, et al. SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. Genome Biol 2013;14:R12.
15. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 2013;14:R36.
16. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 2001;25:402–8.
17. Panagopoulos I, Mertens F, Domanski HA, Isaksson M, Brosjö O, Gustafson P, et al. No *EWS*/*FLI1* fusion transcripts in giant cell tumors of bone. Int J Cancer 2001;93:769–72.
18. Hohenauer T, Moore AW. The Prdm family: expanding roles in stem cells and development. Development 2012;139:2267–82.
19. Fog CK, Galli GG, Lund AH. PRDM proteins: important players in differentiation and disease. Bioessays 2012;34:50–60.
20. Duhoux FP, Ameye G, Montano-Almendras CP, Bahloula K, Mozziconacci MJ, Laibe S, et al. PRDM16 (1p36) translocations define a distinct entity of myeloid malignancies with poor prognosis but may also occur in lymphoid malignancies. Br J Haematol 2012;156:76–88.
21. Wang H, Shen Q, Ye LH, Ye J. MED12 mutations in human diseases. Protein Cell 2013;8:8.
22. Carlsten JO, Zhu X, Gustafsson CM. The multitalented Mediator complex. Trends Biochem Sci 2013;38:531–7.
23. Galbraith MD, Donner AJ, Espinosa JM. CDK8: a positive regulator of transcription. Transcription 2010;1:4–12.
24. Szilagyi Z, Gustafsson CM. Emerging roles of Cdk8 in cell cycle control. Biochim Biophys Acta 2013;9:916–20.

25. Mehine M, Kaasinen E, Mäkinen M, Katainen R, Kämpjärvi K, Pitkänen E, et al. Characterization of uterine leiomyomas by whole-genome sequencing. N Engl J Med 2013;369:43–53.
26. Lim WK, Ong CK, Tan J, Thike AA, Ng CCY, Rajasegaran V, et al. Exome sequencing identifies highly recurrent *MED12* somatic mutations in breast fibroadenoma. Nat Genet 2014;46:877–80.
27. Kampjarvi K, Makinen N, Kilpivaara O, Arola J, Heinonen HR, Bohm J, et al. Somatic MED12 mutations in uterine leiomyosarcoma and colorectal cancer. Br J Cancer 2012;107:1761–5.
28. Chen CM, Bentham J, Cosgrove C, Braganca J, Cuenda A, Bamforth SD, et al.Functional significance of SRJ domain mutations in CITED2. PLoS ONE 2012;7:17.
29. Lau WM, Doucet M, Huang D, Weber KL, Kominsky SL. CITED2 modulates estrogen receptor transcriptional activity in breast cancer cells. Biochem Biophys Res Commun 2013;437:261–6.
30. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 2007;7: 233–45.

Supplementary information for:

Recurrent PRDM10 gene fusions in undifferentiated pleomorphic sarcoma.

Hofvander J, Tayebwa J, Nilsson J, Magnusson L, Brosjö O, Larsson O, Vult von Steyern F, Mandahl N, Fletcher CDM, Mertens F. *Clin Cancer Res*. 2015;21:864-869.



Supplementary Figure 1. Quantitative real-time PCR results for the 5'- and 3'-parts of *PRDM10* in a cohort of 80 soft tissue sarcomas. The *TBP* gene was used as endogenous control, and case No. 73 was used as calibrator. The three gene fusion-positive cases of undifferentiated pleomorphic sarcoma (Cases 1, 2, and 27) showed differential expression of the two parts of *PRDM10*).

Supplementary Table S1: Morphological, Clinical, and Genetic Features of 85 Soft Tissue Sarcomas Analyzed for Gene Fusions Involving the *PRDM10* Gene

Case No	Diagnosis	Age/sex	Site	Size	Grade	FU	Karyotype	qPCR 3'/5'	FISH	RT-PCR	RNA-Seq	Karyotype previously published
1	UPS	30/M	Shoulder/D	1	See results	NED 21 yrs, LR 15yrs	45-46,Y,**t(X;1;11)(q13;p36;q23)**	1.82		*MED12-PRDM10*	*MED12-PRDM10*	
2	UPS	41/F	Shoulder/D	3	See results	NED 61	48,XX,**t(6;11)(q24;q24)**,+7,+16/87-97,idemx2,tas(9;21)(q34;q22)/139-176,idemx3-4	4.15	Pos	*CITED2-PRDM10*	*CITED2-PRDM10*	
3	UPS	68/F	Femur	15	3	DoD 3, M 0	44-47,X,**ins(X;?)(q13;?)**,i(1)(p10),add(2)(q31),add(3)(p25)x2,del(5)(p11),-8,add(9)(q22),der(10)t(3;10)(q11;p15),add(12)(p11),add(13)(p11),i(14)(q10),i(15)(q10),add(18)(q23),inc	0.89				
4	UPS	83/F	Trunk/D	11	3	DwoD 15	71-78,**?add(X)(q11)**,add(1)(q32)x2,add(1)(q21),der(1)del(1)(q12)dup(1)(p21p36),der(2)del(2)(p12)t(2;7)(q3?1;q11)x1-2,add(6)(p23),i(9)(q10),add(11)(p15),add(12)(p11)x2,del(16)(q13),add(18)(q23),add(19)(q13),?add(22)(q13),10dmin,inc	0.98				
5	UPS	83/M	L leg/S	6	3	DoD 47, M 13	43-47,-Y,**add(X)(q22)**,-4,-4,+5,add(5)(q11)x2,-6,+7,+9,+9,-13,add(14)t(5;14)(q13;p13)x2,+add(16)(q13),-18,add(20)(p13),-21,-22,+der(?)t(?;4)(?;q11),+1-2r/83-97,idemx2,+mar	1.03				
6	UPS	73/M	Thigh/S	6	3	NED 21	65-70,XX,-Y,del(1)(q21),-2,del(2)(q3?3),-3,-3,add(4)(q35),-6,**?del(6)(q2?3)**,t(7;12)(q11;q13),-8,-9,-10,-13,+14,-15,?del(16)(q12),-17,-17,?del(17)(p11),-18,-18,-19,+2-4r,inc	1.08				
7	UPS	61/F	Thigh/D	8	3	LTF	75-80,X,-X,-X,?add(1)(p36),del(1)(q11),?del(1)(p12),-5,**?ins(6)(q?;?),add(6)(q1?3)**,-10,add(10)(q22)x1-2,add(11)(p13)x1-2,+der(11)t(1;11)(p21;p14),add(12)(q24)x2,-13,-14,-14,-16,-19,-20,-21,+r,inc	0.75				
8	UPS	70/F	Thigh/D	5	3	NED 74	67-74,XX,-X,add(1)(q21),der(1;17)(p10;q10),+der(1;17),del(2)(p13)x2,+3,del(3)(p12)x2,-4,add(4)(p12)x2,+5,**der(6)t(?4;6)(q27;q23)x2**,del(7)(q11),-8,-9,-10,add(10)(p13)x2,-11,add(11)(p1?1)x2,-12,add(12)(p13),dic(12;?)(p12;?),-13,-13,-13,+14,ins(14;?)(q11;?)x2,+16,der(16)t(1;16)(q23;q13)t(1;8)(q44;q11)x2,-17,add(17)(p11)x2,-18,-19,add(19)(q13),-22,der(22)t(7;22)(q11;p11)x2,+der(?)t(?;9)(?;p13),+r,inc	0.96				
9	UPS	87/M	Thigh/D	14	3	DoD 11, M 9	64-72,XXY,add(1)(p35),del(1)(p11),del(3)(p21),add(5)(p1?4),**add(6)(q?)x1-2**,del(7)(p11),?dup(9)(q13q34),**der(11)add(11)(q23)hsr(11)(q23)x1-2**,del(12)(p12)x2,add(14)(p11),?add(17)(p11)x1-2,add(19)(q13),inc	1.12				
10	UPS	77/M	U arm/S	7	3	AwD 26, M 17	62-66,XX,+X,-Y,add(1)(q32),add(1)(q11),del(1)(q11),-2,-4,-5,-5,-5,-6,-6,**?del(6)(q11)**,-7,-7,-9,-10,**add(11)(q14)**,-12,add(12)(p12),-16,add(16)(q22),-17,-18,-20,-21,-21,-22,+4mar,dmin,inc	1.13				
11	UPS	61/M	Chest wall/D	10	3	NED 56	50-57,XY,add(1)(p36),der(1)add(1)(p32)del(1)(q11),?del(4)(p14),?dic(4;?)(p12;?),del(5)(p14),**add(6)(q21),del(11)(q22)**,del(12)(p12),add(19)(p13),add(20)(q1?1),+r,+2mar,inc/108-110,XY,inc	1.03				
12	UPS	73/M	Thigh/D	9	3	DoD 21, M 9	75-80,X?,del(1)(q32),del(3)(p11),**?del(11)(q23)**,?add(19)(p13),+r,inc	0.76				
13	UPS	65/F	Trunk/D	8	3	DoD 9, M 5	80-84,XXX,del(1)(q32)x3,der(1)del(1)(q32)add(1)(p36),add(6)(p25),add(11)(p13),**del(11)(q23)**,add(13)(p11),add(19)(q13),der(19)t(12;19)(q11;p11),inc	0.99				
14	UPS	41/M	Back/S	9	3	DoD 36, M 20	80-100,X?,?del(1)(p21p34),?i(3)(q10),**add(11)(q24)**,add(19)(p13),+der(?)t(?;7)(?;p13),inc	0.93				
15	UPS	63/M	Thigh/D	10	3	DoD 11, M 2	58-70,XY,-X,add(1)(p36),del(1)(q32),-2,add(3)(p2?6),add(5)(p15),add(6)(p25),add(7)(p22),+add(7)(q34),add(9)(q34),**del(11)(q22),der(11)add(11)(p15)add(11)(q25)**,-16,?add(17)(p11),?add(19)(q13),inc/112-144,idemx2	1.02				
16	UPS	88/F	Chest wall/D	10	3	NED 13	80-81,XXX,+1,+2,+3,add(3)(p11)x2,+4,+?add(7)(p1?),+8,+8,-10,-14,+16,-18,+19,+20,+mar,inc/74-79,XX,-X,del(1)(p31),+2,der(3;11)(p10;q10),+add(3)(p11),add(5)(p11),+?add(7)(p1?),+8,-10,-14,+16,-18,+19,+20,+1-3mar,inc	0.87				

17	UPS	78/M	Shoulder/D	6	3	LTF	40-42,X,-Y,-1,-2,-3,-4,add(5)(q31)x2,-6,del(7)(q22),-8,add(8)(p11),-9,-9,-10,-10,-12,add(12)(q22),-13,-15,+add(16)(p12),-17,-18,add(19)(p13),der(19)t(3;19)(q11;p11),-20,-21,-21,+der(?)t(?;10)(?;q11)x2,+r,+mar,inc/70-79,XY,-X,-1,-2,-4,+5,add(5)x2,-6,add(6)(p12)x2,del(7),-8,add(8)x2,-9,-9,-9,-10,add(10)(q22)x2,+11,-12,add(12)x2,-13,-13,-15,+16,add(16)x2,-17,-18,+19,add(19)x2,+der(19)t(3;19)x2,-20,-21,-21,-21,+22,+der(?)t(?;10)x2,+1-5r,+2mar,inc	0.91				
18	UPS	62/F	Shoulder/D	8	3	NED 106, M 34	85-114,X?,add(1)(p11),der(1)add(1)(p36)add(1)(q31)x2,?del(2)(p12),?add(3)(q12),add(12)(p1?2),?add(15)(p11),add(19)(p13)x1-2,+der(?)t(?;2)(?;q1?)x2,+r,?1-4dmin	0.89				
19	UPS	72/M	Thigh	13	3	DoD 37, M 32	70-83,XXY,add(1)(q21),del(1)(q12)x1-2,+add(1)(q12)x1-2,add(2)(q?),?add(3)(q11),add(10)(q24),?del(11)(p11),+r,inc	0.90				
20	UPS	89/F	L arm/S	8	3	AwD 29	58-64,XX,-X,add(1)(q11),del(1)(q11),-2,add(3)(p11),-4,-5,-7,del(9)(q12),del(10)(p12)x2,+del(11)(p11),-13,-14,-15,-16,-17,-18,-21,add(21)(p11),?add(22)(q11)x2,+1-3r,inc	0.81				
21	UPS	80/M	Thigh/D	7	3	NED 63	40-41,XY,?del(2)(p1?2),-4,?add(5)(q?),-6,-11,-13,-14,-17,-20,-21,-22,+r,+3mar	0.84				
22	UPS	75/F	L arm/D	?	3	AwD 24	46,XX	0.91				
23	UPS	88/F	L leg/D	?	3	DoD 56, M 54	Failure	0.95				
24	UPS	70/M	Thigh/D	15	2	DoD 24, M12	79-86,XY,**del(X)(q22)**,-Y,add(1)(p36),del(1)(q11),del(1)(q42),-2,-2,del(2)(p16),-3,add(4)(p16)x2,-5,add(5)(p15)x2,-6,add(6)(p23),add(7)(p22)x2,-8,**t(8;11)(p13;q21)**,add(9)(q?),-10,dup(10)(q11q26)x2,-11,del(11)(p11),-12,add(12)(p13),-13,-13,-14,-14,-15,add(15)(q?),add(16)(q24),-17,-17,-17,-18,-18,-19,add(19)(q13)x2,-20,-20,-21,-21,-22,add(22)(q13),+6-15mar	0.38		Neg		Ref 3054, Case 22
25	UPS	69/F	L leg/S	2	2	DoD 31, M 18	43,XX,der(1)t(1;9)(p11;q13)ins(1;?)(p11;?),add(2)(p21),del(3)(p12),der(5)t(5;12)(p13;q13),-6,-8,-9,**der(11)t(11;17)(q23;q21)**,-12,-13,-17,+der(?)t(?;8)(?;q13),+r,+mar	1.13				
26	UPS	75/M	Knee/S	1	2	NED 78	43-47,XY,der(1)t(1;7)(q21;q22),add(3)(p13),der(4)t(1;4)(p22;q31),der(12)t(2;12)(q11;q11),add(17)(p11),add(19)(q13),+1-2r,inc	1.00				Ref 5609, Case 49
27	UPS	58/F	Foot/S	3	See results	NED 41	Failure	0.57	Pos	*MED12-PRDM10*	*MED12-PRDM10*	
28	UPS	37/M	Scalp	?	?	DoD 17	48-100,XX,-Y,del(3)(p12),add(7)(p22)x2,der(9)t(1;9)(p11;p22)x2,**add(11)(q25)**,inc	0.87				Ref 8379, Case 944
29	MFS	74/M	Thigh/S	10	3	NED 154	46,Y,**add(X)(q24)**,add(2)(q33),add(12)(q11),add(14)(q24),del(15)(q22),add(17)(q23),-18,der(22)t(18;22)(q11;q12),+mar	0.42	Neg	Neg		Ref 13848, Case 35
30	MFS	77/F	Trunk/D	4	3	DoD 42	61-67,XX,-X,add(1)(p36),add(2)(q21),del(2)(q32),add(3)(q11),add(4)(p16),der(4)add(4)(p16)add(4)(q?),add(5)(q35),**del(6)(q11)**,del(7)(q31),der(7)add(7)(p22)add(7)(q32),add(8)(q24),?del(10)(p12),**der(11)add(11)(p15)add(11)(q22),der(11)t(11;11)(p15;q11)ins(11;?)(p15;?)**,der(12)t(5;12)(q13;q13),dic(16;18)(q13;q23),add(17)(p11),?add(17)(p13),add(19)(q13),der(20)t(1;20)(q44;p13)del(1)(q12),?i(22)(q10),inc	1.05				Ref 5609, Case 33
31	MFS	80/M	Neck/D	7	3	NED 59	51-59,XY,add(1)(q21)x2,-4,-4,-6,?der(6)del(6)(p23),**add(6)(q15)**,+7,-9,add(10)(p1?),**?der(11)add(11)(p13)del(11)(q23)**,del(12)(q24),add(13)(p11),?i(14)(q10),add(15)(p11),?add(16)(q22),-17,-18,-18,inc/95-109,idemx2	0.84				
32	MFS	64/F	Thigh/D	11	3	DoD 60, M 36	54-58,XX,add(2)(q23),+5,+7,+8,**add(11)(q23)**,+17,+19,+20,+2-4mar	0.82				
33	MFS	65/F	Thigh/D	5	3	NED 64	73-79,XXX,del(1)(q32),?add(3)(p25),**del(11)(q21)x2**,add(12)(p11),?add(19)(q13),inc	0.83				
34	MFS	71/M	Thigh/S	25	3	NED 85	81-94,XX,-Y,-Y,-1,-1,-3,-3,-4,-4,-4,+7,add(7)(q31)x2,**dup(11)(q12q25)x2**,-13,-13,+19,+19,+add(19)(q13)x2,+20,+20,+20,+22,+3mar	0.99				Ref 13848, Case 44
35	MFS	43/F	Thigh/D	12	3	DoD 32, M 12	67-78,XXX,i(7)(p10),?add(19)(q13)x2,+1-2r,inc	ND			Neg	
36	MFS	31/M	L leg/D	9	3	NED 84	48-52,X,-Y,+7,+8,?-9,der(14;14)(q10;q10),-16,+17,+18,+2mar,inc	ND			Neg	
37	MFS	84/M	U arm/D	10	3	LTF 30, M 7	43-44,XY,-2,-5,add(7)(q32),add(8)(p11),-9,add(13)(p11),i(14)(q10),-17,-18,-20,inc/81-84,idemx2,add(12)(p11)x2	0.96				Ref 7478, Case 217

38	MFS	79/F	Elbow/D	6	2	NED 18	80-85,XXXX,del(1)(q12),**add(6)(q15)**,add(11)(p15),add(12)(q22),add(18)(p11),inc	0.73					Ref 7478, Case 219
39	MFS	41/M	Thigh/D	12	2	NED 63	85-87,X?,del(1)(q42),add(2)(q11),add(3)(q12)x2,**?add(6)(q15)**,del(7)(q13),del(11)(p13),**?del(11)(q23)**,?del(12)(p11),add(15)(q22),add(17)(q25),add(19)(q13)x1-3,inc	1.08					Ref 10805, Case 27
40	MFS	82/F	U arm/D	13	2	DoD 94	80-86,XX,add(1)(p11),add(1)(q21),del(1)(q32)x2,del(2)(p12),del(7)(p13)x2,add(9)(q22),del(9)(q22),**der(11)del(11)(p13)add(11)(q25)**,add(12)(p11)x2,add(13)(q32),add(16)(p13)x2,add(17)(p11)x2,der(19)hsr(19)(q13)add(19)(q13)x2,add(20)(q13)x2,inc	0.94					
41	MFS	76/F	U arm/S	5	2	120	71-83,XXX,**add(11)(q22)**,**?del(11)(q23)**,+3mar,inc	0.82					
42	MFS	61/M	U arm/S	3	2	NED 70	55-65,XXY,add(1)(q11),add(2)(p13)x2,+der(2)t(2;4)(p21-22;q13-21),add(3)(p13),del(3)(p13p21-23),-4,-4,-6,der(7)t(7;13)(q22;q14)x2,-10,-11,**add(11)(q24-25)**,-12,-13,der(14;21)(q10;q10),-16,+17,add(17)(p12-13)x2,-19,-19,der(19)t(5;19)(q13;p13),-20,-21,-21,add(22)(q13),+4-6r,inc	0.93					
43	MFS	79/F	L leg/D	21	2	DoD 30, M22	65-69,X,-X,-X,+del(1)(q11),+del(1)(q12),del(2)(q31),+del(3)(p12),add(4)(p16),add(4)(p11),+6,+der(7)inv(7)(p15q11)add(7)(q36),+inv(7),del(8)(q?),del(9)(q11)**add(11)(q25)**,del(11)(p11),**+der(11)add(11)del(11)(p15)**,add(13)(q?),+14,+14,+15,add(16)(q23),del(17)(p11),+18,+19,+21,add(22)(q13),+3-7mar	0.81					Ref 3054, Case 29
44	MFS	42/F	Foot/S	4	2	NED 149	40-45,X,-X,add(1)(p11),del(1)(q32),?del(3)(p11),der(4)t(?3;4)(p23;p16),del(5)(p1?3p1?5),add(7)(p22),?i(8)(q10),-9,-10,-11,-13,-14,add(16)(p11),+19,add(19)(p13)x2,-22,+2mar,inc	ND			Neg		
45	MFS	85/F	Thigh/D	10	2	NED 56	48,XX,+7,+8	1.03					
46	MFS	44/M	Chest wall/S	6	2	NED 71	44-45,XY,-6,-8,-9,-10,?inv(10)(p12q24),?dup(17)(q11q21),+2mar	1.11					Ref 13848, Case 43
47	MFS	74/M	Trunk/S	7	1	DoD 82	41-42,XY,add(1)(q23),der(1;2)(q10;q10),-2,-2,der(2;16)(p10;q10),ins(3;?)(p21;?),-4,del(9)(p13),-10,-13,-15,add(17)(q11),-18,+22,+r,+mar/46,XY,t(2;3)(p23;q21)/45,X,-Y	0.95					Ref 13848, Case 34
48	MFS	71/M	Thigh/S	3	1	NED 26	46,XY	0.88					
49	MFS	35/F	Thigh/S	9	?	NED 107	70-120,X?,+r,inc	ND			Neg		
50	MFS	85/F	Hip/D	14	?	LTF	49-50,X,-X,+1,+der(1;14)(p10;q10),der(1)add(1)(p22)add(1)(q32)x2,+5,add(5)(p15)x2,+12,?del(16)(q22),?dup(17)(q21q25),-18,+19,+22,+mar	0.85					Ref 10805, Case 23
51	LMS	57/F	Neck/D	4	3	DoD 12, M 5	67-75,X,**add(X)(q28)**,**del(X)(q23)**,der(1;?)inv(1)(p12p36)dic(1;?)(p12;?)x2,add(3)(p14),del(3)(p14),del(3)(q12),add(4)(p16)x2,add(5)(q35),**del(6)(q15)x2**,der(7)del(7)(p13)ins(7;?)(q32;?),ins(7;?)(q32;?),inv(7)(p15p22),del(8)(p21),add(10)(p15),del(11)(p12)x2,**del(11)(q11)**,add(12)(q24),add(13)(p11),der(14)add(14)(p11)add(14)(q32)x3,add(16)(p13)x2,add(17)(q25),der(18)add(18)(p11)add(18)(q23),add(19)(p13)x2,add(22)(p11),+hsr(?),inc	0.97					Ref 8300, Case 2
52	LMS	80/M	Thigh/D	6	3	DWoD 30	87-97,XY,**?add(X)(q13)**,**der(X)add(X)(p21)add(X)(q22)**,del(1)(q11)x1-2,add(2)(p11),add(3)(q27),del(3)(p11),der(3)del(3)(p13p21)ins(3;?)(p13;?),add(7)(p22),del(9)(q22),add(10)(q22)x2,**del(11)(q22)**,add(13)(q22),add(16)(p13)x2,add(20)(q13)x2,+der(?)t(?;6)(?;p21)x3,+r,inc	0.28	Neg	Neg			
53	LMS	80/M	Groin/S	12	3	DoD 35	40-42,XY,der(1)del(1)(p32)add(1)(q44),add(3)(q11),der(3)del(3)(p13p23)t(3;12)(q29;q13),-4,-5,-5,-9,-9,del(12)(q13),-13,-14,-14,der(15)t(15;15)(p11;q13),der(16)t(12;16)(q13;q24)ins(16;?)(q24;?),-17,-17,-18,**+der(?)t(?;11)(?;q14)hsr(11)(q14)**,+5mar	1.06					Ref 8300, Case 5
54	LMS	63/F	Thigh/D	10	3	NED 54	43-47,X?,+1-4r,inc/83-90,idemx2	1.12					Ref 10805, Case 35
55	LMS	79/M	Thigh/D	6	3	DoD 32, M 7	43-47,X,-Y,der(11;16)t(11;16)(p15;q24)ins(11;?)(p15;?),+1-2r,inc	0.87					
56	LMS	82/M	Thigh/D	9	3	DoD 25, M 6	41-43,XY,del(1)(q11),-2,-6,-9,-10,-12,-13,-17,-18,+r,+3mar/79-85,idemx2,inc/45,X,-Y	0.81					

57	LMS	55/M	Groin/S	7	3	DoD 56, M 21	43-47,XY,-5,+2mar/92-93,idemx2	1.18					
58	LMS	63/F	Perineum/S	4	3	NED 36	75-78,XXX,+X,+1,der(1;6)(q10;p10)x2,-2,+3,-4,+5,+6,+7,+8,+9,-10,+11,+12,-13,add(15)(p13)x2,+16,+18,+19,+20,der(20)t(4;20)(q11;q13)ins(20;?)(q13;?)x2,+21,+22/76-77,idem,add(17)(q25)	0.87					Ref 8300, Case 6
59	LMS	86/F	L leg/S	<5	3	DwoD 51	44,X,add(X)(p22),-10,-13,add(15)(q15),add(16)(q24),der(16)t(15;16)(q15;p13),add(17)(p11),-22,+r/43,idem,-2/45,idem,+8	0.91					Ref 8300, Case 3
60	LMS	63/M	Thigh/D	15	2	NED 70	51,XY,der(1)del(1)(p36)add(1)(q32),add(3)(p25),add(7)(p11),del(7)(q11q22),add(12)(q24),add(17)(q25),add(19)(q13),der(19)t(5;19)(q13;p11)/45-49,idem,-add(3),+del(3)(q11)/81-97,idemx2,-add(3)x2,+del(3)x2/42-44,idem,add(1)(q32),del(4)(p14),add(5)(p15),-del(7),-add(19),-der(19)	0.97					Ref 5609, Case 40
61	LGFMS	34/M	Back/D	9	1-2	LTF 60, M 48	Not done; *FUS-CREB3L2* fusion-positive	0.27		Neg			
62	LGFMS	38/M	Thigh/D	9	1-2	NED 2	46,XY,t(7;16)(q32;p11)	0.95					Ref 10986, 16
63	LGFMS	77/M	Buttock/D	10	1	LTF	30-40,XY,der(1)t(1;5)(p36;q13),-5,-6,t(7;16)(q33;p11),-8,-14,-18,-19,+mar	0.86					
64	LGFMS	46/M	U arm/D	3	1	NED 158	47,XY,+r	0.89					Ref 10582, Case 6
65	LGFMS	23/M	Thigh/D	4	1	NED 132	Not done; *FUS-CREB3L2* fusion-positive	0.96					
66	Myofibroblsarcoma	58/M	Thigh/D	9	3	AwD 109	34-35,X,-Y,add(1)(q12),-2,-4,-5,**add(6)(q23)**,-7,-9,-9,-10,-11,-12,add(12)(p11),-13,add(14)(p11),-15,-16,add(16)(p13),add(17)(p11),add(19)(q13),add(21)(p11),+r,?dmin/63-66,idemx2	0.96					
67	Myofibroblsarcoma	82/M	Knee/S	14	3	LTF	68-75,-X,-X,-Y,del(1)(p22p33)x2,+der(1)t(1;11)(p36;q11)x2,+2,del(2)(p14)x2,add(3)(p25)x2,+del(3)(p21)x2,+4,-5,+7,add(7)(p15)x2,+8,?inv(9)(p13p22),del(10)(p11p11)x2,+12,der(12)add(12)(p11)add(12)(q24)x2,+14,+14,+14,-15,-17,+18,add(19)(p13),+20,+22,inc	0.95					Ref 5609, Case 32
68	Myofibroblsarcoma	30/F	Thigh/D	2	1-2	NED 29	46,XX	0.95					
69	Myofibroblsarcoma	81/F	Chest wall/D	8	1-2	NED 97	46,XX	1.06					
70	Myofibroblsarcoma	74/F	Thigh/D	4	1	NED 72	74-78,XX,-X,-1,del(1)(q42)x2,+2,add(3)(p13),-4,+5,**add(6)(q23)x2**,+7,add(8)(p23)x2,-10,-11,**del(11)(q23)x2**,-12,-13,add(13)(q34)x2,-14,+15,add(15)(p11)x2,-16,-17,-17,-17,+18,-19,dic(21;21)(p12;p12)x2,+der(?)t(?;9)(?;q13),+r,+mar,inc	0.79					
71	MLS	54/M	L leg/D	17	3	NED 84	47,XY,+8,t(12;16)(q13;p11)/48,idem,+13/48,idem,i(7)(q10),+13/46,idem,i(7)(q10),-8	0.83					
72	MLS	34/M	Thigh/D	6	2	NED 4	46,XY,t(12;16)(q13;p11)	1.19					
73	MLS	36/F	Thigh/D	6	2	NED 90	46,XX,t(12;16)(q13;p11)	1.00					Ref 5546, Case 10
74	MPNST	67/F	Groin/S	7	2	NED 4	69-72,XXX,+2,-3,-6,-7,add(7)(q22),add(?8)(p11)x2,-9,-10,**add(11)(q23)**,-13,-17,-17,-18,+19,-20,+der(?)t(?;5)(?;q22)x1-2,+r,inc	0.93					
75	MPNST	62/F	Thigh/D	?	?	NED 85, LR 48	45-46,X,**?add(X)(q11)**,-1,-3,del(3)(q25),-5,-6,-9,-10,-11,-14,-15,add(17)(p11),-19,-20,?add(21)(q22),+der(?)t(?;5)(?;q13),+6-12mar	0.59	Neg	Neg			
76	SFT	35/F	Thigh/D	11	2	NED 144	46,XX; *NAB2-STAT6* fusion-positive	0.96					
77	Spindle cell sarcoma	72/F	Thigh/D	7	3	NED 12	83-87,-X,add(X)(p22),**del(X)(q26),der(X;8)(q10;q10)**,add(1)(q11),del(1)(q11),del(2)(p11),der(2;6)(q10;p10),der(2)del(2)(p11)add(2)(q36),del(3)(q11),-4,-4,del(5)(p11),add(6)(p11)x2,**del(6)(q23)**,add(7)(p11),add(7)(p13),-9,add(9)(q34),der(9)t(9;13)(q13;q12),-10,-10,-10,add(11)(p11),**del(11)(q14),del(11)(q21)**,-12,add(12)(q24),-13,-13,-13,-14,-14,-14,-15,-15,-15,add(?16)(q24)x2,-17,-17,-17,-17,-18,-18,-20,-20,-20,add(20)(q13),-22,-22,+r,inc	1.06					
78	Spindle cell sarcoma	75/M	Thigh/D	25	2	NED 16	43-46,XY,der(3;?8)(p10;q10),+der(3;?)(p14;?)t(?;12)(?;q13),der(5)t(5;12)(p15;q11),**add(6)(q25),add(11)(q13)**,+12,-13,-14,-15,-15,-18,add(19)(p11),-21,+3mar,inc	1.10					

79	Spindle cell sarcoma	73/M	Thigh/D	17	2	DoD 12, M 5	69-82,XXY,add(1)(p36),der(1)add(1)(p36)add(1)(q21)x2,+der(1)add(1)(p12)dup(1)(q32q44),del(2)(p11),add(4)(p16),+5,add(5)(p15)x2,del(9)(q22),del(11)(p11),+r,inc	0.96				Ref 5609, Case 34
80	Spindle cell sarcoma	35/F	Heart	?	?	DoD 4	82-89,XX,-X,**?del(X)(q11)**,der(1;?7)(p10;p10)x2,der(1)t(1;?)(p11;?)t(?;1)(?;q25),add(2)(q3?),del(3)(q12),add(13)(p11),?add(14)(q13),?add(18)(p11)x2,add(19)(q13)x2,+r,inc	0.97				
81	Fibroblastic sarcoma	83/M	Buttock/S	8	2	DoD 51, M 21	41-44,XY,add(1)(p36),del(2)(q31),del(3)(p13p21),-4,add(6)(p21),-9,der(10)t(10;12)(p15;q13),**add(11)(q25)**,-12,-16,inc/79-84,idemx2	0.91				
82	Sarcoma NOS	65/F	Shoulder/D	11	3	DoD 19	45-57,-X,**i(X)(q10)**,tas(1;19)(q44;p13),add(7)(q33),add(12)(p13),+mar,inc	1.06				
83	Sarcoma NOS	70/M	U arm/S	4	3	NED 51	49,XY,+X,-1,del(2)(q33),+5,?del(6)(p11),add(8)(q24),-10,**der(11)t(1;11)(q21;q24)**,+12,+14,-15,add(17)(p11)x2,del(22)(q11),+der(?)t(?;10)(?;q11),+mar	0.98				
84	Sarcoma NOS	77/M	Retroper/D	?	?	NED 1	73-80,Y,-X,**add(X)(q13)**,+2,+4,+add(5)(p15),+6,+add(7)(q11),+8,-9,+10,-11,-12,-12,-13,add(13)(p13),-15,+add(16)(q24)x1-2,+add(17)(q25),+18,+19,**+der(19)t(11;19)(q12;q13)ins(11;?)(q12;?)x1-2**,+20,+add(20)(q13),-21,-22,+3mar	1.02				

Diagnosis: UPS = Undifferentiated pleomorphic sarcoma; MFS= Myxofibrosarcoma; LMS = Leiomyosarcoma; Myofibrobl = Myofibroblastic; MLS = Myxoid liposarcoma; MPNST = Malignant peripheral nerve sheath tumor; SFT = Solitary fibrous tumor;

Sarcoma NOS = Sarcoma not otherwise specified

Site: D = Deep; S = Subcutaneous; L leg = Lower leg; U arm = Upper arm; L arm = Lower arm; Retroper = Retroperitoneal

Size: Largest diameter in cm

Grade: Malignancy grade according to 3-grade scale

FU: Follow-up in months, unless otherwise indicated. NED = No evidence of disease; LR = Local recurrence; DoD = Dead of disease; M = Metastasis; DwoD = Dead without disease; LTF = Lost to follow-up; AwD = Alive with disease

Karyotype: Structural rearrangements of chromosome arms Xq, 6q, and 11q, harboring the *MED12* ,*CITED2*, and *PRDM10* genes, respectively, are indicated in bold type

QRT-PCR 3'/5': Ratio between expression levels of the 3' and 5' parts of *PRDM10* at quantitative real-time PCR analysis

FISH: fluorescence in situ hybridization with BAC probes flanking the *PRDM10* locus. No material was available from Case 1.

RT-PCR: Fusion transcripts detected at RT-PCR for *MED12-PRDM10* and *CITED2-PRDM10*, respectively; Neg = Negative

RNA-Seq: Fusion transcripts detected at transcriptome sequencing; Neg = Negative

Karyotype previously published: Previously published karyotypes are indicated with reference and case numbers in the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (http://cgap.nci.nih.gov/Chromosomes/Mitelman)

Supplementary Table S2. Primers Used for Gene Fusion Verification and Sequencing

Gene	Primer	Location	Sequence	Reference sequence
MED12	MED12-6259-F[a,c]	Exon 42	CTGCAGCAGACACCCATGAT	**NM_005120.2**
	MED12-5725-F	Exon 38	CACCCAGAACCAGCCACTAC	**NM_005120.2**
CITED2	CITED2-883-F[b]	Exon 2	CTGCCGCCCAATGTCATAGA	**NM_006079.4**
	CITED2-401-F	Exon 2	TGGGCGAGCACATACACTAC	**NM_006079.4**
PRDM10	PRDM10-1904-R[a]	Exon 13	GAGATCACAGGTCAGTGGGC	**NM_020228.2**
	PRDM10-2203-R[c]	Exon 14	AGCATGTGGAGTCGCAGTTT	**NM_020228.2**
	PRDM10-2391-R[b]	Exon 15	ACGTGAAGCTGTCGTAGTCTG	**NM_020228.2**
	PRDM10-3130-R	Exon 19	GCTGAGGATCATGGAGCTGG	**NM_020228.2**

[a] Primers that were used for sequencing of the *MED12-PRDM10* fusion in Case 1. [b] Primers that were used for sequencing of the the *CITED2-PRDM10* fusion in Case 2. [c] Primers that were used for sequencing of the the *MED12-PRDM10* fusion in Case 27.

Article III

Article IV

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-018-06098-0 OPEN

Different patterns of clonal evolution among different sarcoma subtypes followed for up to 25 years

Jakob Hofvander[1], Björn Viklund[2], Anders Isaksson[2], Otte Brosjö[3], Fredrik Vult von Steyern[4], Pehr Rissler[5], Nils Mandahl[1] & Fredrik Mertens[1,5]

To compare clonal evolution in tumors arising through different mechanisms, we selected three types of sarcoma—amplicon-driven well-differentiated liposarcoma (WDLS), gene fusion-driven myxoid liposarcoma (MLS), and sarcomas with complex genomes (CXS)—and assessed the dynamics of chromosome and nucleotide level mutations by cytogenetics, SNP array analysis and whole-exome sequencing. Here we show that the extensive single-cell variation in WDLS has minor impact on clonal key amplicons in chromosome 12. In addition, only a few of the single nucleotide variants in WDLS were present in more than one lesion, suggesting that such mutations are of little significance in tumor development. MLS displays few mutations other than the *FUS-DDIT3* fusion, and the primary tumor is genetically sometimes much more complex than its relapses, whereas CXS in general shows a gradual increase of both nucleotide- and chromosome-level mutations, similar to what has been described in carcinomas.

[1] Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, SE-221 84 Lund, Sweden. [2] Science for Life Laboratory, Department of Medical Sciences, Uppsala University, SE-751 23 Uppsala, Sweden. [3] Department of Orthopedics, Karolinska Hospital, SE-171 76 Stockholm, Sweden. [4] Department of Orthopedics, Clinical Sciences, Lund University and Skåne University Hospital, SE-221 85 Lund, Sweden. [5] Department of Clinical Genetics and Pathology, University and Regional Laboratories Region Skåne, SE-221 85 Lund, Sweden. Correspondence and requests for materials should be addressed to J.H. (email: jakob.hofvander@med.lu.se)

Genetic instability is considered an obligate feature of cancer cells[1–3]. This assumption is based on theoretical considerations as well as on extensive observations in tumors and experimental systems. Neoplastic transformation is thought to require more mutations than can be expected to arise from "normal" mutation rates and neoplasms consistently harbor, often numerous, somatic mutations. Furthermore, many neoplasms show extensive intratumoral heterogeneity with regard to mutations and clonal evolution is frequently observed in tumors that are repeatedly sampled during disease progression[4–8]. However, most of the conclusions have been drawn from data on highly malignant epithelial neoplasms in adults, which may develop through mechanisms that differ from other solid tumors or hematopoietic malignancies, and data on neoplasms that have been followed for many years are scarce. Finally, while it is a well-established fact that different tumor types show different mutational profiles and that nucleotide level mutations predominate over chromosomal rearrangements in some tumors and vice versa in others[9,10], it remains poorly investigated to what extent these factors affect clonal evolution.

In this context, sarcomas constitute an interesting group of malignancies. Sarcomas are clinically and genetically heterogeneous and can be arbitrarily subdivided into three main subgroups on the basis of their defining genetic characteristics. One subgroup, comprising about 25% of the entities, is characterized by specific gene fusions, which are thought to function as master switches of transcriptional programs; these sarcomas range in clinical behavior from relatively benign to highly malignant[11]. A second subgroup displays supernumerary ring chromosomes, containing amplified material from large genomic segments; these tumors are typically low-grade malignant and display lipoblastic differentiation, but have the, for sarcomas, unusual potential to progress from low-grade to high-grade malignant lesions[12]. The third and largest subgroup shows various and often extensive combinations of genomic imbalances and point mutations, none of which is specific for any given tumor type; these sarcomas are typically medium-high grade malignant[13]. Few studies on genetic instability and clonal evolution have been performed on sarcomas[14–17], and to our knowledge no attempt has been made to compare patterns of clonal evolution in different genetic subgroups.

In order to assess the type and rate of clonal evolution in different pathogenetic subgroups of sarcoma, we selected the two most common subtypes of liposarcoma: well-differentiated liposarcoma (WDLS, aka atypical lipomatous tumor) and myxoid liposarcoma (MLS). WDLS displays supernumerary ring chromosomes containing amplified material from multiple genomic segments, always including substantial portions of chromosome arm 12q[12]. Extensive inter-cellular genetic variation caused by mitotic instability of the ring chromosomes has been demonstrated[18]. MLS is gene fusion-driven—most cases display a *FUS-DDIT3* chimera, which is considered a strong driver mutation[19]. For comparison, sarcomas representing the third genetic subgroup, with complex genomes (CXS), were included. We also studied multiple samples from some of the primary lesions, in order to evaluate intra-lesional heterogeneity. We show that the extensive single-cell variation in WDLS has little impact on key amplicons in chromosome 12, that MLS displays few mutations other than the *FUS-DDIT3* fusion, and that CXS in general shows a gradual increase of both nucleotide- and chromosome–level mutations.

Results

Amplicon-driven well-differentiated liposarcomas. From five patients with WDLS both chromosome and nucleotide level data were available from 20 samples from 12 lesions. Time interval between first and last sampling was 57–306 months (Table 1). All successfully analyzed samples showed composite karyotypes, united by one or more supernumerary ring chromosomes. The inter-cellular variation was extensive: the number and/or size of ring chromosomes varied considerably (Supplementary Fig. 1), and there were numerical as well as structural non-clonal changes; the latter were found in 42% of the cells (Supplementary Table 1). Neither SNP array analyses nor WES reflected this extensive variation (Fig. 1; Supplementary Fig. 2). When comparing three different samples from the same primary tumor (PT) in four cases, no differences were found (Supplementary Data 1). The 12 lesions showed 22–51 (median 35) GCS at SNP array, almost all of which were gains. When comparing any two lesions from the same patient, 25–83% (median 49%) of the breakpoints were shared (Supplementary Data 2), and the median overlap was 0.57 when the total extension of GCS was compared (Supplementary Table 2). The amplified sequences in chromosome 12, including genes, such as *MDM2*, *HMGA2*, and *CDK4*, displayed greater overlap among different lesions from the same patient both with regard to shared breakpoints (range 31%–89%, median 65%; Supplementary Data 2) and to the extension of GCS (range 0.53–0.99, median 0.71; Fig. 2; Supplementary Table 2). Each WDLS sample had few ESV (range 1–11, median 7), at low allele frequencies (median 21%). Intra-lesional heterogeneity was low, with 82–100% of ESVs being present in all three samples analyzed. With time, however, most mutations were unique for each lesion: only 3/72 mutations that were detected were shared with another lesion (Supplementary Data 3). At relapse, the number of GCS did not increase, and ESV only moderately so (Table 1), and there was no indication that the samples became less similar with time. Actually, both cases from which three samples could be analyzed showed greater similarity with regard to GCS on chromosome 12 between the first and last samples (0.97 and 0.99, respectively) than between the first and second or second and third samples (0.69–0.72).

Gene fusion-driven myxoid liposarcomas. From nine *FUS-DDIT3*-positive MLS, the PT and 1–4 local recurrences (LR) and/or metastases (Met), occurring 12–104 months after diagnosis, were studied (Table 1). The inter-cellular variation at G-banding was small: among the 317 cells from the 15 samples that could be assessed, only 4 cells (1.3%) showed non-clonal structural aberrations and 4 (1.3%) deviated from the stemline chromosome number (Supplementary Table 1); clonal karyotypes were consistently identical when comparing 2–3 samples from the same PT (Supplementary Data 4). Combining cytogenetic and SNP array data, 1–6 chromosome level aberrations were found per PT and there were few differences (range 0–8, median 1) between a PT and its LR or Met. Two LR (cases 1 and 6) had fewer chromosome aberrations than their PT and 6/13 Met had the same number of chromosome level aberrations as the PT (Fig. 1; Supplementary Data 2; Supplementary Fig. 2). WES data on 11 samples from four patients showed 7–165 (median 15.5) ESV per PT. In MLS 1–3, most (61–100%) ESV detected in a PT were present also at relapse, but Case 4 showed a dramatic decrease. That PT had 165 ESV; the large number of ESV was confirmed at independent WES and targeted re-sequencing. Its four Met, occurring 19–74 months after diagnosis, had only 11–24 ESV. However, the clonal relationship between the PT and the Met was unquestionable, with six ESV being shared by all samples. In addition, they all shared the six chromosome level aberrations seen in the PT, with only 1–2 new aberrations per Met (Fig. 1; Supplementary Tables 3 and 4; Supplementary Fig. 2).

Sarcomas with complex genomic aberrations. For comparison with gene fusion- and amplicon-driven liposarcomas, we

Table 1 Longitudinal genetic study of three different pathogenetic subgroups of sarcoma

Case no.[a]	Material[b]	Dx[c]	ESV[d]	GCS[e]	G-band[f]
1A	PT	MLS	7	4	5
1B	LR2 (22)		9	4	4
2A	PT	MLS	18	0	1
2B	Met1 (30)		20	0	1
2C	Met2 (92)		ND	ND	2
2D	Met3 (98)		ND	ND	2
3A	PT	MLS	12	0	1
3B	Met1 (19)		23	8	1
3C	Met2a (28)		ND	ND	1
3D	Met2b (28)		ND	ND	1
3E	Met3 (32)		ND	ND	1
4A	PT	MLS	165	2	5
4B	Met1 (19)		11	3	5
4C	Met2a (20)		16	3	5
4D	Met2b (20)		14	3	5
4E	Met3 (74)		24	4	5
5A	PT	MLS	ND	0	1
5B	LR1 (104)		ND	1	1
6A	PT	MLS	ND	5	3
6B	LR1 (12)		ND	0	1
7A	PT	MLS	ND	1	1
7B	LR1 (25)		ND	1	1
8A	PT	MLS	ND	ND	4
8B	Met1 (48)		ND	ND	4
9A	PT	MLS	ND	ND	1
9B	Met1 (42)		ND	ND	1
10A	PT	WDLS	5	38	8
10B	LR1 (197)		1	35	3
10C	LR2 (306)		8	37	2
11A	PT	WDLS	ND	ND	ND
11B	LR1 (84)		4	35	1
11C	LR2 (124)		7	22	>10
11D	LR3 (141)		9	35	6
12A	PT	WDLS	6	38	4
12B	LR1 (124)		4	51	>3
13A	PT	WDLS	11	27	11
13B	LR2 (215)		7	23	4
14A	PT	WDLS	7	35	3
14B	LR1 (211)		8	39	5
15A	PT	MFS	32	91	
15B	LR1 (47)		46	99	
15C	Met1 (114)		68	84	
16A	PT	MFS	ND	ND	
16B	LR1 (60)		29	33	
16C	LR2 (100)		31	22	
16D	LR6 (152)		33	28	
17A	PT	MFS	5	151	
17B	Met3 (77)		19	114	
18A	PT	MFS	15	27	
18B	LR2 (110)		13	30	
19A	PT	MFS	25	80	
19B	Met1 (86)		27	141	
20A	PT	Myoep	10	110	
20B	LR2 (294)		9	99	

The highly complex karyotypes in CXS tumors precluded any attempt to calculate the number of aberrations
[a]Multiple samples were analyzed from the primary tumors of cases 5, 7, 10,12–14, and 18–20. The figures for each case denote the combined number of changes in all samples
[b]PT = primary tumor, LR = local recurrence, Met = metastasis. Time in months from diagnosis is indicated in parentheses
[c]Diagnosis. MLS = myxoid liposarcoma, WDLS = well-differentiated liposarcoma, MFS = myxofibrosarcoma, Myoep = myoepithelial tumor
[d]ESV = No. of non-synonymous exonic variants detected at whole-exome sequencing. ND = not done. The values for the PT for which >1 sample was analysed represent the median for all samples
[e]GCS = No. of chromosomal imbalances detected at SNP array analysis
[f]No. of clonal chromosome aberrations detected at G-banding analysis. ND = not done

investigated 6 CXS with 2–3 lesions per case and an interval of 77–294 months between first and last sampling (Table 1). In two cases, 2 or 3 samples from the PT could be analyzed with regard to intra-lesional heterogeneity using both SNP array and WES; in case 18, no differences were seen between the samples, whereas one of the three samples in case 20 had 7 additional imbalances at SNP array analysis (Supplementary Data 1). In most cases the clonal aberrations detected at banding analysis could only be partly resolved and were hence not sufficiently informative for comparisons between samples, but three samples from the PT of Case 19 were analyzed cytogenetically, showing extensive variation, including a ploidy shift in one sample, in clonal aberrations (Supplementary Data 4). SNP array analysis identified 22–151 (median 87.5) GCS per sample, and the fraction of shared breakpoints in samples from the same patient was 6–83% (median 42%). The median overlap of GCS was 0.58 (range 0.24–0.93). The number of ESV per sample (median 26, range 5–68) was higher for CXS than for liposarcomas (median 7 in WDLS and 16 in MLS), and in all but one patient there was a steady increase with time (Table 1; Supplementary Data 3).

Discussion

Studies of genetic variation and its role for clonal evolution in tumor cell populations face several problems. For example, the initial driving force(s) for neoplastic transformation may provide different prerequisites for which routes are available and what is needed to sustain and optimize continued proliferation. In the present study, we evaluated the type and degree of stemline variation in multiple lesions from liposarcomas—amplicon-driven WDLS and gene fusion-driven MLS—and other sarcomas characterized by complex genomic rearrangements (CXS) that had been followed for long time periods. Apart from this longitudinal aspect of clonal heterogeneity, we could study intra-lesional heterogeneity at the genome and nucleotide levels in four WDLS and two CXS, as well as inter-cellular (single cell) variation at the chromosome level in all WDLS and 15 MLS lesions. A caveat of the present study is, of course, that the patients were selected on the basis of having late relapses, and it cannot be excluded that rapidly relapsing sarcomas would have yielded different results. Still, the cohort that was analyzed constitutes a rare selection of solid tumors followed for exceptional time periods, and the data provide some interesting clues to the longitudinal clonal dynamics in sarcomas.

WDLS, driven by amplification of parts of chromosome 12, with *MDM2*, *CDK4*, and *HMGA2* as the most important targets[12], displays great inter-cellular variation at the chromosome level, as shown in the present study (Supplementary Table 1). In spite of follow-up periods for up to 25 years, this variation had, however, a minor impact on the composition of the tumor stemlines. Furthermore, the small number of mutations, the low allele frequencies, the small number of shared mutations among different lesions from the same patient, and the absence of mutations shared by different patients all strongly imply that ESV are of little or no significance in WDLS development. Of the 70 genes that displayed mutations, only 8 are included in COSMIC's Cancer Gene Census (https://cancer.sanger.ac.uk/census), and none of these mutations has been reported before in soft tissue tumors. The low frequency or absence of ESV that were shared by all lesions from a patient also suggests that WDLS either develop early in life or that the progenitor cell has undergone far fewer cell divisions before neoplastic transformation than a typical precursor cell in a carcinoma. Clonal dynamics in WDLS instead concern larger copy number changes. The results of the present study show that the genotype in WDLS fluctuates around a set of core amplicons in chromosome 12. The only region amplified in

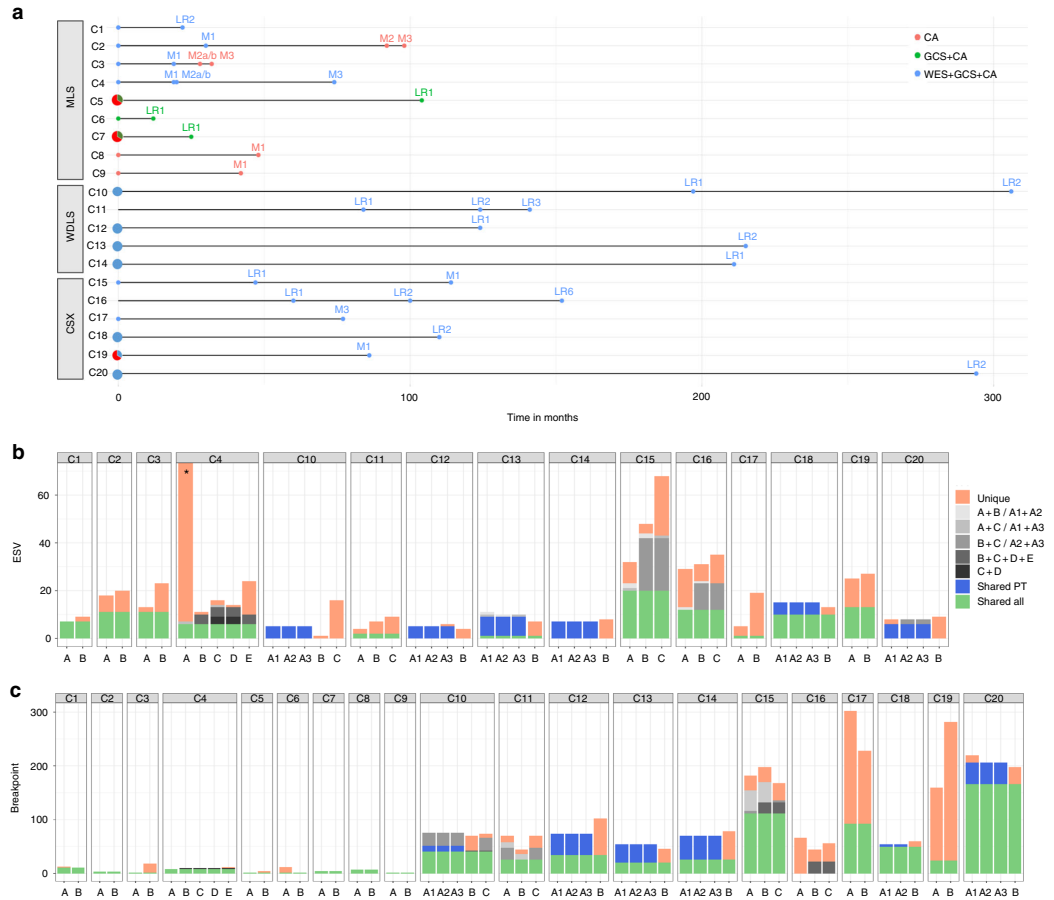


Fig. 1 Schematic illustration of clonal evolution in 20 sarcomas (C1–C20). C1–C9 are gene fusion-driven myxoid liposarcomas (MLS), C10–C14 are amplicon-driven well-differentiated liposarcomas (WDLS), and C15–C20 are sarcomas with complex genotypes (CXS). **a** Time intervals (in months) between lesions that were analyzed with regard to chromosomal aberrations and nucleotide level mutations. Each sample is indicated by a filled circle; blue samples were analyzed by whole-exome sequencing (WES), SNP arrays (GCS), and chromosome banding analysis (CA), green samples by GCS and CA, and red samples only by CA; larger filled circles represent lesions from which multiple samples were analyzed for assessment of intratumoral heterogeneity. Each line starts with the primary tumor, followed by local recurrences (LR) and/or metastases (M). **b** Diagram showing the number of non-synonymous exonic variants (ESV) detected at WES, as well as the extent of shared mutations among different samples and lesions from the same patient. **c** Diagram showing the number of clonal chromosomal breakpoints detected at GCS and, for MLS also including CB, as well as the extent of shared aberrations among different samples and lesions from the same patient. Figures for C8 and C9 are based on CB only

all 12 samples was a discontinuous 856 Kb sequence in 12q14–15, including six functional genes (Supplementary Data 5), suggesting that at least some of them, notably *MDM2* and the first three exons of *HMGA2*, are essential for tumorigenesis, in line with previous data[12]. Bearing in mind the mitotic instability of ring chromosomes it is highly surprising not only that the follow-up samples were so similar to the first sample, but also that the total extension of chromosome 12 amplification remained 20 times larger than the minimal shared region of amplification. For instance, in Case 10, the total length of the amplified material from chromosome 12 was 19.1 Mb in the PT, 20.7 Mb in the 19 cm large LR1 16 years later, and 19.6 Mb in the 20 cm LR2 another 9 years later (Supplementary Data 2). These results are in line with the suggestions by Lloyd et al. that tumors, as long as their microenvironment remains stable, relatively early might reach a genetic fitness maximum[20]; additional mutations occur but are not selected for or even deleterious, and transient clones could be attributed to genetic drift facilitated by the bottlenecks caused by the surgical excisions. Indeed, all WDLS follow-up samples analyzed were LR, arising in the same location as the PT and none of the patients had received any chemotherapy that could have shifted the selection pressure.

In MLS, expression of *FUS/DDIT3* has been shown to be sufficient for neoplastic transformation in various experimental models[18], which is in agreement with cytogenetic and sequencing data showing that there are few recurrent chromosomal imbalances, notably trisomy 8 and idic(7)(p11), or exonic SNVs, none of which is consistent[21,22]; the only frequent secondary mutation

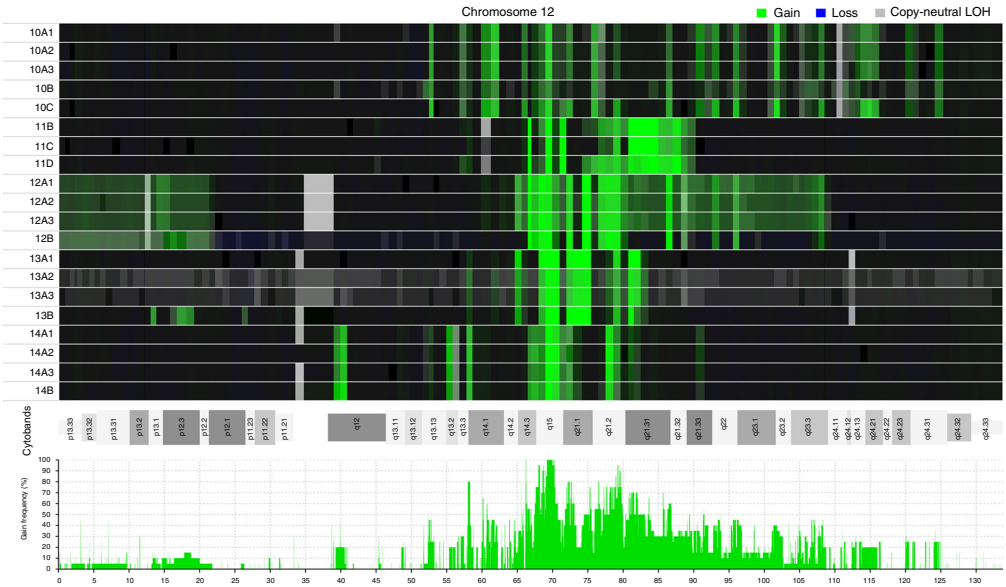


Fig. 2 Heat map and frequency distribution of amplicons in chromosome 12. Twenty samples from 12 lesions from five patients with well-differentiated liposarcomas, representing amplicon-driven sarcomas, were analyzed. **a** The upper panel, based on the log ratios, shows that the extension of gains (green) and copy-neutral loss of heterozygosity (LOH) is highly similar among different samples from the same patient. Note that samples 10A1–A3 (three samples from primary tumor) and 10C (local recurrence 2; LR2) are more similar to each other than 10A1–3 to 10B or 10B to 10C; the same is true for samples 11B (LR1) and 11D (LR3) in comparison to 11C (LR2). **b** The lower panel, based on the copy number segmentation, shows the frequency of distinct amplicons in chromosome 12 among the 20 samples. Only two segments in 12q14–15, with a combined length of 856 kb, were amplified in all samples

identified so far affects the promoter region of the *TERT* gene, which is seen in some 70–90% of the tumors[23,24]. Despite the relative lack of secondary mutations, the clinical behavior of MLS varies substantially. Some 35% of the patients develop metastases and it has been suggested that certain mutations, e.g., in *PIK3CA* and *TP53*, are associated with aggressive behavior[25–27]. Our results show that clonal evolution in MLS is usually very slow at the chromosome level with few deviations from the stemline, even in metastatic lesions. Less than 5% of the cells showed non-clonal aberrations at G-banding analysis and only 1/4 LR and 7/13 Met showed chromosomal aberrations, as assessed by cytogenetics and/or SNP array, which deviated from the set of shared aberrations (Fig. 1; Supplementary Table 1). Admittedly, 7/13 metastases could only be analyzed by G-banding, but also the six metastases analyzed by high-resolution SNP array showed few (0–8, median 1) additional imbalances compared to the mutational trunk. In contrast, there was a more pronounced accumulation of ESV among the four cases that could be analyzed also by WES, and as expected the relapse samples more often had more ESV than the PT, with the PT of Case 4 as an extreme exception (Fig. 1). That PT had 165 ESV, including in well-known cancer-associated genes such as *BCOR*, *CHEK2*, and *TP53* that have also been implicated in MLS progression[22,25,27]. Only six ESV were shared by all samples, and these occurred at allele frequencies around 5–10% in the PT. Thus, the cell population that gave rise to all metastases had been replaced by a subclone with a much higher level of nucleotide level instability; the allele frequencies of *CHEK2* (54–68%) and *TP53* (36–43%) mutations in this subclone suggest that they occurred early and may have triggered the massive accumulation of ESV.

Case 4 notwithstanding, the results show that MLS cells are genetically relatively stable, and that clonal evolution in MLS is mainly driven by nucleotide level mutations. The slow accrual of new mutations, or even reduction of genetic complexity, in MLS with time and tumor progression has several important implications. First, as already suggested by Reiter et al., cells that eventually form metastases may arise relatively early in the primary tumor; studying pancreatic carcinomas, they showed that metastases share most if not all important driver mutations with their PT[28]. Second, although we cannot exclude an impact of mutations in non-coding sequences, much of the morphological and clinical variation in MLS, such as the transition from a low-grade to a high-grade tumor in cases 3, 6, and 9, could be caused by epigenetic factors. Furthermore, the findings in case 4 demonstrate that therapeutic decisions based on genetic findings in a single sample may not be relevant for all tumor sites. In contrast to the more common notion that mutations in small subclones of a PT might be overlooked, the present case demonstrates that analysis of the PT might suggest therapeutic targets that are not present in the metastatic lesions.

The CXS group of sarcomas was included for comparison with the amplicon-driven WDLS and gene fusion-driven MLS. While the pathogenetic mechanisms in CXS sarcomas still remain relatively poorly investigated, it is well known that there exists an extensive genetic and clinical variation not only among subtypes but also within morphologic subgroups[8,13,29]. In general, our findings in CXS were in good agreement with recent comprehensive genetic data on sarcomas in adults[30]. That study showed that myxofibrosarcomas, which was the most common CXS subtype studied here, have complex copy number changes but few

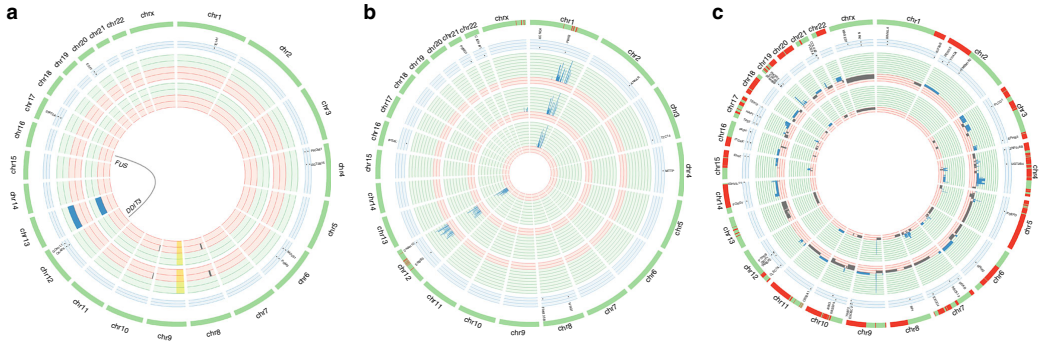


Fig. 3 Circos plots illustrating different modes of clonal evolution in sarcomas with different genetic backgrounds. **a** A fusion-driven myxoid liposarcoma (MLS), **b** an amplicon-driven well-differentiated liposarcoma (WDLS), and **c** a myxofibrosarcoma (MFS) with a complexly rearranged genome. The red/green inner circles represent the location and amplitude of the allelic imbalances; blue is gain, gray is loss and yellow background indicates loss of heterozygosity. The number of red fields can vary between lesions depending on what is considered the expected number of copies for that lesion in relation to the ploidy level (2*n*–3*n*), the number of green fields varies between patients and is determined by the gain with highest number of copies in that patient. The circles are ordered chronologically, starting from the center with the first lesion. The light blue circles represent the location of the variants reported by the whole-exome sequencing (WES) in the same order. Red on the schematic green chromosomes represents differences in genomic changes at SNP array (GCS) between lesions. Both the primary tumor (PT) and the local recurrence (LR) 22 months later from the MLS (case 1) displayed few and mostly identical GCS and ESV. In three lesions from a WDLS (case 10), the GCS of the PT were more similar to those in the second LR occurring after 306 months than to those in the first LR occurring after 197 months. The MFS (case 19) had no less than 209 GCS, but only 39 ESV. While many of the ESV were shared by the PT and the metastasis occurring 86 months later, the GCS overlap was only 0.39. Circos plots for all 20 sarcomas analyzed by both SNP array and WES are shown in Supplementary Fig. 2

significant single nucleotide variants. Indeed, of the 87 mutations that were shared by at least two lesions from the same patient in the present study, only 4 are included in COSMIC's Cancer Gene Census (https://cancer.sanger.ac.uk/census): *EGF*, *IDH2*, *PTPRB*, and *TP53*. Of these, only *TP53* mutations have been implicated in sarcoma development before.

Although, the CXS samples had, on average, higher GCS (median 87.5 compared to 33 in WDLS and 2 in MLS) and ESV (median 26 compared to 7 in WDLS and 16 in MLS) levels than liposarcomas the CXS analyzed here were highly heterogeneous, both with regard to rate and type of clonal evolution. For instance, case 20 showed only 7 ESV in the PT and 11 in the LR obtained 24.5 years later, none of which was shared, but the GCS overlap was 0.79. In contrast, LR1 of case 16 had 29 ESV while LR6 (8 years later) had 35, 12 of which were shared with LR1; at the same time, there were massive changes at the chromosome level, with a GCS overlap of only 0.24 in the two samples (Supplementary Data 2). Thus, more cases of CXS, including other morphologic subtypes than myxofibrosarcoma and myoepithelial tumors, need to be analyzed to draw any firm conclusions on the longitudinal clonal dynamics in these malignancies.

Although it is known that local relapse in sarcoma patients is associated with an increased risk for distant spreading it has been debated whether this should be explained by inherent differences in aggressiveness, i.e., some sarcomas have a higher risk for both local and distant relapse, or whether some locally relapsed tumors actually beget metastases[31]. In the present study, there was only one patient (case 15) with data on both types of relapse: an LR after 47 months and a Met after 114 months. While the PT, LR, and Met all displayed highly complex, incomplete karyotypes, the GCS overlap was higher for the LR-Met comparison (0.73) than for the PT-Met comparison (0.62). In addition, out of 68 ESV occurring at frequencies >5% in the Met, only 1 was uniquely shared with the PT while 20 were uniquely shared with the LR. Thus, the molecular data strongly argue for the LR begetting the Met in this particular case.

In conclusion, the present study shows that the rate by which new mutations become predominant and that the type of clonal evolution, i.e., whether nucleotide or chromosome level mutations prevail, vary considerably among sarcomas caused by different pathogenetic mechanisms (Fig. 3). It also demonstrates, as exemplified by WDLS, that marked genetic instability, i.e., great variation at the single cell level, does not necessarily translate into major changes in the tumor stemline. Whereas, the development of new mutations at the chromosome and nucleotide levels in many CXS fit well with data on carcinomas, both types of liposarcoma displayed a remarkable paucity of clonal evolution at the DNA level. This scenario is similar to what has been suggested for some pediatric tumors and leukemias[32,33], but it should be pointed out that all liposarcoma patients were adults (39–77 yrs). Thus, in some sarcomas the genetic alterations needed for metastatic seeding are present well before the diagnosis of the primary tumor suggesting that they obtain a genetic fitness maximum early in tumor development. As sarcomas are highly heterogeneous from a biological point of view it remains to be investigated whether also other subtypes display similar patterns of clonal evolution. Furthermore, the slow accumulation of DNA level mutations in some sarcomas does not exclude that epigenetic changes could be important in tumor progression.

Methods

Tumors. To assess type and rate of clonal evolution in sarcomas with different pathogenetic mechanisms, we selected patients from which more than one lesion –PT, LR, and/or Met—had been analyzed, and in which at least 1 year had elapsed between first and last sampling. Information on tumors, samples, and analyses performed are given in Table 1 and Fig. 1, and in more detail in Supplementary Data 4. We then combined data from chromosome banding, high-resolution SNP array, and whole-exome sequencing (WES) analyses to assess the spectrum and distribution of genetic aberrations that may develop with time. The study included 20 sarcoma patients from which 2–5 lesions had been obtained with 12–306 months between first and last sampling. Five patients had WDLS, representing amplicon-driven sarcomas, nine had MLS, representing gene fusion-driven sarcomas, and six had myxofibrosarcoma (MFS, $n = 5$) or myoepithelial tumor ($n = 1$), representing CXS. Due to the retrospective, longitudinal nature of the

study, with tumors dating back to the early 1980s, only one sample was available from most lesions. However, in nine cases, 2–3 samples from the PT could be studied separately, allowing us to correlate the longitudinal variation with intra-tumoral heterogeneity at the chromosome and/or the nucleotide level. Tumors were diagnosed according to established criteria[29], and the *FUS-DDIT3* fusion transcript in MLS was detected using standard RT-PCR protocols[34]. Gene fusions in Case 20 were excluded through transcriptome sequencing, using previously described methods[35]. Samples were obtained after informed consent and the study was approved by the local review board (diary number 2017/796).

Chromosome banding and SNP array analysis. Chromosome preparations were made from short-term cultured cells obtained from disaggregated tumor tissue from 54 samples from 20 patients and stained for G-banding as previously described[36]. SNP array analysis was performed as described[37]. In brief, tumor DNA was extracted from fresh frozen tumor tissue from 54 samples from 43 lesions from 18 patients and analyzed using the Affymetrix CytoScan HD array (Affymetrix, Santa Clara, CA, USA), containing more than 2.6 million markers, or the Illumina HumanOmni1-Quad Genotyping BeadChip (Illumina Inc, San Diego, CA), containing 1.2 million markers. Genomic aberrations were identified by visual inspection using the Chromosome Analysis Suite version 1.2 (Affymetrix) or the GenomeStudio Data Analysis Software (Illumina) combined with bioinformatic analysis regarding copy numbers and segmentation using Rawcopy and the Tumor Aberration Prediction Suite (TAPS)[38,39]. For calculations of intra- and inter-lesional heterogeneity in WDLS and CXS only the genomic changes detected at SNP array analysis (GCS), here including copy-neutral loss of heterozygosity, that extended >500 kb were included. Breakpoints were considered shared when the copy number shift or copy-neutral loss of heterozygosity occurred between the same two probes in two or more samples. For MLS, G-banding and SNP array data were combined to calculate the number of chromosome level aberrations. The human reference sequence used for alignment was the GRCh37/hg19 assembly. Constitutional copy number variations were excluded through comparison with the Database of Genomic Variants (http://projects.tcag.ca/variation/).

Jaccard index. The Jaccard index was used to measure the similarity within and between different lesions based on the overlap of their GCS. The index is calculated by taking the ratio of the number of overlapping base pairs between two samples and the length of the union; the union is the length of the GCS in both samples minus the number of overlapping bases. The value of the index can range from 0 to 1, where 0 represents no overlap and 1 represents complete overlap. The Jaccard index was calculated on the genomic intervals listed in Supplementary Table 2 using bedtools (v2.26.0) with the jaccard subcommand.

Whole-exome sequencing (WES). DNA was extracted from fresh frozen tumor biopsies as described[40]. Whole-exome libraries were prepared from a total of 49 tumor samples from 37 lesions and 15 blood samples from 15 patients using the Nextera Rapid Caputre Exome Kit (Illumina) according to the manufacturer's recommendations. Paired 2 × 76 bp or 2 × 151 bp reads were generated from the exome libraries using a NextSeq 500 (Illumina). First, remaining adapter sequences were removed from the FASTQ files using Trim-galore (v0.4.1). The trimmed reads were aligned to the human reference genome hg19 using BWA-MEM (v0.7.10). Duplicate reads where marked using Picard (v2.2.4) and the BAM files were further processed using GATK (v3.5) according to the best practice pipeline for tumor-normal pairs. Somatic SNVs were called using MuTect[41] (v1.1.7) with default settings and somatic indels were detected using Strelka[42] (v1.0.15) with default settings. Variants were annotated using VEP[43]. The WES generated an average coverage of ×98 of the target bases. The total number of somatic SNVs and indels among all samples were 16,968 and 1428, respectively. In order to enrich for true somatic missense mutations, and limit the number of sequencing artifacts known to be generated by WES, variants were further filtered as follows: read depth of ≥20 in tumor and ≥10 in corresponding normal sample, average base quality ≥20, mutated allele frequency (MAF) of ≥10% in tumor and <1% in the normal sample, and only non-synonymous exonic somatic variants (ESV) were kept. In addition, each variant was visually inspected using IGV (http://software.broadinstitute.org/software/igv/) and a minimum of 2 reads in each orientation was demanded. However, if the same ESV was present in more than one sample from the same patient, only one of the ESV had to fulfill the above criteria and the additional identical ESV needed only 3 reads to be counted. After filtering, a total of 564/861 (unique/total) ESV were retained. An additional"non-somatic" variant caller, Freebayes (v1.0.1) (https://github.com/ekg/freebayes), was run on all samples using the list of filtered variants as targets. This was done for additional verification but primarily to acquire read depth information for positions where no ESV had been reported. The pathogenetic relevance of detected ESV was evaluated with Polyphen (http://genetics.bwh.harvard.edu/pph2/) and SIFT (http://sift.jcvi.org/) and by assessment of COSMIC's Cancer Gene Census database (https://cancer.sanger.ac.uk/census).

Amplicon sequencing. In order to verify some of the mutations detected at WES, a TruSeq Custom Amplicon (TSCA) panel (Illumina) was designed. Library preparation was performed according to the manufacturer's recommendations using the TruSeq Custom Amplicon Low Input Kit (Illumina). Paired-end 2 × 151 bp reads were generated from the Amplicon libraries using a NextSeq 500 (Illumina). Paired reads were merged using Pear (v0.9.6)[44] and aligned to the human reference genome hg19 using BWA-MEM. SNVs and indels were called on the positions reported from the WES using Freebayes. The TSCA generated an average coverage of ×347 and a total number of 181 variants could be analyzed with sufficient coverage (≥×50). Out of these, 176 (97%) were confirmed and four additional variants missed by the WES were detected.

Data availability

The data on which the study is based are presented in full in the Supplementary files. The raw data files that support the findings of this study are available from the corresponding author upon reasonable request. Please note that WES data are available for academic purposes by contacting the corresponding author, as the patient consent does not cover depositing data that can be used for large-scale determination of germline variants.

Received: 9 March 2018 Accepted: 14 August 2018
Published online: 10 September 2018

References

1. Cahill, D. P., Kinzler, K. W., Vogelstein, B. & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends Cell Biol.* **9**, M57–M60 (1999).
2. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
3. Loeb, L. A. Human cancers express a mutator phenotype: hypothesis, origin, and consequences. *Cancer Res.* **76**, 2057–2059 (2016).
4. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194**, 23–28 (1976).
5. Loeb, L. A. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* **51**, 3075–3079 (1991).
6. Shah, S. P. et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
7. Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **501**, 338–345 (2013).
8. Heim, S. & Mitelman, F. *Cancer Cytogenetics* 4th ed. (Wiley Blackwell, Oxford, 2015).
9. Nowak, M. A. et al. The role of chromosomal instability in tumor initiation. *Proc. Natl Acad. Sci. USA* **99**, 16226–16231 (2002).
10. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
11. Mertens, F., Antonescu, C. R. & Mitelman, F. Gene fusions in soft tissue tumors: recurrent and overlapping pathogenetic themes. *Genes Chromosomes Cancer* **55**, 291–310 (2016).
12. Italiano, A. et al. Clinical and biological significance of CDK4 amplification in well-differentiated and dedifferentiated liposarcomas. *Clin. Cancer Res.* **15**, 5696–5703 (2009).
13. Chibon, F. et al. Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. *Nat. Med.* **16**, 781–787 (2010).
14. Örndal, C., Rydholm, A., Willen, H., Mitelman, F. & Mandahl, N. Cytogenetic intratumor heterogeneity in soft tissue tumors. *Cancer Genet. Cytogenet.* **78**, 127–137 (1994).
15. Francis, P. et al. Intratumor versus intertumor heterogeneity in gene expression profiles of soft-tissue sarcomas. *Genes Chromosomes Cancer* **43**, 302–308 (2005).
16. Desai, J. et al. Clonal evolution of resistance to imatinib in patients with metastatic gastrointestinal stromal tumors. *Clin. Cancer Res.* **13**, 5398–5405 (2007).
17. Chen, L. et al. Clonality and evolutionary history of rhabdomyosarcoma. *PLoS Genet.* **11**, e1005075 (2015).
18. Gisselsson, D. et al. Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc. Natl Acad. Sci. USA* **97**, 5357–5362 (2000).
19. Riggi, N. et al. Expression of the FUS-CHOP fusion protein in primary mesenchymal progenitor cells gives rise to a model of myxoid liposarcoma. *Cancer Res.* **66**, 7016–7023 (2006).
20. Lloyd, M. C. et al. Darwinian dynamics of intratumoral heterogeneity: not solely random mutations but also variable environmental selection forces. *Cancer Res.* **76**, 3136–3144 (2016).
21. Mandahl, N. et al. Nonrandom secondary chromosome-aberrations in liposarcomas with t(12;16). *Int. J. Oncol.* **4**, 307–310 (1994).
22. Barretina, J. et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nat. Genet.* **42**, 715–721 (2010).

23. Killela, P. J. et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc. Natl Acad. Sci. USA* **110**, 6021–6026 (2013).
24. Koelsche, C. et al. TERT promoter hotspot mutations are recurrent in myxoid liposarcomas but rare in other soft tissue sarcoma entities. *J. Exp. Clin. Cancer Res.* **33**, 33 (2014).
25. Joseph, C. G. et al. Exomic analysis of myxoid liposarcomas, synovial sarcomas, and osteosarcomas. *Genes Chromosomes Cancer* **53**, 15–24 (2014).
26. Kilpatrick, S. E., Doyon, J., Choong, P. F., Sim, F. H. & Nacimento, A. G. The clinicopathologic spectrum of myxoid and round cell liposarcoma. A study of 95 cases. *Cancer* **77**, 1450–1458 (1996).
27. Oda, Y. et al. Frequent alteration of p16(INK4a)/p14(ARF) and p53 pathways in the round cell component of myxoid/round cell liposarcoma: p53 gene alterations and reduced p14(ARF) expression both correlate with poor prognosis. *J. Pathol.* **207**, 410–421 (2005).
28. Reiter, J. G. et al. Reconstructing metastatic seeding patterns of human cancers. *Nat. Commun.* **8**, 14114 (2017).
29. Fletcher, C. D. M., Bridge, J. A., Hogendoorn, P. C. W. & Mertens, F. *WHO Classification of Tumours of Soft Tissue and Bone* 4th ed. (IARC, Lyon, 2013).
30. The Cancer Genome Atlas Research Network. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* **171**, 950–965 (2017).
31. Trovik, C. S. Local recurrence of soft tissue sarcoma. A Scandinavian Sarcoma Group Project. *Acta Orthop. Scand. Suppl.* **72**, 1–31 (2001).
32. Mullighan, C. G. & Downing, J. R. Global genomic characterization of acute lymphoblastic leukemia. *Semin. Hematol.* **46**, 3–15 (2009).
33. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
34. Panagopoulos, I. et al. Fusion of the EWS and CHOP genes in myxoid liposarcoma. *Oncogene* **12**, 489–494 (1996).
35. Walther, C. et al. A novel SERPINE1-FOSB fusion gene results in transcriptional up-regulation of FOSB in pseudomyogenic haemangioendothelioma. *J. Pathol.* **232**, 534–540 (2014).
36. Mandahl, N. et al. Three major cytogenetic subgroups can be identified among chromosomally abnormal solitary lipomas. *Hum. Genet.* **79**, 203–208 (1988).
37. Walther, C. et al. Genetic heterogeneity in rhabdomyosarcoma revealed by SNP array analysis. *Genes Chromosomes Cancer* **55**, 3–15 (2016).
38. Rasmussen, M. et al. Allele-specific copy number analysis of tumor samples with aneuploidy and tumor heterogeneity. *Genome Biol.* **12**, R108 (2011).
39. Mayrhofer, M., Viklund, B. & Isaksson, A. Rawcopy: improved copy number analysis with Affymetrix arrays. *Sci. Rep.* **6**, 36158 (2016).
40. Nord, K. H. et al. Concomitant deletions of tumor suppressor genes MEN1 and AIP are essential for the pathogenesis of the brown fat tumor hibernoma. *Proc. Natl Acad. Sci. USA* **107**, 21122–21127 (2010).
41. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
42. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
43. McLaren, W. et al. The Ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
44. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).

Acknowledgements

The technical assistance of Jenny Nilsson is gratefully acknowledged. This study was supported by the Swedish Cancer Society (CAN2017/269) and Region Skåne (ALFS-KANE-430191).

Author contributions

J.H., N.M. and F.M. designed research; J.H., B.V., A.I., N.M. and F.M. performed research; O.B., F.V.v.S. and P.R. provided clinical and histopathological data; and all authors assisted with drafting and revising the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-06098-0.

Competing interests: The authors declare no competing interests.

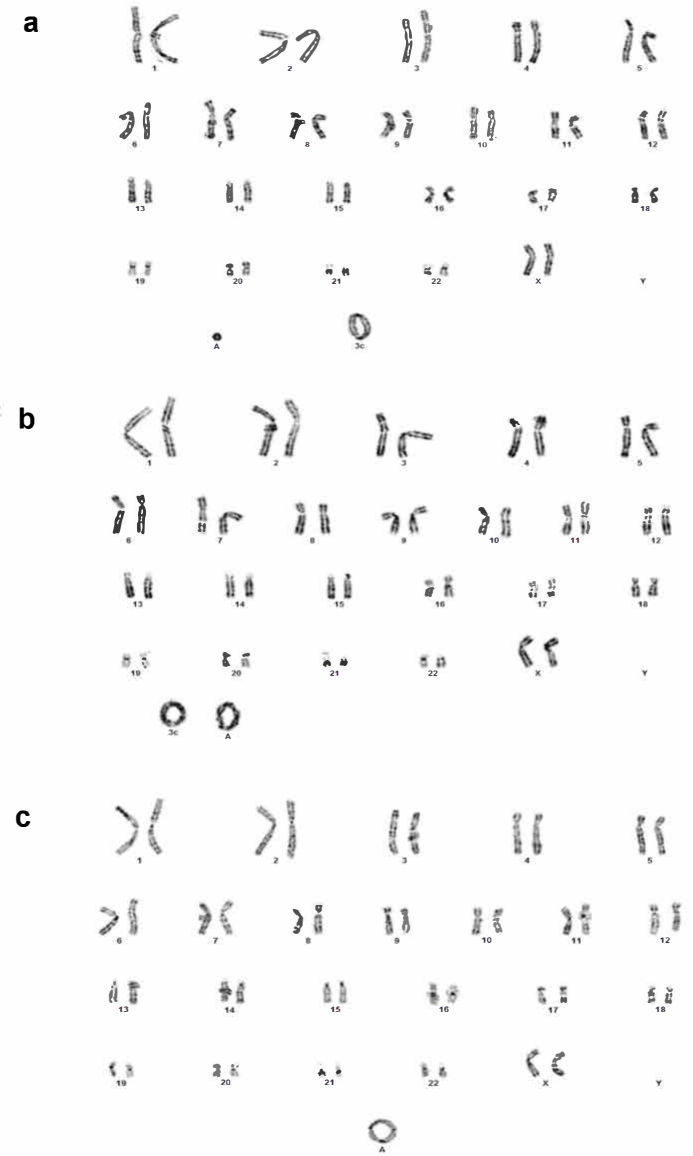
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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

Supplementary Information

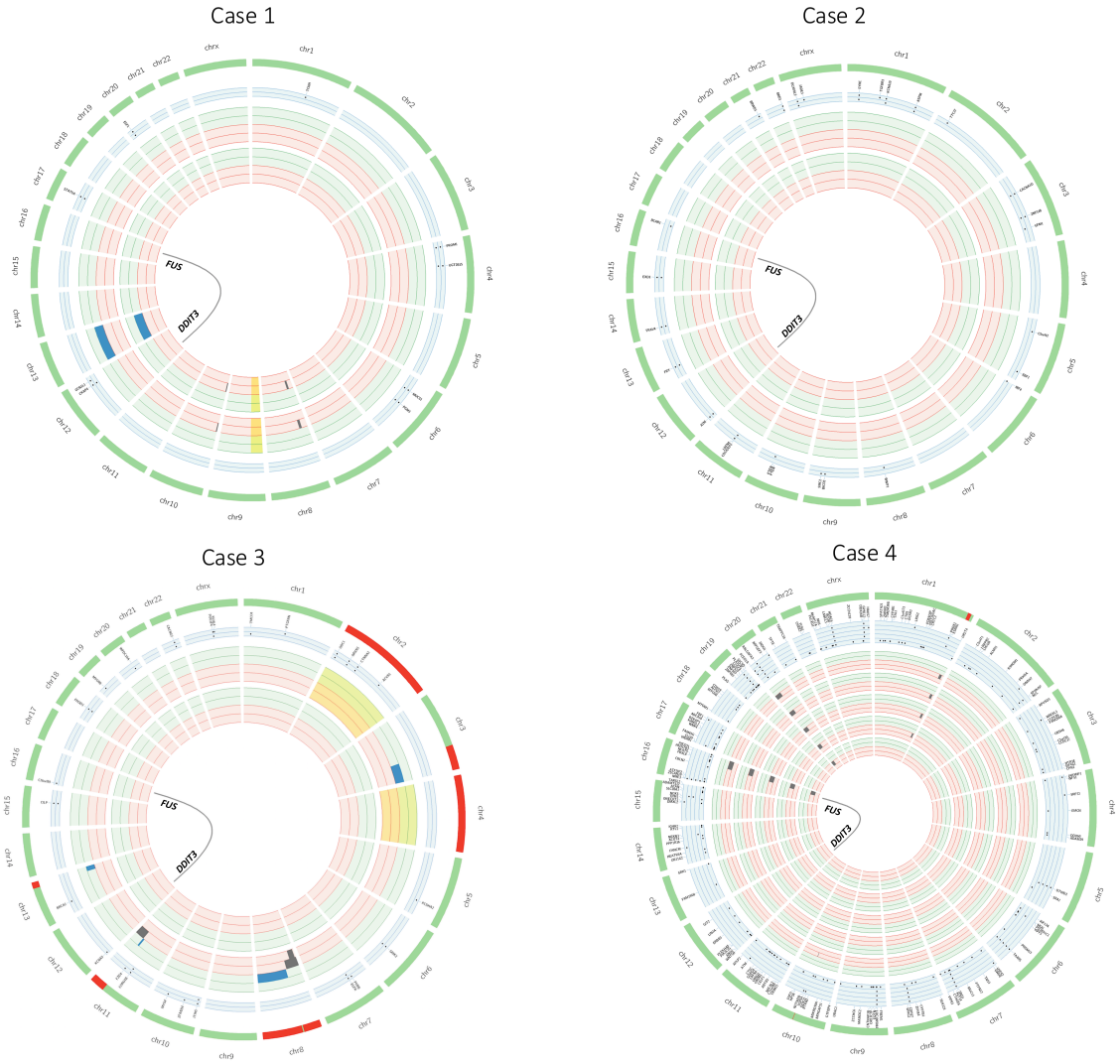
Different patterns of clonal evolution among different sarcoma subtypes followed for up to 25 years

Hofvander et al.



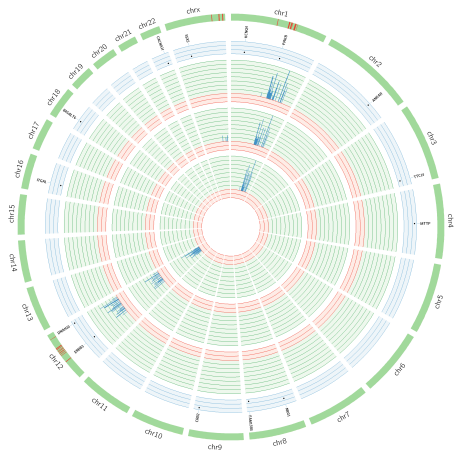
Supplementary Fig. 1. Single cell variation.
Variation at the single cell level in a well-differentiated liposarcoma (local recurrence 2 of Case 10), as demonstrated by G-banding analysis. (a) Cell with one large and one small ring chromosome; (b) Cell with two large ring chromosomes; (c) Cell with only one ring chromosome.

a

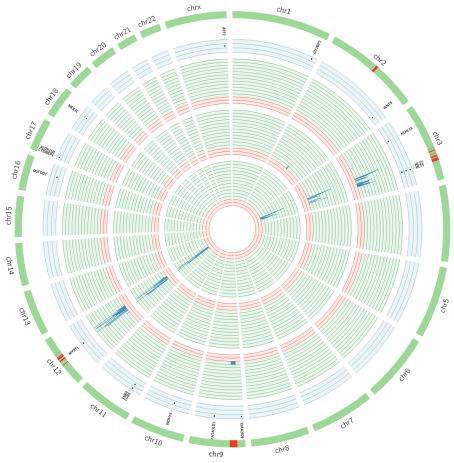


b

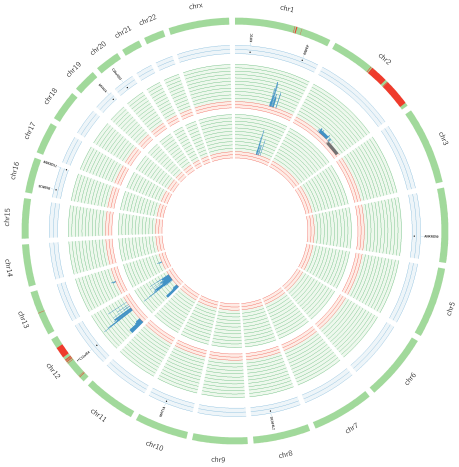
Case 10



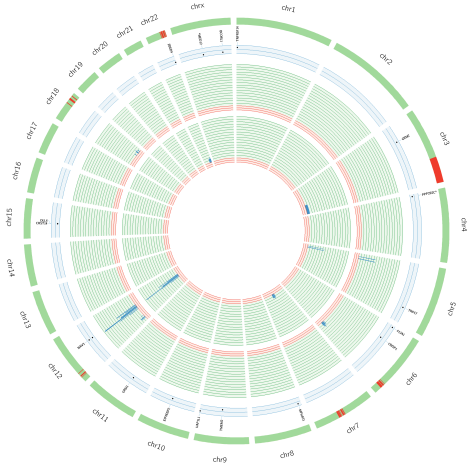
Case 11



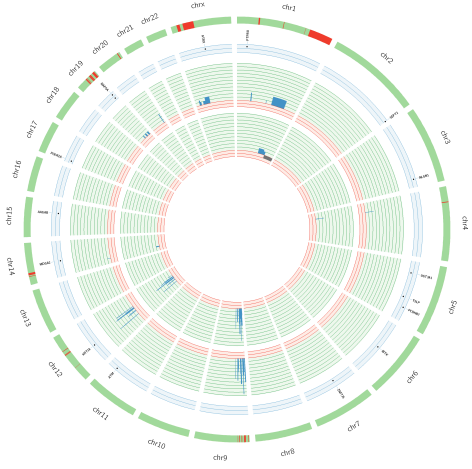
Case 12



Case 13

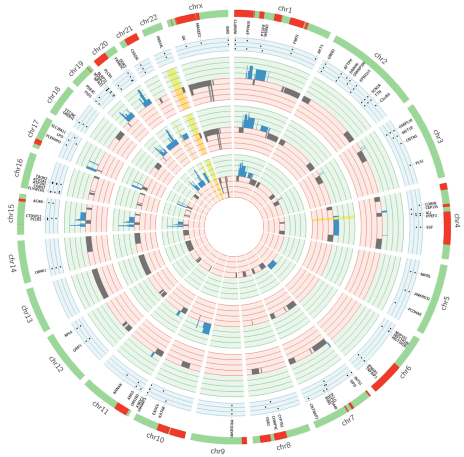


Case 14

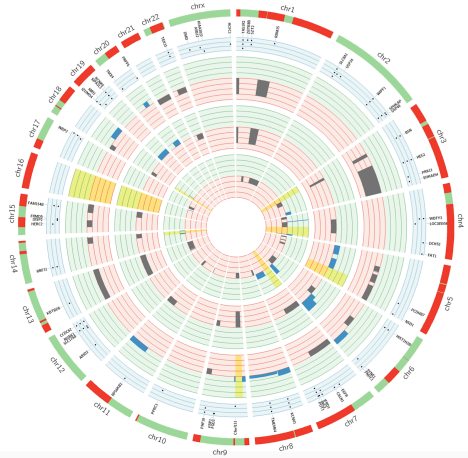


C

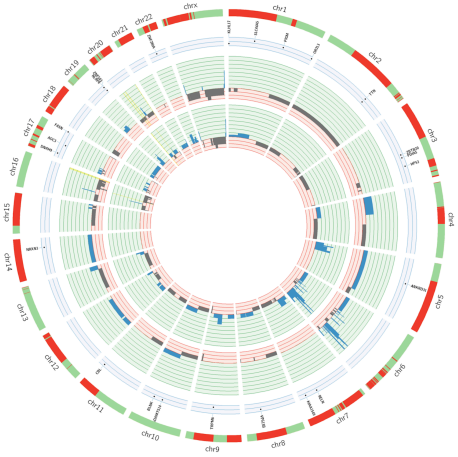
Case 15



Case 16

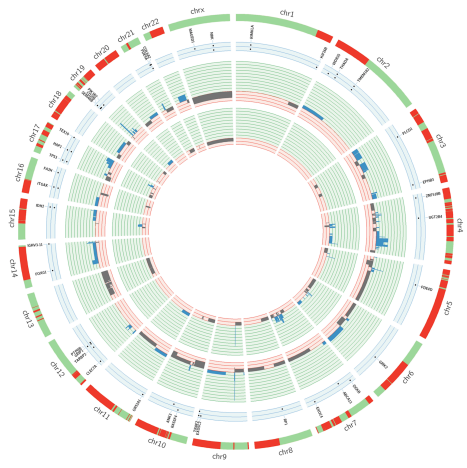


Case 17

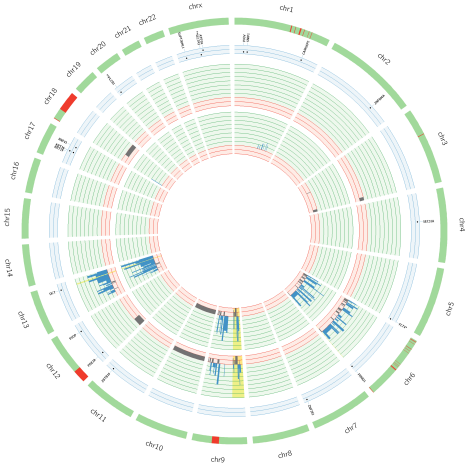


Case 18

Case 19



Case 20



Supplementary Fig. 2. Circos plots summarizing SNP array and whole exome sequencing results. The red/green inner circles represent the location and amplitude of the genomic changes at SNP array (GCS); blue is gain, grey is loss and yellow background indicates LOH. The circles are ordered chronologically, starting from the center with the first lesion. The light blue circles represent the location of the variants reported by the WES in the same order. Red on the green schematic chromosomes represents differences in GCS between lesions. (a) Fusion-driven sarcomas; (b) amplicon-driven sarcomas; (c) sarcomas with complex genomes.

Supplementary Table 1. Single cell variation in myxoid and well-differentiated liposarcomas, assessed by chromosome banding

					No. of chromosomes in AC[c]						
Case No.	Sample[a]	Chromosome number in stemline and sidelines[b]	No. of Ac[c]	No of cells with NCSA[d]	<45	45	46	47	48	49	>49
Myxoid liposarcoma											
1A	PT	46	11	0			11				
1B	LR2	46	11	0			11				
2A	PT	46	25	0			25				
2B	Met1	46	25	0		2	23				
2C	Met2	46	10	0			10				
2D	Met3	46	4	1		1	3				
5A1	PT	46	6	0			6				
5B	LR1	46	17	3	1		16				
7A1	PT	46	75	0			75				
7A2	PT	46	25	0			25				
7A3	PT	46	24	0			24				
8A	PT	47	23	0				23			
8B	Met1	47	25	0				25			
9A	PT	46	17	0			17				
9B	Met1	46	19	0			19				
Total			**317**	**4**	1	3	265	48	0	0	0
Well-differentiated liposarcoma											
10A1	PT	47-49/47/46	20	9			2	4	8	6	
10B	LR1	49	1	0						1	
10C	LR2	47-48	15	1		1		8	6		
11B	LR1	47	8	2				5	1		2
11C	LR2	45-50/88-94	17	14		1	2	4	5	2	3
11D	LR3	46-49/45-46	14	0		1	7	4	2		
12A1	PT	43-49/44-45/84-89	20	20	2	2				1	15
12B	LR1	76-88	11	3	1		1		1		8
13A1	PT	47/48-49/49-52/51	20	9				2	3	4	11
13B	LR2	45-47	22	2		2		12	4	1	3
14A1	PT	47-48/80-89	10	6	1		1	1	4		3
14B	LR1	48-50	9	4				1	4	3	1
Total			**167**	**70**	4	7	13	41	38	18	46

[a] PT = primary tumor; LR = local recurrence

[b] Only the modal chromosome number in each clone is shown here. Clones are separated by /. For full karyotypes, see Supplementary Table 6.

[c] AC = abnormal cells

[d] NCSA = non-clonal structural aberrations

Supplementary Table 2: Jaccard Index

WDLS

Sample	10A	10B	10C	11B	11C	11D	12A	12B	13A	13B	14A	14B
10A	1,00											
10B	0,69	1,00										
10C	0,94	0,78	1,00									
11B	0,07	0,08	0,07	1,00								
11C	0,06	0,07	0,07	0,57	1,00							
11D	0,05	0,07	0,06	0,67	0,39	1,00						
12A	0,19	0,17	0,20	0,23	0,18	0,23	1,00					
12B	0,08	0,07	0,08	0,05	0,03	0,05	0,25	1,00				
13A	0,02	0,02	0,02	0,15	0,07	0,13	0,05	0,02	1,00			
13B	0,03	0,04	0,04	0,08	0,08	0,06	0,10	0,04	0,11	1,00		
14A	0,17	0,15	0,17	0,03	0,02	0,04	0,09	0,06	0,01	0,02	1,00	
14B	0,16	0,14	0,16	0,02	0,01	0,04	0,08	0,06	0,02	0,02	0,40	1,00

WDLS: Chr12

Sample	10A	10B	10C	11B	11C	11D	12A	12B	13A	13B	14A	14B
10A	1,00											
10B	0,72	1,00										
10C	0,97	0,72	1,00									
11B	0,15	0,19	0,16	1,00								
11C	0,12	0,15	0,14	0,70	1,00							
11D	0,15	0,19	0,17	0,99	0,69	1,00						
12A	0,16	0,14	0,17	0,29	0,22	0,28	1,00					
12B	0,12	0,11	0,13	0,21	0,14	0,20	0,53	1,00				
13A	0,10	0,11	0,10	0,20	0,20	0,20	0,11	0,12	1,00			
13B	0,08	0,09	0,09	0,18	0,17	0,17	0,16	0,20	0,71	1,00		
14A	0,11	0,09	0,12	0,18	0,14	0,18	0,11	0,11	0,21	0,17	1,00	
14B	0,09	0,10	0,09	0,14	0,11	0,14	0,08	0,10	0,24	0,19	0,81	1,00

CXS

Sample	15A	15B	15C	16B	16C	16D	17A	17B	18A	18B	19A	19B	20A	20B
15A	1,00													
15B	0,68	1,00												
15C	0,62	0,73	1,00											
16B	0,36	0,33	0,26	1,00										
16C	0,28	0,24	0,23	0,32	1,00									
16D	0,29	0,20	0,23	0,24	0,53	1,00								
17A	0,36	0,41	0,33	0,38	0,30	0,31	1,00							
17B	0,27	0,35	0,29	0,23	0,27	0,23	0,47	1,00						
18A	0,32	0,30	0,27	0,27	0,17	0,18	0,25	0,17	1,00					
18B	0,34	0,32	0,29	0,24	0,18	0,19	0,25	0,18	0,93	1,00				
19A	0,22	0,30	0,23	0,27	0,19	0,18	0,33	0,27	0,19	0,16	1,00			
19B	0,29	0,34	0,30	0,37	0,31	0,35	0,56	0,51	0,15	0,14	0,39	1,00		
20A	0,12	0,09	0,09	0,20	0,16	0,15	0,21	0,16	0,15	0,15	0,23	0,19	1,00	
20B	0,16	0,12	0,13	0,17	0,15	0,18	0,23	0,17	0,13	0,13	0,23	0,22	0,79	1,00

Supplementary Data 1. SNP array intrasample heterogeneity.

Case	Diagnosis[a]	Ploidy level[b]	Chromosomal location[c]	A1	A2	A3	Aberration (No of copies)[d]	Position first abnormal SNP	Position last abnormal SNP	Genes affected[e]
10A1+A2+A3	**WDLS**	**2n**								
			1p11		1	1	1 Gain (3)	120 527 434	121 144 961	
			1q21		1	1	1 Gain (4)	147 832 189	149 898 950	
			1q21		1	1	1 Gain (4-5)	149 920 615	150 600 180	
			1q21		1	1	1 Gain (4)	150 873 782	151 984 957	
			1q21		1	1	1 Gain (3-9)	151 992 360	152 674 012	
			1q21-q22		1	1	1 Gain (3-4)	152 674 189	157 511 658	
			1q23		1	1	1 Gain (8-9)	157 511 738	158 296 404	
			1q23		1	1	1 Gain (4)	159 315 207	161 409 185	
			1q23-q24		1	1	1 Gain (3)	166 713 418	168 157 907	
			1q24		1	1	1 Gain (4-6)	169 563 568	170 637 677	
			1q24		1	1	1 Gain (4-9)	170 637 888	171 182 122	
			1q24		1	1	1 Gain (4)	171 182 169	171 806 983	
			1q24		1	1	1 Gain (7)	171 808 058	172 735 507	
			1q25		1	1	1 Gain (6-7)	180 866 711	181 519 139	
			1q31		1	1	1 Gain (8-9)	193 213 777	193 777 851	
			1q31		1	1	1 Gain (10)	193 842 116	194 530 427	
			1q32		1	1	1 Gain (7)	200 916 367	201 733 443	
			1q32		1	1	1 Gain (5)	202 165 531	203 365 745	
			1q32		1	1	1 Gain (5-7)	207 664 023	208 235 430	
			12q13		1	1	1 Gain (2-5)	52 724 077	53 275 334	
			12q13		*1*	*1*	*1 Gain (6-8)*	*57 792 433*	*58 296 523*	*CDK4: 6 copies*
			12q14		1	1	1 Gain (2-8)	60 678 349	62 250 017	
			12q14		*1*	*1*	*1 Gain (3-5)*	*66 142 050*	*66 389 967*	*HMGA2: 4-5 rearranged copies*
			12q14-q15		1	1	1 Gain (3)	67 392 970	68 206 890	
			12q15		1	1	1 Gain (2-7)	68 590 866	70 277 883	*MDM2*: 7 copies
			12q21		1	1	1 Gain (2-5)	71 600 673	72 705 758	
			12q21		1	1	1 Gain (2-6)	75 900 129	76 869 994	
			12q21		1	1	1 Gain (2-4)	79 207 917	79 989 266	
			12q21		1	1	1 Gain (4-6)	86 283 733	86 837 267	
			12q21		1	1	1 Gain (2-8)	89 987 170	90 801 897	
			12q22		1	1	1 Gain (6)	92 761 820	93 894 028	
			12q22		1	1	1 Gain (2-7)	95 262 678	96 816 253	
			12q23		1	1	1 Gain (2-6)	101 069 782	102 610 497	
			12q23		1	1	1 Gain (2-7)	107 574 000	108 306 855	
			12q24		1	1	1 Gain (2-5)	112 381 017	113 055 705	
			12q24		1	1	1 Gain (2-6)	113 417 472	116 018 360	
			12q24		1	1	1 Gain (7)	119 982 524	120 491 560	
			12q24		1	1	1 Gain (2-5)	123 986 854	124 774 710	
					38	**38**	**38**			
12A1+A2+A3	**WDLS**	**2n**								
			1q21		1	1	1 Gain (3-4)	143 932 349	144 914 801	
			1q21		1	1	1 Gain (6-9)	145 966 116	147 266 857	
			1q21		1	1	1 Gain (4-5)	147 391 922	148 216 112	
			1q21		1	1	1 Gain (5-6)	148 513 853	149 711 554	
			1q21		1	1	1 Gain (6-13)	149 713 775	151 598 196	
			1q21-q22		1	1	1 Gain (4-6)	156 175 733	156 911 190	
			1q23		1	1	1 Gain (4-6)	160 642 450	162 581 345	
			1q24		1	1	1 Gain (2-12)	171 765 963	172 375 257	
			1q24		1	1	1 Gain (6-8)	172 388 895	173 045 568	
			1q24		1	1	1 Gain (5-8)	173 047 673	173 738 477	
			12pter-p13		1	1	1 Gain (3-10)	1	14 078 487	
			12p13-p12		1	1	1 Gain (4)	14 078 633	15 370 009	
			12p12		1	1	1 Gain (3-4)	15 383 036	17 828 985	
			12p12		1	1	1 Gain (3-4)	17 834 307	21 766 931	
			12q14		1	1	1 Gain (2-7)	60 093 229	60 741 807	
			12q14		1	1	1 Gain (3-6)	64 170 858	65 884 346	
			12q14		1	1	1 Gain (10-15)	66 204 695	67 234 031	*HMGA2*: 11, 11 complete copies
			12q14-q15		1	1	1 Gain (3-7)	67 234 298	68 288 407	
			12q15		1	1	1 Gain (9-12)	68 288 577	68 936 263	
			12q15		1	1	1 Gain (8-9)	68 936 266	69 977 396	*MDM2*: 9 copies
			12q15		1	1	1 Gain (3-14)	69 978 701	70 745 073	
			12q15-q21		1	1	1 Gain (3)	70 751 623	71 760 021	
			12q21		1	1	1 Gain (5-10)	71 760 470	73 095 868	
			12q21		1	1	1 Gain (2-15)	74 155 068	75 354 206	
			12q21		1	1	1 Gain (5-11)	75 954 402	79 899 470	
			12q21		1	1	1 Gain (3-7)	79 899 520	86 450 295	

			12q21	1	1	1 Gain (2-14)	86 450 307	87 005 031
			12q21	1	1	1 Gain (3-6)	87 005 763	88 733 208
			12q21	1	1	1 Gain (3-6)	88 733 541	89 578 981
			12q21-q22	1	1	1 Gain (3-8)	89 580 796	92 751 711
			12q22	1	1	1 Gain (2-9)	92 754 017	93 533 469
			12q22	1	1	1 Gain (3-5)	93 551 462	95 885 589
			12q22-q23	1	1	1 Gain (8-10)	95 885 634	96 343 078
			12q23	1	1	1 Gain (3-6)	96 354 454	106 639 980
			12q23	1	1	1 Gain (3-9)	106 640 053	108 070 380
			12q23	1	1	1 Gain (6-11)	108 070 521	108 778 440
			13q14	1	1	1 Gain (3)	53 858 301	55 731 925
			13q21	1	1	1 Gain (3-4)	57 374 335	58 915 843
				38	**38**	**38**		
13A1+A2+A3	**WDLS**	**2n**						
			3q22-qter	1	1	1 Gain (3)	132 086 845	197 851 985
			5p15	1	1	1 Gain (2-15)	13 779 187	14 561 695
			5p14	1	1	1 Gain (2-15)	22 706 229	24 170 517
			7q21	1	1	1 Gain (2-5)	78 148 006	78 675 710
			7q21	1	1	1 Gain (2-5)	78 791 648	79 524 680
			7q21	1	1	1 Gain (2-5)	81 234 143	82 099 837
			7q21	1	1	1 Gain (2-5)	82 256 890	83 656 047
			7q21	1	1	1 Gain (3-5)	84 060 810	84 605 376
			7q21	1	1	1 Gain (2-5)	87 431 883	87 943 694
			7q21	1	1	1 Gain (2-5)	89 262 668	91 670 748
			7q21	1	1	1 Gain (2-5)	92 358 624	93 698 285
			7q21	1	1	1 Gain (2-5)	94 211 119	95 545 246
			12q14	*1*	*1*	*1 Gain (5-15)*	*58 014 571*	*58 450 790* *CDK4: 13, 9*
			12q14	1	1	1 Gain (5-7)	61 243 900	61 765 489
			12q14	1	1	1 Gain (7-14)	65 131 417	65 737 048
			12q14	*1*	*1*	*1 Gain (7)*	*66 186 663*	*66 240 759* *HMGA2: 7, 7 truncated copies*
			12q15	1	1	1 Gain (4-8)	69 001 286	69 032 458
			12q15	1	1	1 Gain (15-20)	69 049 534	71 158 276 *MDM2*: 15, 15
			12q21	1	1	1 Gain (2-17)	72 120 280	75 508 419
			12q21	1	1	1 Gain (2-12)	78 051 757	79 257 295
			12q21	1	1	1 Gain (8-14)	81 629 444	82 404 576
			12q21	1	1	1 Gain (8-14)	82 404 576	83 095 018
			22q13	1	1	1 Gain (2-5)	37 571 113	38 290 585
			22q13	1	1	1 Gain (2-5)	40 060 609	41 732 758
			22q13	1	1	1 Gain (2-5)	41 852 788	42 593 916
			22q13	1	1	1 Gain (2-5)	42 833 719	43 747 685
			22q13	1	1	1 Gain (2-6)	45 468 751	47 047 404
			22q13	1	1	1 Gain (2-5)	47 875 345	48 775 868
			22q13	1	1	1 Gain (2-5)	50 335 108	51 234 518
				27	**27**	**27**		
14A1+A2+A3	**WDLS**	**2n**						
			1q21-q25	1	1	1 Gain (3-4)	143 932 349	177 690 761
			1q25-q31	1	1	1 Gain (3)	177 690 761	189 326 616
			1q31-qter	1	1	1 Loss (1)	189 326 616	249 228 414
			4p14	1	1	1 Gain (4-5)	38 501 036	39 393 720
			4p14	1	1	1 Gain (3)	39 393 720	40 641 468
			9p24	1	1	1 Gain (4)	66 016	1 871 708
			9p24	1	1	1 Gain (4-10)	1 871 708	2 975 938
			9p24	1	1	1 Gain (4-8)	4 562 034	6 264 342
			9p24	1	1	1 Gain (5-9)	6 264 342	8 594 984
			9p24-p23	1	1	1 Gain (10-13)	8 594 984	9 378 800
			9p23	1	1	1 Gain (6-8)	9 378 800	10 622 626
			9p23	1	1	1 Gain (3)	10 622 626	11 199 164
			9p23	1	1	1 Gain (4-6)	11 199 164	12 791 960
			9p23	1	1	1 Gain (4-6)	12 791 960	13 479 842
			9p22	1	1	1 Gain (8-11)	14 801 808	16 312 906
			9p22	1	1	1 Gain (9-10)	16 979 587	17 767 716
			9p22	1	1	1 Gain (5-8)	17 767 716	19 080 370
			9p22	1	1	1 Gain (4-5)	19 080 370	19 784 020
			9p21	1	1	1 Gain (5-7)	27 242 352	30 048 939
			9p13	1	1	1 Gain (7-9)	38 133 338	38 698 686
			12q12	1	1	1 Gain (5-8)	39 269 336	39 853 527
			12q12	1	1	1 Gain (4-10)	39 853 527	40 544 633
			12q13	1	1	1 Gain (2-8)	55 368 738	56 225 782
			12q13-q14	1	1	1 Gain (2-7)	58 014 571	58 795 788 *CDK4*: 6, 7
			12q14	*1*	*1*	*1 Gain (3-4)*	*66 176 097*	*66 255 696* *HMGA2: 3 truncated copies*
			12q14-q15	1	1	1 Gain (2-8)	67 236 784	68 199 228

			12q15	1	1	1 Gain (2-15)	68 900 116	70 244 093 *MDM2*: 8, 13
			12q15	1	1	1 Gain (5)	70 244 093	70 784 878
			12q15	1	1	1 Gain (2-14)	70 784 878	71 456 004
			12q21	1	1	1 Gain (2-8)	72 636 501	73 372 953
			12q21	1	1	1 Gain (6-7)	77 274 064	77 844 894
			12q21	1	1	1 Gain (2-11)	77 844 894	78 587 040
			12q21	1	1	1 Gain (2-9)	79 230 474	79 934 619
			12q21	1	1	1 Gain (2-8)	81 028 851	81 542 981
			12q21	1	1	1 Gain (2-5)	90 800 000	91 578 028
			14q12	1	1	1 Gain (3)	25 024 558	29 859 588
				35	**35**	**35**		
18A1+A2	**MFS**	**2n**						
			1pter-p32	1	1	Loss (1)	1	57 317 492
			1p32-p13	1	1	Gain (4)	57 331 721	110 570 083
			1q21-q24	1	1	Gain (4)	151 473 155	174 193 251
			1q24-qter	1	1	Loss (1)	174 193 251	249 228 414
			2q32-qter	1	1	Loss (1)	197 162 328	243 096 490
			3q13	1	1	Loss (1)	105 628 142	109 956 046
			7q11-q22	1	1	Gain (3)	77 500 338	106 814 233
			7q22-q31	1	1	Gain (3-4)	106 814 233	114 204 094
			7q31	1	1	Loss (1)	114 204 094	115 872 698
			7q31	1	1	Gain (2-7)	115 872 698	119 529 848
			7q31	1	1	Loss (1)	119 529 848	121 255 217
			7q31	1	1	Gain (4-7)	121 255 217	122 157 852
			7q31	1	1	Gain (4)	122 157 852	124 682 285
			7q31	1	1	Loss (1)	125 122 286	125 792 854
			7q31-q36	1	1	Gain (4)	125 792 854	155 464 930
			7q36-qter	1	1	Loss (1)	155 464 930	159 119 732
			9pter-p21	1	1	Loss (1)	1	21 894 052
			9p21	1	1	hom del	21 894 052	24 305 038 *CDKN2A, CDKN2B, MTAP*
			9p21	1	1	Loss (1)	24 305 038	26 245 326
			9p21-p13	1	1	Gain (3)	26 245 326	35 321 622
			9p13	1	1	Loss (1)	35 323 697	38 787 479
			11p13-q12	1	1	Loss (1)	34 201 460	61 048 625
			13	1	1	Loss (1)	1	115 108 502
			15q11-q21	1	1	Loss (1)	22 318 372	43 895 366
			16q11-qter	1	1	Loss (1)	46 463 674	90 289 024
			20pter-p12	1	1	Loss (1)	1	17 517 726
			20p12-p11	1	1	Gain (4)	17 517 726	22 341 939
				27	**27**			
20A1+A2+A3	**Myoepith**	**2n**						
			1q21	**0**	**0**	**1 Gain (3)**	**146 101 790**	**147 823 369**
			1q23	**0**	**0**	**1 Gain (3)**	**157 511 738**	**158 246 699**
			1q24	**0**	**0**	**1 Gain (3-4)**	**169 835 949**	**172 744 542**
			1q25	**0**	**0**	**1 Gain (3-4)**	**180 870 203**	**181 519 034**
			1q31	**0**	**0**	**1 Gain (4)**	**193 430 422**	**193 751 815**
			1q32	**0**	**0**	**1 Gain (3-4)**	**201 020 381**	**201 754 528**
			3p14	1	1	1 Loss (1)	68 047 486	68 873 856
			3q27-qter	1	1	1 Loss (1)	185 324 972	198 000 000
			6p25	1	1	1 Gain (4-6/3-5)	1	1 083 918
			6p25	1	1	1 Gain (4-5)	1 120 025	1 575 692
			6p25	1	1	1 Loss (1)	1 575 692	2 866 616
			6p25	1	1	1 Gain (5)	2 866 616	3 048 519
			6p25	1	1	1 Loss (1)	3 048 519	**3 298 769**
			6p25	1	1	1 Gain (3)	3 604 240	4 462 036
			6p25-p24	1	1	1 Loss (1)	**4 532 034**	**10 473 741**
			6p24-p23	1	1	1 Gain (6/5)	10 778 654	14 209 231
			6p23-p22	1	1	1 Loss (1)	14 209 231	16 119 003
			6p22	1	1	1 Gain (4-6/4)	16 119 003	21 969 728
			6p22	1	1	1 Loss (1)	**24 797 942**	**25 415 244**
			6p22	1	1	1 Loss (1)	26 285 545	28 818 171
			6p22	1	1	1 Loss (1)	30 817 684	30 959 185
			6p21	1	1	1 Loss (1)	32 704 898	32 933 990
			6p21	1	1	1 **Gain (5)**	**37 168 935**	43 419 731
			6p21	1	1	1 Gain (6/5)	44 366 884	45 717 810
			6p12	1	1	1 Gain (7-9/4-6)	45 717 810	47 708 022
			6p12	1	1	1 Gain (5/4)	47 708 022	47 842 890
			6p12	1	1	1 Loss (1)	47 842 890	54 375 558
			6p11	1	1	1 Loss (1)	57 605 632	57 890 311
			6q12	1	1	1 Loss (1)	64 152 925	69 225 007
			6q13-q14	1	1	1 Gain (3)	70 066 868	72 086 260

6q14	1	1	1 Gain (3)	72 295 170	77 875 202
6q14	1	1	1 Loss (1)	81 536 238	81 629 789
6q14	1	1	1 Gain (6-7/5)	81 629 789	85 433 675
6q15	1	1	1 Gain (6-8/4-6)	89 029 272	90 940 618
6q15	1	1	1 Gain (3)	90 994 044	91 977 339
6q15-q16	1	1	1 Gain (3)	92 500 950	95 417 611
6q16	1	1	1 Loss (1)	99 535 538	101 963 505
6q16-q21	1	1	1 Gain (6-7/5-6)	103 088 705	110 578 467
6q21	1	1	1 Gain (3)	113 735 151	115 617 759
6q22	1	1	1 Gain (6/5)	121 780 969	124 578 716
6q22-q23	1	1	1 Gain (3)	124 589 302	126 317 684
6q23	1	1	1 Gain (5-9/4-6)	126 317 684	130 731 309
6q23-q25	1	1	1 Gain (3)	131 364 766	157 857 739
6q25	1	1	1 Gain (3)	158 572 391	159 465 858
6q25-q27	1	1	1 Gain (5-6/4-5)	159 541 772	166 620 670
6q27-qter	1	1	1 Loss (1)	168 836 750	171 000 000
9pter-p13	1	1	1 Copy neutral LOH	1	38 787 479
9p24-p23	1	1	1 Gain (5-6/4)	7 383 988	10 406 382
9p23	1	1	1 Loss (1)	10 406 382	10 860 265
9p23	1	1	1 **Gain (5/4)**	10 860 265	12 802 215
9p23	1	1	1 Gain (8)	12 802 215	13 199 594
9p23-p22	1	1	1 Gain (5)	13 199 594	16 106 014
9p22	1	1	1 Gain (8)	16 106 014	18 865 754
9p22	1	1	1 Gain (7)	18 865 754	19 326 112
9p22	1	1	1 Gain (4)	19 326 112	19 435 274
9p22-p21	1	1	1 Hom del (0)	19 435 274	28 643 576
9p21	1	1	1 Loss (1)	28 643 576	29 370 527
9p21	1	1	1 Loss (1)	30 432 336	32 566 127
9p13	1	1	1 Gain (5/4)	33 464 832	34 993 024
9p13	1	1	1 Loss (1)	34 993 024	35 298 697
9p13	1	1	1 Gain (5/3)	35 298 697	37 251 776
9p13	1	1	1 Loss (0-1)	37 251 776	38 800 620
9q21-q22	1	1	1 Gain (6/5)	71 012 048	71 296 385
9q21	1	1	1 Gain (6/5)	**72 338 410**	**75 507 288**
9q21	1	1	1 Gain (5-6/4)	**75 512 990**	90 017 860
9q21-q22	1	1	1 Gain (8/6-8)	90 017 860	94 666 047
9q22	1	1	1 Gain (5-6/4)	94 666 047	95 369 385
9q22	1	1	1 Loss (1)	95 369 385	95 674 962
9q22	1	1	1 Loss (1)	97 165 318	101 596 977
9q22	1	1	1 Gain (6-7/5)	102 310 439	106 204 395
9q31	1	1	1 Gain (3)	106 204 395	106 936 518
9q31	1	1	1 Gain (3)	108 311 066	111 205 333
9q31	1	1	1 Gain (6/4-5)	111 205 333	112 395 108
9q31-q33	1	1	1 Gain (3)	112 395 108	118 131 630
9q33	1	1	1 Loss (1)	118 862 694	123 486 424
9q33	1	1	1 Gain (5-8/4-6)	123 486 424	124 433 230
9q33	1	1	1 Loss (1)	124 433 230	124 627 672
9q33	1	1	1 Gain (6/4)	125 900 562	127 954 729
9q34	1	1	1 Loss (1)	130 549 616	132 116 760
9q34	1	1	1 Loss (1)	133 030 889	133 837 389
9q34	1	1	1 Loss (1)	138 527 244	139 390 258
10	1	1	1 Loss (1)	1	134 000 000
12q24	**0**	**0**	**1 Gain (3)**	**119 982 524**	**120 541 162**
13q12	1	1	1 Gain (3)	19 280 035	20 410 527
13q12	1	1	1 Gain (4)	20 410 527	23 501 972
13q12	1	1	1 Gain (3)	23 501 972	24 688 165
13q12	1	1	1 Gain (3)	24 688 165	27 891 388
13q12	1	1	1 Gain (6/4)	27 891 388	28 032 259
13q13	1	1	1 Loss (1)	29 664 520	30 321 425
13q13	1	1	1 Gain(3)	34 032 903	36 273 577
13q13	1	1	1 Gain (6/5)	36 556 117	37 143 478
13q13-q14	1	1	1 Gain (7-9/6)	37 143 478	41 982 332
13q14	1	1	1 Gain (5-6/4-5)	41 982 332	46 933 739
13q14	1	1	1 Gain (9/6)	46 933 739	49 574 610
13q14	1	1	1 Gain (5-6/5)	49 574 610	50 012 084
13q14	1	1	1 Loss (1)	53 216 428	53 596 196
13q14-q21	1	1	1 Gain (5-6/3-4)	53 596 196	56 807 502
13q21	1	1	1 Loss (1)	56 807 502	60 433 351
13q21	1	1	1 Gain (5-6/4)	60 433 351	62 884 089
13q21	1	1	1 Gain (8/5)	62 884 089	65 416 850
13q21	1	1	1 Gain (5/3)	65 416 850	70 344 288
13q21-22	1	1	1 Gain (3)	71 193 691	74 135 604
13q22-q31	1	1	1 Gain (5-6/5)	74 645 126	89 444 402

13q31	1	1	1	Gain (9-10/7)	89 444 402	94 517 230
13q31-q32	1	1	1	Gain (6/5)	94 517 230	95 784 654
13q32	1	1	1	Copy neutral LOH	95 793 870	99 175 012
13q32-q33	1	1	1	Gain (6-7/5)	99 178 857	105 186 746
13q33	1	1	1	Gain (8-11/6-8)	105 186 746	111 359 147
13q34	1	1	1	Gain (5-6/4-5)	111 367 607	114 784 386
18pter-p11	**1**	**1**	**1**	**Gain (5)**	**1**	**1 125 826**
	103	**103**	**110**			

[a] WLDS = well-differentiated liposarcoma; MFS = myxofibrosarcoma; Myoep = myoepithelial tumor.

[b] A slash denotes different ploidy levels in different samples.

[c] Chromosomal bands affected by the imbalance; at the end, the total number of AI among all samples is given. Regions in bold represent imbalances that were different among the samples from the same patient. Regions in normal font indicate shared imbalances. Regions in italics indicate imbalances that were <500 kb, but that were either homozygous deletions or affected pathogenetically important genes in WLDS.

[d] Gains and losses are in relation to the estimated ploidy level of the tumor. Numbers in parentheses indicate the number of copies present; copy number differences between samples A and B in Cases 19 and 21 are indicated by /. Hom del = homozygous deletion.

[e] Only genes affected by homozygous deletions and the 3 critical genes in 12q in WLDS (*CDK4*, *HMGA2*, and *MDM2*) are indicated.

[f] SB = fraction of shared breakpoints among all samples (SBA) or between specific samples from the same patient.

Supplementary Data 2. Genomic imbalances detected by SNP array analysis in multiple samples from sarcomas.

Table too large for printing.

Supplementary Data 3. Results of whole exome (WES) and targeted re-sequencing (TSCA)

Table too large for printing.

Supplementary Data 4. Clinical data and summary of genetic analyses performed.

Case No	Genetic subgroup	Material[a]	Diagnosis[b]	Grade[c]	Age/Sex	Location/ Depth[d]	Size[e]	Local recurrence[f]	Metastasis[g]	Adjuvant treatment[h]	Outcome[i]	WES[j]	TSCA[k]	SNP array[l]	Karyotype[m]	Gene fusion	Clinical comments
1A	Fusion-driven	PT	MLS	?	51/F	Knee/D	7	9,**22**	No	RT after LR2	NED 119	68x	1787x	Illumina	46,XX,t(4;16;12)(q1?1;p11;q1?3),r(8),der(10)add(10)(p?)t(8;10)(q2?;q2?),i(13)(q10),der(15;?22)(q10;q10)[11]	*FUS/DDIT3*	
1B		LR2		?								67x	943x	Illumina	46,XX,t(4;16;12)(q1?1;p11;q1?3), ?r(8),der(10)add(10)(p1?1)t(8;10)(q2?;q2?),i(13)(q10)[11]		
2A	Fusion-driven	PT	MLS	3	60/M	Thigh/D	13	No	**30**,**92**,**98**, 136	RT after PT, Met1, and Met4. CT after Met3	153 DoD	67x	934x	Illumina	46,XY,t(12;13;16)(q13;q12;p11)[25]		
2B		Met1		3								67x	423x	Illumina	idem[25]	*FUS/DDIT3*	
2C		Met2		3								ND	ND	ND	46,XY,t(2;18)(q35;q11),t(12;13;16)(q13;q12;p11)[10]		
2D		Met3		3								ND	ND	ND	46,XY,t(2;18)(q35;q11),t(12;13;16)(q13;q12;p11)[4]		
3A	Fusion-driven	PT	MLS	2	40/M	Thigh/D	14	No	**19**,**28**,**32**	No	90 DoD	117x	645x	Illumina	46,XY,t(12;16)(q13;p11)[25]	*FUS/DDIT3*	
3B		Met1		3								104x	626x	Illumina	idem[25]		
3C		Met2a		3								ND	ND	ND	idem[25]		
3D		Met2b		3								ND	ND	ND	idem[25]		
3E		Met3		2								ND	ND	ND	idem[21]		
4A	Fusion-driven	PT	MLS	3	42/F	L leg/D	8	No	**19**,**20**,**74**	No	78 DoD	94x, 290x	222x	Illumina	45,XX,der(2)t(2;12)(q35;q13),-9,der(12)t(12;16)(q13;p11),der(16)t(9;16)(p13;p11),der(20)t(9;20)(q11;p13)[21]	*FUS/DDIT3*	
4B		Met1		?								107x	455x	Illumina	idem[25]		
4C		Met2a		2								102x	84x	Illumina	idem[24]		
4D		Met2b		2								94x	411x	Illumina	idem[16]		
4E		Met3		3								71x	107x	Affymetrix	idem[21]		
5A1	Fusion-driven	PT	MLS	3	43/M	L leg/D	14	**104**	No	No	252 NED	ND	ND	Illumina	46,XY,t(12;16)(q13;p11)[6]	*FUS/DDIT3*	
5A2		PT		3								ND	ND	ND	idem[7]		
5B		LR1		3								ND	ND	Illumina	idem[17]		
6A	Fusion-driven	PT	MLS	2	77/F	Foot/D	10	**12**	16	No	24 DoD	ND	ND	Illumina	46,XX,idic(7)(p11),del(10)(p12),t(12;16)(q13;p11)[24]	*FUS/DDIT3*	
6B		LR1		3								ND	ND	Illumina	46,XX,t(12;16)(q13;p11)[18]		
7A1	Fusion-driven	PT	MLS	3	51/M	Knee/D	7	**25**	No	No	195 NED	ND	ND	Illumina	46,XY,t(12;16)(q13;p11)[75]	*FUS/DDIT3*	
7A2		PT		3								ND	ND	ND	46,XY,t(12;16)(q13;p11)[25]		
7A3		PT		3								ND	ND	ND	46,XY,t(12;16)(q13;p11)[24]		
7B		LR1		3								ND	ND	Illumina	idem[?]		
8A	Fusion-driven	PT	MLS	2	48/M	Knee/S	7	96	**48**	No	213 NED	ND	ND	ND	47,XY,der(1)add(1)(p32)add(1)(q42),+8,der(10)t(1;10)(p34;p11)ins(10;?)(p11;?),t(12;16)(q13;p11)[23]	*FUS/DDIT3*	
8B		Met1		2								ND	ND	ND	idem[25]		
9A	Fusion-driven	PT	MLS	2	40/F	Groin/D	13	No	**42**	No	89 NED	ND	ND	ND	46,XX,t(12;16)(q13;p11)[17]	*FUS/DDIT3*	
9B		Met1		3								ND	ND	ND	idem[19]		
10A1	Amplicon-driven	PT	WDLS	1	51/F	Thigh/D	?	**197**,**306**	No	No	338 NED	87x	483x	Affymetrix	47-49,XX,+1-3r[16]/47,XX,der(1)t(1;19)(q11;q11),add(7)(q11),der(19)t(7;19)(q11;q11),+r[2]/46,XX,der(1)t(1;10)(q11;q11),-10,+r[2]		LR1: 19 cm, shelled out with intact capsule; LR2: 20 cm, shelled out with marginal margin
10A2		PT		1								108x	ND	Affymetrix	ND		
10A3		PT		1								123x	ND	Affymetrix	ND		
10B		LR1		1								79x	152x	Affymetrix	49,XX,+3r[1]		

10C		LR2		1							**78x**	265x	Affymetrix	47-48,XX,+1-2r[15]	
11A	Amplicon-driven	PT	WDLS	1 39/M	U arm/S	? **84**,**124**,**141**	No	No	141 NED	ND	ND	ND	ND		PT: resected with marginal margin; LR1: 10 cm, resected with unknown margin; LR2: 4 cm, resected with intralesional margin; LR3: 6 cm, shelled out
11B		LR1		1							82x	349x	Affymetrix	47,XY,+r[5]	
11C		LR2		1							**84x**	489x	Affymetrix	45-50,XY,add(1)(q11),add(5)(p11),tas(12;21)(p13;p13),-17,-21,+1-3r,+mar,inc[15]/88-94,idemx2[2]	
11D		LR3		1							99x	456x	Affymetrix	46-49,XY,+1-3r[7]/45-46,XY,-13,-15,+r,+mar[7]	
12A1	Amplicon-driven	PT	WDLS	1 52/M	Thigh/D	21 **124**	No	RT after LR1	160 NED	68x	91x	Affymetrix	43-49,XY,+mar[2]/44-45,XY,+r[2]/84-89,XXYY,+1-2r[4]		PT: intralesional excision; LR1: 17 cm
12A2		PT		1							133x	ND	Affymetrix	ND	
12A3		PT		1							106x	ND	Affymetrix	ND	
12B		LR1		1							88x	454x	Affymetrix	76-88,XXY,+1-2r,inc[6]	
13A1	Amplicon-driven	PT	WDLS	1 46/F	Thigh/D	8 192, **215**, 228	No	RT after LR3	228 LTF	78x	207x	Affymetrix	47,XX,add(7)(p22),+mar[6]/48-49,XX,der(13;14)(q10;q10),+3-4r[5]/49-52,XX,i(14)(q10),+3-6r[3]/51,XX,der(14)t(8;14)(q21;p?),+5r[2]		LR2: 17 cm; LR3: 10 cm
13A2		PT									133x	ND	Affymetrix	ND	
13A3		PT									99x	ND	Affymetrix	ND	
13B		LR2		1							77x	63x	Affymetrix	45-47,XX,+1-4r[22]	
14A1	Amplicon-driven	PT	WDLS	1 54/F	Thigh/D	14 **211**	No	No	211 LTF	111x	93x	Affymetrix	47-48,XX,+1-2r[7]/80-89,XXX,-X,+1-2r[3]		LR1: 10 cm. Seen at MRI 18 months before; then 5 cm
14A2		PT		1							109x	ND	Affymetrix	ND	
14A3		PT		1							110x	ND	Affymetrix	ND	
14B		LR1		1							76x	57x	Affymetrix	48-50,XX,+1-3r,+1-2mar[9]	
15A	Complex	PT	MFS	3 70/M	L arm/S	5 **47**, 76	**114**, 115	RT after PT	127 DoD	137x	12x	Affymetrix	41-49,XY,-1,add(2)(p2?),add(3)(p?23),-4,-7,-10,add(10)(q26),add(11)(p15),-14,-19,?add(20)(q11),+der(?)t(?;12)(?;q13),inc[14]/88-97,idemx2[8]		
15B		LR1		3							65x	12x	Affymetrix	45-55,XY,inc[11]/93-96,?Y,-X,add(1)(p36)x2,add(1)(p13),der(2)add(2)(p2?1)add(2)(q37)x2,del(3)(p12),add(6)(q12),add(10)(q26),add(20)(q11)x2,+der(?)t(?;12)(?;q13),inc[13]		
15C		Met1		3							102x	143x	Affymetrix	81-100,-X,-X,-Y,?der(1;15)(q10;q10),der(2)add(2)(p2?1)add(2)(q37)x2,add(6)(q12),add(15)(p11),inc[8]		
16A	Complex	PT	MFS	? 74/M	Thigh/?	? **60**,**100**,125, 134, 147,**152**, 157	112	RT after PT and LR6	157 LTF	ND	ND	ND	ND		LR3: 4 cm. LR4: 4.5 cm. LR5: 5cm
16B		LR1		3							103x	122x	Affymetrix	46,Y,add(X)(q24),add(2)(q33),add(12)(q11),add(14)(q24),del(15)(q22),add(17)(q23),-18,der(22)t(18;22)(q11;q12),+mar[6]	
16C		LR2		3							107x	189x	Affymetrix	39-46,XY,add(1)(p11),-2,-9,?add(9)(q32),-10,add(13)(q22),add(14)(q1?3),add(15)(p11),der(17)t(1;17)(p1?1;p13)ins(17;?)(p13;?),-18,-20,inc[6]		
16D		LR6		3							124x	59x	Affymetrix	Failure	
17A	Complex	PT	MFS	3 58/M	Thigh/S	5 No	43, 66,**77**	CT after Met2	81 DoD	99x	214x	Affymetrix	67-74,XY,-X,add(1)(p34),del(1)(q12),del(1)(q21),der(1;7)(q44;p22)del(1)(p11),der(3)t(3;?18)(q21;q11),add(5)(q11),add(6)(q11)x2,add(8)(p11),add(10)(p15),add(11)(p15),?del(11)(p11),add(13)(q3?),add(15)(q22),add(19)(p13),?add(22)(q13),inc[10]		
17B		Met3		3							79x	66x	Affymetrix	64-77,X?,del(1)(q12),del(1)(q21),add(3)(q21),add(6)(q11)x2,der(1;7)(q44;p22)del(1)(p11),add(8)(p11),?del(9)(q12),add(11)(p15),del(11)(p11),add(13)(q32),?del(15)(q22),add(19)(p13),?add(21)(p11),inc[5]/129-154,idemx2[5]		
18A1	Complex	PT	MFS	3 73/F	L leg/S	4 36,**110**	No	RT after LR1	123 NED	78x	30x	Affymetrix	46,XY		LR1: 1-2 cm, resected with intralesional margins; LR2: 2.5 cm, resected with wide margins

18A2		PT		3							97x	ND	Affymetrix	ND		
18A3		PT		3							132x	ND	Failure	ND		
18B		LR2		3							85x	55x	Affymetrix	46,XX,add(1)(q32),der(3)t(3;5)(q22;q14),add(5)(q14),add(7)(q11),add(10)(p12),der(20)t(??;20)(q21;q13)[6]		
19A1	Complex	PT	MFS	3	69/M	L arm/D	12	No	**86**, 106	No	109 DoD	89x	108x	Affymetrix	40-44,XY,add(5)(p15),der(15)t(11; 15)(q13;p11),der(19)t(15;19)(q12;p11),inc[3]		
19A2		PT											ND	ND	37-40,XY,del(1)(p11),del(1)(q11),add(5)(p15),del(5)(p14),der(15)t(11;15)(q13;p11),der(19)t(15;19)(q12;p11)inc[5]		
19A3		PT											ND	ND	38-41,XY,del(2)(p14),del(15)(p14),der(6)t(6;11)(q15;q13),der(10;12)(q10;q10),der(19)t(15;19)(q12;p11),inc[3]/69-79,idemx2[3]		
19B		Met1		3								91x	766x	Affymetrix	45,X,-Y[13]		
20A1	Complex	PT	Myoep	3	44/F	L leg/D	4	92, **294**	No	No	LTF 295	104x	192x	Affymetrix	45-46,XX,add(3)(p13),-9,-9,-10,-13,-13,der(15;15)(q10;q10),+?der(22)t(14;22)(q13;p11),+ider(?)t(?;4)(q12)add(4)(q35),+der(?)t(?;4)(?;q12)add(4)(q35),+2-3mar[37]	RNA-seq negative	
20A2		PT		3								85x	ND	Affymetrix	ND		
20A3		PT		3								93x	ND	Affymetrix	ND		
20B	Complex	LR2		3								95x	252x	Affymetrix	Failure		

[a] PT = primary tumor; LR = local recurrence; Met = metastasis.

[b] MLS = myxoid liposarcoma; WDLS = well-differentiated liposarcoma; MFS = myxofibrosarcoma; Myoep = myoepithelial tumor.

[c] Malignancy grade according to 3-grade scale.

[d] L = lower; U = upper; D = deep-seated; S = superficial.

[e] Largest diameter in cm.

[f] Months to local recurrence. Samples that were analyzed are marked in bold.

[g] Months to metastasis. Samples that were analyzed are marked in bold.

[h] RT = radiotherapy; CT = chemotherapy.

[i] Outcome in months. NED = no evidence of disease; DoD = dead of disease; LTF = lost to follow-up.

[j] Average coverage at whole exome sequencing (WES).

[k] Average depth at targeted re-sequencing of mutations detected at WES. ND = not done.

[l-m] ND = not done.

Supplementary Data 5. Minimally gained segments in chromosome 12 in 12 samples from well-differentiated liposarcomas.

Chromosome	Start	End	Width (nt)	Ref seq genes
chr12	65586379	65586806	427	*LEMD3* (intron 1)
chr12	66208121	66240759	32638	*HMGA2* (exons 1-3)
chr12	69200588	69233068	32480	*MDM2*
chr12	69233068	69234034	966	*MDM2*
chr12	69234034	69235051	1017	*MDM2*
chr12	69235051	69260584	25533	*CPM*
chr12	69260584	69292810	32226	*CPM*
chr12	69292810	69361958	69148	*CPM*
chr12	69361958	69502490	140532	
chr12	69502490	69705232	202742	*CPSF6*
chr12	69705232	69708631	3399	
chr12	69708631	69751520	42889	*LYZ*
chr12	69758134	69831664	73530	*YEATS4* (exons 2-7)
chr12	69831664	69854068	22404	
chr12	69854068	69910432	56364	*FRS2*
chr12	69910432	69942858	32426	*FRS2*
chr12	69942858	69978017	35159	*FRS2*
chr12	70122843	70129622	6779	*LOC101928002* (exon 2)
chr12	70129622	70146943	17321	*LOC101928002* (exon 1), *RAB3IP* (exon 1)
chr12	70146943	70166168	19225	*RAB3IP* (exon 2-3)
chr12	70260147	70268818	8671	
Total			855876	

[a] Positions are according to GRCh37/hg19.

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903

The cover photo is a circular heatmap displaying copy number changes on chromosome 12 in liposarcomas. The background color, ranging from light to dark blue indicates a diploid state whereas green denotes gain of genomic material. The lighter the green the higher the level of amplification. The dots surrounding the circle indicate the location of breakpoints identified by a structural variant caller. The figure was generated using the circos software and the above image, portraying the author of this thesis, was generated using photoshop.

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Clinical Genetics

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2019:7
ISBN 978-91-7619-736-3
ISSN 1652-8220