THE GENETIC AND CELL FREE DNA CHARACTERISTICS OF SOFT TISSUE SARCOMAS

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Mr Nicholas Charles Eastley

MBChB BMedSci MRCS

Department of Genetics and Genome Biology

University of Leicester

August 2018

Abstract:

Soft tissue sarcomas (STSs) are a group of rare, malignant tumours with a relatively poor prognosis. Although they have been broadly classified as complex or simple based on their karyotype, only a few studies have investigated the genetic characteristics of STSs in detail. To further explore STS genomics we analysed a group of primary STSs using Illumina Next Generation Sequencing. This analysis revealed several characteristics of the analysed tumours including 1) a low single nucleotide variant and insertion/deletion mutational burden, 2) a high number of recurrent amplifications/deletions, 3) significant inter-tumoural heterogeneity regardless of histopathological classification and 4) complex genotypes in the vast majority of STSs analysed.

Currently no circulating STS biomarkers exist. Short fragments of DNA termed cell free DNA (cfDNA) can be found in the bloodstream. Some cancer patient cfDNA is tumour derived (circulating tumour derived DNA / ctDNA) and in several malignancies this ctDNA appears to correlate with disease behaviour. Despite this, very few studies have investigated STS patient cfDNA/ctDNA. To address this paucity of work we next used quantitative PCR, semiconductor targeted NGS and digital droplet PCR to characterise the cfDNA/ctDNA characteristics of two groups of metastatic and non-metastatic STS patients. This analysis revealed elevated cfDNA levels in the metastatic patients, which weakly correlated with disease burden suggesting a potential diagnostic role. Overall ctDNA was also identified in 27% (non-metastatic) - 36% (metastatic) of the analysed patients suggesting either that 1) the experimental approach used was not specific enough to detect ctDNA in tumours as genetically heterogeneous as STSs, or that 2) not all STSs shed ctDNA. Moving forwards although this analysis highlights a potential role for the use of ctDNA profiling in STS patients, it has also identified several significant challenges that must be addressed before ctDNA can be proposed as a realistic source for novel biomarkers.

Acknowledgements

Firstly I would like to thank my supervisors Dr Nicola Royle, Associate Professor Robert Ashford, Professor Jacqui Shaw and Dr Claire Esler for their continual guidance and enthusiasm over the last three years. Without this support this project would have certainly failed, and their constant encouragement has helped even the murkiest bioinformatic challenges appear surmountable.

Next I would like to thank all of the members of Dr Royle and Professor Shaw's laboratory teams for their help at various stages throughout the project – your advice and guidance has been essential and I thank each of you. In particular I would like to thank Dr Jin-li Luo and Dr Rob Hasting for their help analysing my sequencing data, Dr Barbara Ottolini for her guidance designing our targeted sequencing panels, and Rita Neumann for her help in many aspects of the project.

This project would not have been possible without the help of many members of the East Midland Sarcoma Service, the Nottingham Health Science Biobank or Associate Professor Ashford's administrative and clinical teams at the Leicester Royal infirmary. There are far too many individuals to name here, but I thank you all for your help.

Finally, but most importantly, I would like to thank my wife Lu for her unwavering support throughout this project, and my surgical career in general - sorry for the all the late nights, missed dates, and thank you for the countless sacrifices you've made.

Student Input

To facilitate sample collection with assistance from my supervisors I wrote the bulk of our initial IRAS ethics application, our sponsorship application to the University of Leicester NHS trust and the project's multiple successful funding applications. Throughout the project I maintained the site file and facilitated our sponsor's internal auditing process. I also successfully applied to the National Institute Health Research (NIHR) to have our project added to their research portfolio, allowing us to receive support from NIHR research nurses at Nottingham University Hospital NHS Trust.

Throughout the project I attended the East Midlands Sarcoma Service weekly MDT and personally enrolled over 90% of participants. I also attended the majority of participants' surgeries to ensure their intraoperative samples were collected and processed successfully. Following surgery I attended every post-operative outpatient appointment of those patients seen in Leicester (>70% participants) where I collected and processed the required whole blood samples. To facilitate the collection and processing of follow up whole blood samples from patients followed up in Nottingham I liaised with and trained i) staff from the Nottingham Health Science Biobank and ii) NIHR research nurses employed by the Nottingham City Hospital.

Prior to analysis I extracted all cfDNA, FFPE tumour DNA, fresh frozen tumour DNA and Buffy Coat DNA and quantified all of this DNA using the techniques outlined in chapter 2.8. To understand the SNV bioinformatic pipeline used during the project I processed one patient's sequencing data from its original raw FASTQ format to a useable VCF file as outlined in 2.14. The remaining samples were processed by Dr Jin-Li Luo (Bioinformatician at the University of Leicester). With close guidance from Dr Rob Hastings (Bioinformatician at the University of Leicester) I performed copy number analysis on patients' sequencing data as outlined in 2.14.3, and processed the resulting data using Maftools as outlined in section 2.14.4. For our targeted NGS analysis, after observing Dr Barbara Ottolini (Research scientist, The University of Leicester) design our first Iontorrent Panel (Sarcoma V1) I created our second panel (Sarcoma V2) independently. I also created all of the tNGS libraries analysed during the project, and alongside Dr Rita Neumann helped prepare and enrich several templates for sequencing. I also analysed all of the tNGS data as outlined in 2.16.5. For our ddPCR analysis I designed and optimised the cycling conditions for each ddPCR assay created in house as outlined in section 2.17. I also performed all of the patient sample analyses including SNV and copy number analysis, and processed/ interpreted all of the resulting raw data.

TABLE OF CONTENTS:

Chapter 3: The genetic characteristics of soft tissue sarcomas

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Chapter 2: Methods

Chapter 3: The genetic characteristics of soft tissue sarcomas

LIST OF ABBREVIATIONS

ALT	Alternative Lengthening of Telomeres
APBs	ALT associated PML bodies
APC	Adenomatous Polyposis Coli
ATM	ATM Serine/Threonine Kinase
ATRX	Alpha-thalassemia/mental retardation X-linked
AWD	Alive with disease
BAF	B allele frequency
BC	Buffy Coat
BCL2L1	BCL2-like 1
bp	Base pair
BRAF	B-Raf Proto-Oncogene
BRD9	Bromodomain containing 9
BRIP1	BRCA1 interacting protein C-terminal helicase 1
Ca 15-3	Cancer antigen 15-3
CCNE1	Cyclin E1
CDK2	Cyclin Dependent Kinase 2
CDK4	Cyclin Dependent Kinase 4
CDKN2A/B	Cyclin Dependent Kinase Inhibitor 2A/B
cfDNA	Cell-free DNA
CN	Copy number
CNAs	Copy Number Alterations
COSMIC	Catalogue of Somatic Mutations in Cancer
CR	Complete response
CT	Computer Tomography
CTCs	Circulating tumour cells
ctDNA	Circulating tumour-derived DNA
DACH1	Dachshund family transcription factor 1
DAXX	Death-domain associated protein
ddPCR	Droplet digital PCR
DDR	DNA damage response mechanism
DNA	Deoxyribonucleic Acid
DOC	Died from other causes
DOD	Died from disease
DSBs	Double strand DNA breakages
EGFR	Epidermal Growth Factor Receptor
EMSS	East Midlands Sarcoma Service
EPHB6	EPH receptor B6
ERAP2	Endoplasmic reticulum aminopeptidase 2
F	Female
FERMT3	Fermitin family member 3
FFPE	Formalin Fixed Paraffin Embedded

FH	Fumarate hydratase
FISH	Fluorescence in situ hybridization
FLT4	Fms-related tyrosine kinase 4
FRG1B	FSHD region gene 1 family member B
FRS2	Fibroblast Growth Factor Receptor Substrate 2
GISTIC	Genomic Identification of Significant Targets in Cancer algorithm
GISTs	Gastro-intestinal tumours
GRB2	Growth factor receptor bound protein 2
GRB2	Growth factor receptor bound protein 2
H3F3A/B	H3 histone family member 3A/B
HMGA2	High Mobility Group AT-Hook 2
HMGN3	High mobility group nucleosomal binding domain 3
HRAS	HRas proto-oncogene
hTERT	human telomerase reverse transcriptase
hTR	Human telomerase RNA
IGH	Intratumoural Genetic Heterogeneity
IGV	Integrated Genomics Viewer
IHC	immunohistochemistry
Indel(s)	Nucleotide Insertion/Deletion
IRAK1	Interleukin 1 receptor associated kinase 1
ISPs	Ion Sphere Particles
ITGB4	Integrin subunit beta 4
ITGB4	Integrin subunit beta 4
KIT	KIT Proto-Oncogene Receptor Tyrosine Kinase
LMS	Leiomyosarcoma
LNAs	Locked Nucleic Acids
LOH	Loss of heterozygosity
M	Male
MAPK7	Mitogen activated protein kinase 7
Mb	Megabase
MCL1	Myeloid cell leukaemia 1
MDM2	MDM2 Proto-Oncogene
MELK	Maternal embryonic leucine zipper kinase
MET	MET proto-oncogene
MFS	Myxofibrosarcoma
MPNST	Malignant peripheral nerve sheath tumour
MRI	Magnetic Resonance Imaging
MTAP	Methylthioadenosine Phosphorylase
MUC1/2/4/16	Mucin 1/2/4/16
MYO5C	Myocin VC
NED	No evidence of disease
NGS	Next Generation Sequencing

NHEJ	Non-homologous end-joining DNA repair pathway
NTC	Non-Template Control
PARP4	Poly(ADP-ribose) polymerase family member 4
PCR	Polymerase chain reaction
PD	Progressive disease
PDGFR	Platelet-derived growth factor receptor
PDGFRA	Platelet derived growth factor receptor alpha
PIK3CA	Phosphatidylinositol-4, 5-bisphosphate 3-kinase catalytic subunit alpha
PML	Promyelocytic leukaemia
PR	Partial response
PRB1	Proline rich protein BstNI subfamily 1
PTCH1	Patched 1
PTEN	Phosphatase And Tensin Homolog
qPCR	Quantitative Polymerase Chain Reaction
R1	Microscopically incomplete resection
R2	Macroscopically incomplete resection
RAC1	Rac family small GTPase 1
RASAL3	RAS protein activator like 3
RB1	RB transcriptional corepressor 1
RNA	Ribonucleic acid
RO	Wide resection
RPPH1	Ribonuclease P RNA component H1
SD	Stable disease
SEM	Standard error of the mean
SMAP1	Small ArfGAP 1
SNP	Single Nucleotide Polymorphism
SNVs	Single Nucleotide Variants
STK11	Serine/threonine kinase 11
STSs	Soft tissue sarcoma
TCGA	The Cancer Genome Atlas
TERT	Telomerase Reverse Transcriptase
TMMs	Telomere Maintenance Mechanisms
tNGS	Targeted Next Generation Sequencing
TOP3A	DNA Topoisomerase III Alpha
TP53	Tumour suppressor 53
TPM3	Tropomyosin 3
TTN	Titin
UK	United Kingdom
UPSs	Undifferentiated Pleomorphic Sarcomas
VAF	Variant allele frequency
WES	Whole Exome Sequencing

Chapter 1: Introduction

1.1 Soft Tissue Sarcomas

1.1.1 Definition

Soft tissue sarcomas (STSs) are a diverse group of malignant tumours that originate from tissues of a mesenchymal lineage. Based on their clinical, histological and genetic characteristics STSs have been classified extensively by the World Health Organisation, with over 50 different histological subtype currently recognised (Fletcher et al. 2013).

1.1.2 Epidemiology

An ageing population and increasing awareness of STSs in the medical profession has led to a steady increase in the incidence of STSs over the last 15 years (Beckingsale & Shaw 2017). Despite this, STSs remain relatively rare, with only 13,100 new cases diagnosed annually in the European Union (equating to just 1% of all new adult cancer diagnoses)(Weitz et al. 2003; Cormier & Pollock 2004).

In contrast to the majority of cancers STSs overall follow a bimodal age distribution (Burningham et al. 2012). From birth up until the age of 5 their incidence rises steadily before falling back down to its lowest level throughout adolescence. Following this, from the age of 20 upwards their incidence starts to escalate again, particularly above the age of 50 when 85% of cases are diagnosed. The mean age of a STS patient at diagnosis is 58 years, although individual subtypes vary considerably in their age distribution. This is clearly shown by the predisposition of certain STSs to children and adolescents (rhabdomyosarcoma), young adults (synovial sarcomas) or the elderly (angiomyosarcomas, leiomyosarcomas and liposarcomas)(Corey et al. 2014). The overall incidence of STSs is higher in males than females with a relative risk of 1.2:1(Corey et al. 2014; Hsieh et al. 2013; Pastore et al. 2006) Only a small number of STS subtypes buck this trend. These include myxofibrosarcoma and rhabdomyosarcoma which show no predilection to either sex, and malignant haemangiopericytoma and haemangiosarcoma which are more common in females with relative risks of 1.27:1 and 1.08:1 respectively (Corey et al. 2014). The incidence of STSs also varies between ethnicities. Data from 1973-2008 shows the highest incidence is seen in Afro-Caribbeans (5.1 per 100,000), followed by Caucasians (4.5 per 100,000) then American Indian/Asian Pacific Islanders (2.8 per 100,000) (Burningham et al. 2012). These racial and ethnic disparities are also evident in adolescents (Hsieh et al. 2013), and despite working

investigating many of the variables seen between different ethnics populations (including prenatal characteristics, birth characteristics, parental characteristics and drinking water components) the reasons behind them are not yet fully understood.

1.1.3 Aetiology

Although the majority of STSs occur sporadically with no identifiable cause, a small number of specific aetiologies have been recognised. Several rare genetic syndromes have been linked with STSs. These include Li–Fraumeni syndrome (in which STSs make up 25% of the malignancies)(Thomas & Ballinger 2015), type 1 Neurofibromatosis (which holds a 10% lifetime risk of malignant peripheral nerve sheath tumour (MPNST) (Uusitalo et al. 2016; Ferrari et al. 2007)) and hereditary retinoblastoma (which results in a STS relative risk of 184:1(Kleinerman et al. 2007). Several genotoxic environmental agents have also been linked with certain STSs. These include Human Herpes virus 8 (Kaposi's sarcoma), Epstein–Barr virus (leiomyosarcomas), radiotherapy (cutaneous angiosarcoma(Depla et al. 2014)) and vinyl chloride (hepatic angiosarcoma (Ward et al. 2001)). Finally certain STS subtypes also appear to preferentially occur in certain physical/anatomical environments. Examples include cases of lymphangiosarcoma in limb subjected to chronic lymphoedema following mastectomy (Stewart–Treves syndrome (Sharma & Schwartz 2012)) and cutaneous angiosarcoma adjacent to a non-functioning arteriovenous fistulae in renal dialysis patients (Ahmed & Hamacher 2002).

1.1.4 Clinical characteristics

The classical presentation for an extremity STS is a palpable lump, which may occur at almost any anatomical location (Table 1.1). In an attempt to hasten diagnosis, several groups have investigated which clinical characteristics of a lump are most predictive of malignancy. This work has shown that large (defined as > 5cm (Johnson et al. 2001)), growing and painful lumps should all be considered malignant unless proven otherwise, as well as lumps found beneath the deep fascia (Dangoor et al. 2016). Occasionally patients may present complaining of neurological symptoms resulting from STS invasion into nearby neurovascular structures. Although this is generally rare, in cases of malignant peripheral nerve sheath tumour these symptoms (pain +/- paraesthesia/weakness) are almost always present (Ogose et al. 1999). Finally patients may present with symptoms secondary to metastatic lesions. Most commonly these will involve the lung (where pulmonary compromise can result) although may also involve

other soft tissues (characteristic of myxoid liposarcoma (Estourgie et al. 2002)), intra-abdominal organ including the liver (myxoid liposarcoma (Estourgie et al. 2002) / leiomyosarcoma (Jaques et al. 1995)) or the skeleton (alveolar soft-part sarcoma / dedifferentiated liposarcoma / angiosarcoma / rhabdomyosarcoma (Yoshikawa et al. 1997)).

Site	Incidence (%)
Lower limb and girdle	40
Upper limb and girdle	20
Retroperitoneum	15
Trunk	10
Head and neck	10
Other	5

Table 1.1. Anatomical distribution of Soft tissue sarcomas. Adapted from Cormier JN, Pollock RE, *CA Cancer J Clin* 2004;54:94–109; Clark M, et al., *N Engl J Med* 2005;353(7): 701–11.

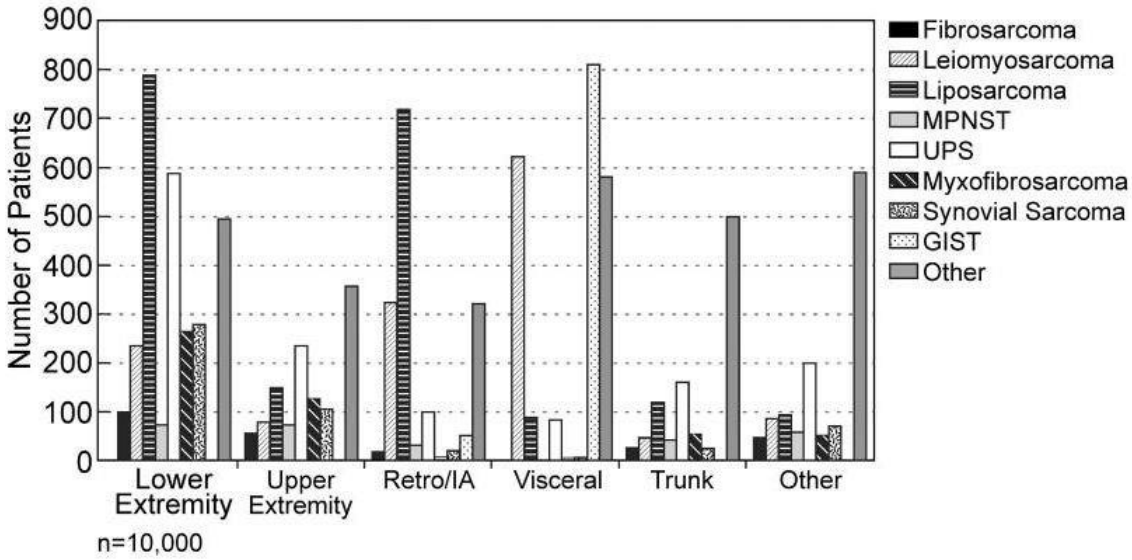


Figure 1.1: A histogram showing the relationship between soft tissue sarcoma location and histological subtype. This figure highlights the site dependent relationship with histopathology seen in STSs. **MPNST** - malignant peripheral nerve sheath tumour; **UPS** - undifferentiated pleomorphic sarcoma; **GIST** - gastrointestinal stromal tumour; **IA** – intraabdominal; **Retro** - retroperitoneal. Taken from Brennan MF, Antonescu CR, Moraco N, et al. Lessons learned from the study of 10,000 patients with soft tissue sarcoma. *Ann Surg* 2014;260(3):419. N=10,000.

1.1.5 Genetic characteristics of STSs

Within every cancer cell genetic mutations often accompanied by epigenetic changes result in certain 'hallmarks' that are essential for the development and progression of cancer (Hanahan & Weinberg 2011) (Figure 1.1). This highlights the importance of developing a detailed understanding of the genomic characteristics of all malignant tumours (including STSs) if the current rising incidence of cancer it to be addressed (Mistry et al. 2011). In a group of tumours as heterogeneous as STSs many different kinds of mutation are likely to play a role in the process of tumourigenesis. These may include Single Nucleotide Variants (SNVs) (a single nucleotide variation from the wild-type reference sequence), Indels (an insertion or deletion of nucleotides into the genomic sequence), Copy Number Alterations (CNAs) (a somatically acquired change in the copy number of a genomic region and the genes it contains) or even more complex chromosomal rearrangements such as unbalanced translocations.

1.1.5.1 Karyotype and genome complexity

As molecular techniques have developed a better understanding of the genetic characteristics of STSs has emerged (Mertens et al. 2010). Using cytogenetic techniques STSs have broadly been classified into one of two groups based on their karyotypes and general genome complexity. The first of these groups contains around 15 different STS subtypes and around 20% of STS cases overall (Mertens et al. 2010). These tumours are characterised by their relatively simple genomes, near diploid karyotypes and the simple nature of the mutations they contain (predominantly consisting of discrete amplifications (Mertens et al. 2009) and simple activating (oncogene) or inactivating (tumour suppressor) point mutations.) Several STSs within this group also contain subtype specific reciprocal translocations, which in some cases hold key diagnostic roles in a clinical setting (see section 1.4.5). The second group of STSs contains the remaining 80% of subtypes which are characterised by more complex, unstable genomes. In addition to simple point mutations these tumours contain many less specific, more extensive alterations, some of which have been identified as poor prognostic factors (Mertens et al. 2006). These may include significant amplifications and/or deletions, unbalanced translocations, or gross abnormalities in chromosome number suggestive of widespread genome instability (polyploidy or aneuploidy).

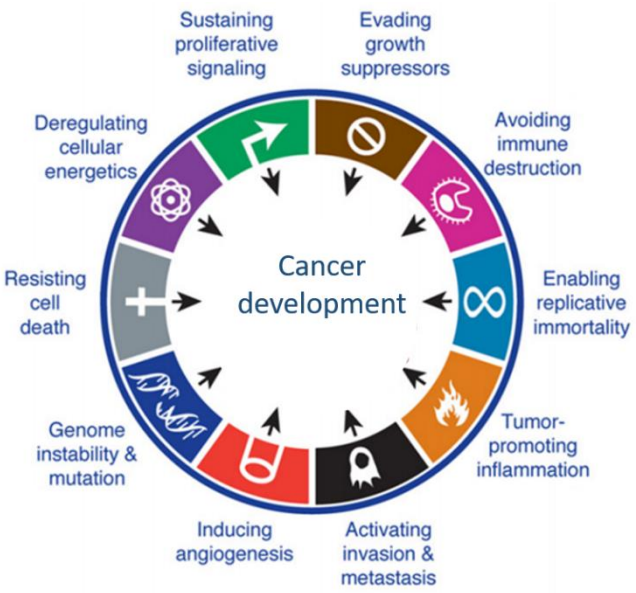


Figure 1.2: Hanahan's proposed hallmarks of cancer. The hallmarks outlined in this figure represent the biological characteristics of a malignant tumour that are key for it to develop and progress. The initial tumour development is facilitated by the chronic uncontrolled proliferation of its cells. This is facilitated by the genetic instability of these cells combined with their production of sustained proliferative signals, evasion of growth suppression, resistance to cell death and immune destruction, and capacity to maintain their own telomeres. To manage their own microenvironment to cope with this proliferation these cells also stimulate localised inflammation and angiogenesis, before finally invading their surrounding tissues and metastasising. Adapted from figure 6 in Hanahan et al (Hanahan & Weinberg 2011).

1.1.5.2 Single Nucleotide Variants

In recent years Next Generation Sequencing (NGS) techniques have revolutionised the field of cancer genetics, particularly as NGS costs have reduced and workflow speeds increased. The ability of NGS technologies to process millions of reactions in parallel allows many genomic regions to be sequenced simultaneously without compromising on read depth. This translates to a much higher sensitivity when calling variants compared with other older 1st generation sequencing techniques, and has propelled NGS forward as a fundamental tool for investigating the genetic characteristics of solid tumours.

Several groups have used NGS to investigate the SNV characteristics of STSs, with two particularly important sources of data currently publically available. The first and largest of these was performed in 2015 by Movva et al (Movva et al. 2015). This group analysed DNA isolated from 591 Formalin Fixed Paraffin Embedded (FFPE) STS samples using the Illumina TruSeq Amplicon Cancer Hotspot panel to sequence a total of 47 cancer associated genes on an Illumina MiSeq platform (see supplementary 1.1 for genes). In the same study another 1250 STSs were analysed using Sanger sequencing to sequence known hotspot regions in *BRAF, KRAS, KIT*, *EGFR,* and *PIK3CA*. Together these analyses revealed SNVs in 47% of the 1940 cases analysed involving 35 different cancer associated genes. Perhaps unsurprisingly *TP53* was the most commonly mutated of these (altered in 26% of cases) following by *BRCA2* (18%) then *BRCA1, PIK3CA, PTEN, ATM* and *APC* (all mutated in <5% of cases.)

The second substantial source of STS NGS data is provided by The Cancer Genome Atlas (TCGA) (National Cancer Institute 2018). This collaboration between the National Cancer Institute and National Human Genome Research Institute has created a comprehensive database of sequencing data in an attempt to uncover the underlying genomics of primary untreated malignant tumours. Several criteria govern which samples are selected for analysis by TCGA. These are in place to ensure that only tumours with a poor prognosis and public health impact are analysed, and that high quality data is produced form this analysis. Currently TCGA only accept fresh frozen tumour resection samples that have not been subjected to neoadjuvant treatment. In addition all tumour samples accepted are analysed alongside a source of germline DNA also collected from patients. This germline DNA is usually obtained from blood (or less commonly a normal tissue sample collected from the surgical field) and facilitates a more detailed genomic characterisation of tumours. Currently TCGA provides NGS SNV data for 254 STSs including over 6 different subtypes. Within this dataset *TP53* is the gene

most commonly found to contain SNVs (altered in 33% of tumours), although contrasting with Movva et al. the next most commonly mutated genes seen are *ATRX* (14%) (an inhibitor of the Alternative lengthening of Telomeres mechanism (Napier et al. 2015)), *TTN* (11%) (a gene central in the assembly and functioning of striated muscles previously identified as mutated in multiple cancers (Greenman et al. 2007)) , *RB1* (9%) (a common tumour suppressor (Chinnam & Goodrich 2011)) and *MUC16* (9%) (a recognised biomarker linked with several oncogenic processes including cellular proliferation and migration (Liu et al. 2016)).

A third significant source of STS sequencing data was produced by Barretina et al. in 2010 (Barretina et al. 2010). Using sanger sequencing and single nucleotide polymorphism (SNP) arrays the group analysed 207 STSs including 6 different subtypes reporting on both SNVs and CNAs. For their SNV analysis Barretina et al. sequenced 722 genes in total and successfully identified 37 mutations (28 SNVs and 9 indels) in 21 different genes. In contrast to Movva et al. and TCGA the most commonly mutated genes identified were *KIT* and *PIK3CA*, which were interestingly still only mutated in <3% of cases.

A key finding of Barretina and Movva et al's analyses is the low rate of SNVs identified. Barretina et al. reported an overall somatic mutation rate of 1.00 per megabase (Mb) of sequenced DNA in their cohort. With the exception of renal cancers (0.74/Mb), testicular cancers (0.12/Mb) and Acute Lymphoblastic Leukaemia (0.57/Mb) the majority of other malignancies contain a higher concentration of SNVs including melanoma (18.54/Mb), lung (4.2/Mb), gastric (2.1/Mb), breast (2.7/Mb), ovarian (1.85/Mb) and colorectal (1.21/Mb) cancers (Greenman et al. 2007). Both Barretina and Movva et al failed to identify any somatic SNVs in over 50% of the cases they analysed. If they were indeed truly absent in these cases (and not missed by the groups' targeted sequencing approaches) this suggests that other more complex alterations drove the process of sarcomagenesis in these tumours.

A second important observation of Barretina and Movva et al's studies is the wide variety of genes over which the identified SNVs were spread. Although subgroup analyses showed that certain genes appear to show a propensity to mutate in certain STS subtypes (for example *PIK3CA* in myxoid liposarcoma) neither group identified any recurrent SNVs in more than one of the tumours analysed. This suggests that significant SNV heterogeneity can be found between different tumours of the same STS subtype, as well as different STS subtypes.

1.1.5.3 Somatic Copy Number Alterations

Somatic copy number alterations (CNAs) affect a larger proportion of the cancer genome than any other type of genetic mutation and so are likely to play a critical role in activating oncogenes and inactivating tumour suppressors (Zack et al. 2013). Considering their complex karyotypes but relatively low somatic SNV burdens (Vogelstein et al. 2013) this is particularly likely to be the case for STSs.

The largest publically available sources of STS copy number data are again provided by Barretina et al. and TCGA (Barretina et al. 2010; National Cancer Institute 2018). Taken together these 2 studies highlight the high rate of CNAs seen in STSs, and the heterogeneity seen in the size and location of these abnormal segments. By identifying CNAs frequently involving the MDM2-p53 and the p16-CDK4-RB1 pathways they also eloquently show how characterising STS copy number profiles can have potential therapeutic implications (Abeshouse et al. 2017).

Barretina et al analysed their cohort of 207 STSs using an Affymetrix SNP array and the RAE and Genomic Identification of Significant Targets in Cancer (GISTIC) algorithms to look for statistically significant somatic CNAs. Using this approach they identified a total of 419 genomic segments with an abnormal copy number consisting of 144 amplifications and 275 deletions. These segments varied significantly in size from just 1 amplified /deleted gene up to 91 amplified or 272 deleted genes. Those genes most commonly amplified were *MDM2* (27%) (a key negative regulator of p53 commonly amplified in dedifferentiated liposarcomas (Boltze et al. 2001)), *FRS2* (26%) (a component of the mitogen-activated protein kinase pathway) *CDK2* (24%) (an oncogenic regulator of the cell cycle) and *HMGA2* (21%) (a transcription factor linked with adipogenesis and mesenchymal differentiation.) The most commonly deleted genes were *TP53* (11%) and *RB1* (8%).

The Cancer Genome Atlas reports on somatic CNAs in a total of 257 STSs. Significant CNAs were identified by analysing NGS data using the Genomic Identification of Significant Targets in Cancer (GISTIC) 2.0 algorithm (Mermel et al. 2011). In total 59 genomic regions with an abnormal copy number were identified in this way (22 amplifications, 37 deletions) which again showed great diversity in their size ranging from 1 amplified/deleted gene up to 507 amplified genes /337 deleted genes. *MDM2* was again the most commonly amplified gene seen (19%), again followed by *FRS2* (19%), *CDK2* (18%) and *HMGA2* (15%). *RB1* (16%), *CDKN2A/B* (14%) (two well recognised tumour suppressors), *MTAP* (12%) (a component of polyamine metabolism often lost in

cancer due to co-deletion with *CDKN2A*) and *TP53* (11%) were the most commonly deleted genes in the cohort.

1.1.5.4 Germline variants

Several STS characteristics suggest the involvement of germline genetic variants. These include their disproportionate incidence in children and adolescents, the increased risk of second cancers in STS survivors, and the association between STSs and several genetic syndromes such as Li-Fraumeni syndrome. To investigate these germline variants further and to characterise the individual and familial genetic determinants of sarcoma risk Ballinger et al. carried out a review of 1162 adult sarcoma patients (Ballinger et al. 2016). After performing targeted exon sequencing of 72 genes and a case-control rare variant burden analysis the group identified an excess of pathogenic germline variants in 55% of these patients, with 227 known or expected pathogenic variants identified in 19% of individuals. These genetic characteristics may explain the multiple primary cancers seen in 15% of analysed cohort, and the recognisable cancer syndromes identified in 17% of the 911 families reviewed. However, although Ballinger et al. concluded that a large clinically significant burden of genetic risk was present in STSs based on these findings, the absence of a direct measure of gene expression is a recognised limitation of their work, and precludes any concrete conclusions from being drawn necessitating further analysis.

1.1.5.5 Pathological assessment of STSs

The histopathological diagnosis of a STS is initially based upon an assessment of cell morphology. This assessment is key to rule out certain reactive processes that mimic STSs. Examples of such conditions include nodular and ischemic fasciitis, which may only be distinguishable from a soft tissue neoplasm by the presence of distinct zonal characteristics such as peripheral cuffs of proliferating fibroblasts, or the absence of any atypical mitosis or nuclear atypia. Following this initial morphological assessment STSs are usually broadly categorised as epithelioid, pleomorphic, small round cell or spindle-cell in nature. Despite these specific phenotypic groups, microscopic similarities many different STS subtype's morphologies and cellular polymorphism often make this morphology alone inadequate to differentiate between specific STS subtypes. To address this issue molecular techniques including immunohistochemistry (IHC), fluorescence in-

situ hybridisation (FISH), polymerase chain reaction-based techniques and expression arrays are often implemented by histopathologists. In recent years these techniques have resulted in significant changes in the way STSs are classified, as well as the formal recognition of several new entities such as desmoplastic small round cell tumours and intimal sarcomas. Importantly they also key to help rule out certain non-mesenchymal tumours such as metastatic sarcomatoid carcinoma or melanoma, and to define a STS's exact mesenchymal cell lineage. Many STS subtype-specific mutations are now routinely targeted using the methods outlined above and include the overexpressed proteins STAT6 (solitary fibrous tumours) and MDM2 (dedifferentiated liposarcoma), and the chimeric genes *EWSR1-FLI1* (Ewing's sarcoma (May et al. 1993)), *FUS-DDIT3* (myxoid/round cell liposarcoma (Crozat et al. 1993)) and *SS18-SSX1/2* (Synovial sarcoma (Crew et al. 1995)). Unfortunately despite this and the progress made in the histopathological assessment of mesenchymal tumours in recent years, there remains a significant risk of STS subtype misdiagnosis which may be as high as 60% outside specialised STS centres (Lehnhardt et al. 2009). Furthermore, the diagnostic complexity of STSs and small number of known subtype specific alterations means that a significant proportion of cases remain unclassified, and are simply termed 'undifferentiated pleomorphic sarcomas' (UPS).

1.1.6 The management of soft tissue sarcomas

The principals of the management for primary and recurrent STSs are outlined below.

1.1.6.1 Grading, imaging and staging

Following a tissue diagnosis every STS must be formally graded to allow each patient to be staged appropriately. In the UK this is generally done using the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) grading system which categorises a STS as one of three grades based on the degree of its cellular differentiation, mitotic cells and necrosis (Trojani et al. 1984). Each newly diagnosed patient must also undergo complete local and systemic radiological imaging to allow for staging. This usually consists of a Magnetic Resonance Imaging (MRI) scan of the tumour (to accurately delineate its boundaries) and a Computer Tomography (CT) of the chest (to rule out metastases). Occasionally additional imaging may also be performed to include other atypical sites of metastases. This may include a whole body MRI and abdomino-pelvic CT in cases of myxoid liposarcoma, or a brain CT/MRI in alveolar soft part or clear cell sarcoma. Soft tissue sarcomas are classically staged according to the American Joint Committee on Cancer (AJCC)/International Union against Cancer staging system (Table 1.2). This system assigns a STS a stage based on its grade, location, size and the presence or absence of any distant disease, and is key in order to predict a patient's prognosis, and instigate appropriate management thereafter.

<table>
<tr><th colspan="2">Grade</th><th colspan="2">Tumour</th><th colspan="2">Node</th></tr>
<tr><td rowspan="2">Gx</td><td rowspan="2">Grade cannot be assessed</td><td>Tx</td><td>Tumour cannot be assessed</td><td>Nx</td><td>Regional lymph nodes cannot be assessed</td></tr>
<tr><td>T1</td><td><5cm largest dimension</td><td>N0</td><td>No evidence of node metastases</td></tr>
<tr><td>G1</td><td>Well Differentiated</td><td>T1a</td><td>Superficial to deep fascia</td><td>N1</td><td>Regional node metastases</td></tr>
<tr><td>G2</td><td>Moderately differentiated</td><td>T1b</td><td>Deep to deep fascia*</td><td colspan="2">Metastasis</td></tr>
<tr><td rowspan="3">G3</td><td rowspan="3">Poorly differentiated</td><td>T2</td><td>> 5cm in largest dimension</td><td rowspan="2">M0</td><td rowspan="2">No Evidence of metastases</td></tr>
<tr><td>T2a</td><td>Superficial to deep fascia</td></tr>
<tr><td>T2b</td><td>Deep to deep fascia*</td><td>M1</td><td>Distant metastases</td></tr>
<tr><th colspan="6">Stage</th></tr>
<tr><td>Ia</td><td>G1/X</td><td colspan="2">T1a/b</td><td colspan="2" rowspan="5">N0 /M0</td></tr>
<tr><td>Ib</td><td>G1/X</td><td colspan="2">T2a/b</td></tr>
<tr><td>IIa</td><td>G2/3</td><td colspan="2">T1a/b</td></tr>
<tr><td>IIb</td><td>G2</td><td colspan="2">T2a/b</td></tr>
<tr><td rowspan="2">III</td><td>G3</td><td colspan="2">T2a/b</td></tr>
<tr><td>Any G</td><td colspan="2">Any T</td><td colspan="2">N1 /M0</td></tr>
<tr><td>IV</td><td>Any G</td><td colspan="2">Any T</td><td colspan="2">N0-1 / M1</td></tr>
</table>

Table 1.2: The American Joint Committee on Cancer (AJCC)/International Union against Cancer grading system (Edge et al. 2010)**.** Both the parameters for the individual components of the staging system (Tumour/Grade/Tumour/Node/Metastasis) and the parameters for AJCC stage 1-4 tumours themselves are shown.

1.1.6.2 Surgery and histological margins

The keystone of non-metastatic STS treatment is surgery to remove the tumour in its entirety (Dangoor et al. 2016). To assess the adequacy of surgery and judge the need for any adjuvant treatment STS resection margins are classically described using the Enneking classification (Enneking et al. 2003). In this system surgical margins are described as intralesional, marginal, wide or radial depending on the plane of the surgical resection (Figure 1.2). In brief, a resection that enters the STS is termed intralesional, a resection that removes all of a STS but travels through its pseudocapsule (or reactive zone) is termed marginal, a resection that removes the STS and its pseudocapsule is termed wide, and a resection that removes the entire anatomical compartment containing a STS is termed radical. An alternative way to describe resection margins is to label them as R0, R1 or R2 depending on whether there is no evidence of tumour cells at the margin (R0), microscopic evidence of a tumour at the margin (R1) or gross tumour contamination at the margin (R2) (Dangoor et al. 2016). One key factor that must be considered when interpreting a positive (marginal / R1) surgical resection is whether the positive margin was planned or not. It is clear that local recurrence rates following a planned marginal/R1 STS resection (usually performed to preserve a key neurovascular structure (Figure 1.3) are significantly lower than those seen following an unintentional positive resection (O'Donnell et al. 2014). This combined with the excellent local control rates seen following planned marginal resections (Gerrand et al. 2001) has led to a fall in the number of amputations being performed for primary STSs in recent years to 4.1% (Smith et al. 2018).

1.1.6.3 (Neo)adjuvant radiotherapy

Radiotherapy is also often used to reduce the risks of local recurrence in case of large, high grade or incompletely excised STS. In the UK radiotherapy is usually administered as external beam therapy. In certain cases when the precise delivery of radiation is necessitated by the presence of key radiosensitive anatomical structures Intensity-Modulated Radiation Therapy (IMRT) or proton therapy may be considered. Although neoadjuvant (preoperative) and adjuvant (postoperative) radiotherapy result in comparable local control rates (O'Sullivan et al. 2002) each have their respective advantages and disadvantages. In short, neoadjuvant radiotherapy requires a lower radiation dose than adjuvant radiotherapy (reducing the risks long term complications) although results in an increased risk of wound complications than adjuvant treatment.

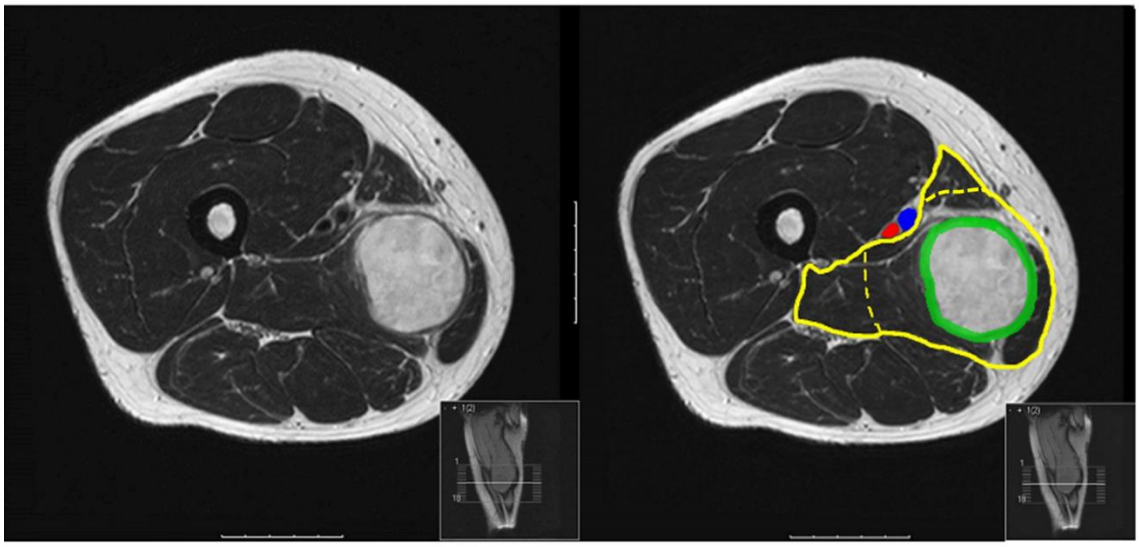


Figure 1.3: Axial MRI scan images showing a STS in the adductor compartment of the right thigh. The image on the right highlights the STS (green), femoral artery (red) femoral vein (blue) and the planes of dissection that would result in a radical (yellow line) or smaller wide (dashed yellow line adjustments) margin of resection as defined by Enneking.

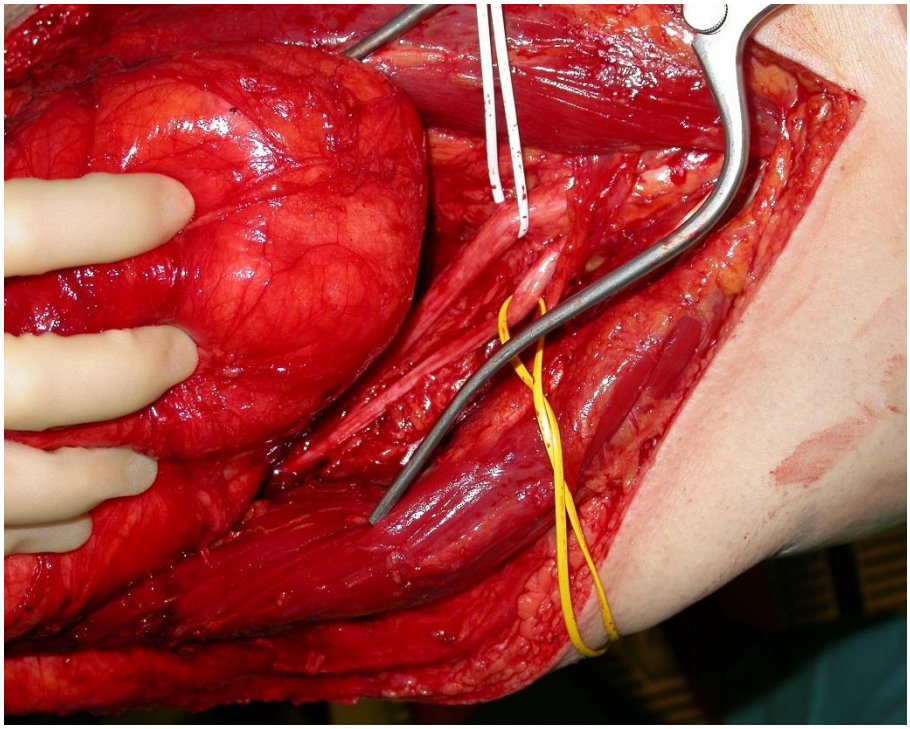


Figure 1.4: A clinical photograph of a large well differentiated liposarcoma in the posterior compartment of the thigh. As shown the tumour was intimately associated with the common peroneal and tibial nerves (shown in white and yellow vascular slings) and so a planned marginal resection with nerve preservation was performed.

1.1.6.4 (Neo)Adjuvant chemotherapy

Systemic chemotherapy is not routinely recommended for cases of non-metastatic STS in the UK (Dangoor et al. 2016). This is based on the currently available evidence which has failed to show any clear benefit from its use in terms of local control or long term survival rates (Nathenson & Sausville 2016; Woll et al. 2012). Despite this there is a growing recognition that certain STS subtypes may be more chemo-sensitive than others. As a result (neo)adjuvant chemotherapy may be considered in those individual patients with potentially chemo-sensitive, particularly high risk tumours (Dangoor et al. 2016). Examples include those with cardiac STSs, in whom local recurrence would have a particularly poor prognosis.

1.1.6.5 Local and metastatic STS recurrence

Unfortunately following treatment a significant proportion of high grade STSs will recur either locally (17%) or with metastatic disease (24%) (Trovik 2001; Sabolch et al. 2012). In the absence of simultaneous metastases local recurrence should be treated with curative intent using the same surgical principals as a primary STS. Pulmonary metastatic recurrence carries an extremely poor prognosis with a 5 year survival of just 15% (Gunar K. Zagars et al. 2003). Although a thoracic metastasectomy may be considered in those patients with a small number of lesions and an indolent disease course, there is no clear evidence that this improves long term survival (Dangoor et al. 2016). Other less invasive techniques that may be adopted for pulmonary metastases include radiofrequency ablation and stereotactic ablative radiotherapy, but again significant long term benefits on survival are yet to be reported (Nakamura et al. 2009).

1.1.6.6 Advanced metastatic STS disease

Generally once STSs have metastasised the approach to treatment becomes palliative. Although surgery and radiotherapy may be adopted to remove or reduce the size of a symptomatic lesion, systemic chemotherapy is the most commonly used treatment in these cases.

1.1.6.6.1 Chemotherapy

The rarity of STSs has hindered the development of standardised chemotherapy protocols based on quality, randomised evidence. Unless contraindicated the majority of metastatic STS patients in the UK are initially treated using doxorubicin (an anthracycline) in isolation or in combination with ifosfamide (an alkylating agent) (Tacar et al. 2013; Furlanut & Franceschi 2003). Unfortunately a limited clinical response is seen in the majority of STS patient to these regimes, with objective improvements in disease volume seen in just 14%-35% (single agent doxarubicin) and 26% (doxarubicin and ifosphamide) of cases, and overall median survival periods of just 12.8 and 14.3 months reported (Tierney 1997; Ravi et al. 2015; Schöffski et al. 2014; Judson et al. 2018). Several chemotherapy regimens are routinely offered as second line treatment for metastatic STS patients including gemcitabine (a nucleoside analog (Mini et al. 2006)) and docetaxel (a taxane that impairs centrosome and spindle microtubule function (Dumontet & Jordan 2010)), ifosfamide or trabectedin (another alkylating agent). Again, despite their assorted effects the objective response rates seen to these regimes are also disappointingly low, with response rates of just 16% (Maki et al. 2007) - 20% (Seddon et al. 2017), 8% (Van Oosterom et al. 2002) and 9.9% (Demetri et al. 2016) reported respectively.

1.1.6.6.2 Targeted Agents

As a better understanding of the molecular mechanisms involved in sarcomagenesis develops it is hoped that targeted therapies will play a growing role in the management of STSs (Judson 2010; Steppan et al. 2017). Several families of drugs are currently under investigation including tyrosine kinase inhibitors (of which pazopanib (van der Graaf et al. 2012) and regorafenib (Mir et al. 2015) have both showed particular promise), mammalian target of rapamycin (mTOR) pathway inhibitors and immunotherapy agents such as anti-PD-1 agents (Steppan et al. 2017). Unfortunately despite this work the improvements in patient outcome seen in several other malignancies has not been mirrored in STS patients. In the case of certain immunological treatments such as PD-1 inhibitors this is likely a result of the low number of somatic SNVs seen in STSs, as it is these mutations which drive T cell activation via the transcription of abnormal cell surface proteins in those tumours sensitive to these treatments, and regulation by the immune system in general.

1.1.7 STS outcome and challenges

Soft tissue sarcomas are a challenging group tumours with a relatively poor prognosis compared to many others solid malignant tumours. Long term STS patient outcome is heavily dependent on disease stage with 5 year survival rates of 10%, 52%, 72% and 86% reported for stage 4, 3, 2 and 1 tumours respectively (Edge et al. 2010). For the majority of STS subtypes there has been no significant improvement in these survival figures in the last 25 years (Public health England 2011). Four key barriers must be overcome to address this:

1.1.7.1 Delays in diagnosis

A significant delay can often be retrospectively identified between the initial presentation of many STS patients and their referral to a specialist STS centre (Brouns et al. 2003). A direct consequence of this is that the average size of a STS when diagnosed increases (currently 10cm in the UK) (Grimer 2006). This worsens a patient's prognosis significantly, with every 1cm increase in a STS's size at diagnosis reducing their chance of cure by 3-5% (Grimer & Briggs 2010). Any delays in referral will also invariably result in a higher proportion of patients already having metastatic disease at diagnosis (a dismal prognostic factor as discussed in section 1.1.6.5). To try to reduce any delays in patient referral efforts have been made to increase STS awareness in a primary care setting. Despite these, the great rarity of STSs compared to their benign soft tissue tumour counterparts (1:100) means addressing them will be challenging.

1.1.7.2 Incomplete understanding of STS genomics

Despite the work outlined in section 1.1.4 there remains an incomplete understanding of the genomic characteristics of STSs, and in particular which mutations are most important to the process of STS development and progression (termed 'Driver mutations') (Vogelstein et al. 2013). This is an issue compounded by STSs' low prevalence, genomic heterogeneity and the wide variety of mutations that they possess. If a better understanding of these characteristics was developed the molecular mechanisms central to sarcomagenesis may be uncovered, which themselves may provide novel therapeutic targets.

1.1.7.3 Lack of Biomarkers

The current post-operative follow up regime for STS patients that undergo attempted curative treatment consists of serial clinical examinations to detect local recurrence and chest radiographs to detect pulmonary metastases. Unfortunately despite this surveillance local STS recurrence is often so extensive when diagnosed that to secure acceptable surgical margins an amputation is required (needed in 22% of cases (Trovik 2001)). In addition, metastatic recurrence is often too extensive when diagnosed to facilitate any attempts at curative treatment. Currently no sensitive biomarkers of STS tissue exist. If developed and successfully incorporated into STS patients' follow up these may allow local and metastatic STS recurrence to be diagnosed earlier. This may in turn reduce the need for as many disabling surgeries for local recurrence, and increase the number of metastatic patients that can be offered curative treatment.

1.1.7.4 Ineffective systemic treatments

As outline above (see section 1.5.6) the current systemic therapies available for STS patients are ineffective in the majority of cases. An ideal approach to addressing this issue would be to develop and run prospective multi-centre, randomised studies testing new therapies, highlighted as potential candidates for use in STS patients based on a robust understanding of sarcomagenesis. In the absence of this understanding however, a more realistic approach at present may be to collate all of the data currently available from previous phase II/III trials in a meta-analytical way. Creating more powerful subtype, patient and treatment specific datasets in this way may subsequently allow therapies already available to be offered to STS patients in a more targeted manner. Finally, a third approach may be to adopt a truly 'patient specific' strategy by testing a broad range of therapies (potentially including repurposed drugs) on cultured tumour samples collected from individual STS patients. Treatment in this scenario could then be offered to individual patient based on the responses shown by their tumour in culture.

1.2 Circulating nucleic acids

In recent years increasing emphasis has been placed on the importance of the early diagnosis of cancer patients, and the ability to monitor patients with more advanced disease throughout a course of treatment or follow up. This has led to a sustained drive to identify novel circulating cancer biomarkers, and considerable interest in the field of circulating nucleic acids (Thierry et al. 2016).

1.2.1 Circulating cell free DNA

Circulating cell free DNA (cfDNA) is defined as extracellular nucleic acids that circulate freely in the blood stream. In healthy individuals cfDNA fragments are approximately 165 base pairs (bp) in length (Underhill et al. 2016), although in cases of malignancy may shorten considerably to <100bp (Mouliere et al. 2011). Based on this length (which is consistent with the process of internucleosomal cleavage (Nagata et al. 2003)) and the higher levels of cfDNA found in apoptotic environments compared with necrotic or phagocytic ones (Wang et al. 2003) circulating nucleic acids are thought to predominantly originate from apoptotic cells. Despite this the exact mechanism of cfDNA release is yet to be elucidated, with some evidence still suggesting that it is actively secreted (Stroun et al. 2001) or released following necrosis (Jahr et al. 2001). After entering the circulation cfDNA is rapidly excreted, with clearance times of 30-120 minutes reported (Lo et al. 1999; Fatouros et al. 2010). Although the exact route via which this occurs is again unknown early evidence suggested that the kidneys (and to a lesser extent the liver and spleen) play a role (Tsumita & Iwanaga 1963). Circulating cfDNA levels have been shown to correlate with patient outcome in many pathologies including sepsis (Rhodes & Cecconi 2012), trauma (Gögenur et al. 2017), systemic lupus erythematosus (Hendy et al. 2016) and even following organ transplant (where high levels appear predict organ rejection (Gielis et al. 2015)). Although these are all interesting findings, the field of oncology is where cfDNA has shown its most exciting potential as a biomarker.

1.2.2 Cell free DNA levels and cancer

Cell free DNA was first identified in cancer patients over 45 years ago (Koffler et al. 1973). Soon after this cfDNA levels were found to be significantly elevated in many

cancers including lymphoma, lung, breast, genitourinary, colorectal, head and neck and central nervous system malignancies, particularly following the development of metastatic disease (Zaher et al. 2013; Leon et al. 1977; Spindler et al. 2015; Shao et al. 2001; Fournié et al. 1995). Subsequent studies investigating breast, lung and colorectal cancers have gone onto show that cfDNA levels correlate well with multiple clinical prognostic factors including tumour size, regional metastases and disease stage (Shao et al. 2001; Sozzi et al. 2001; Fournié et al. 1995; Spindler et al. 2015). Furthermore, cfDNA levels in numerous cancer types have been shown to accurately reflect response to various oncological treatments including radiotherapy, chemotherapy and targeted agents (Leon et al. 1977; Feng et al. 2013; Agostini et al. 2011; Dowler Nygaard et al. 2014). Taken together these characteristics suggest that cfDNA levels hold much potential as a diagnostic biomarker or means to monitor tumour burden throughout a course of treatment. One obvious issue that complicates monitoring cfDNA levels in cancer patients is the effect of nucleic acids shed from extra-tumoural sources. This component of cfDNA will invariably increase or decrease in response to physiological changes unrelated to a cancer patient's tumour, and so may potentially cloud any subtle changes in tumour burden. To overcome this issue and obtain a more accurate reflection of tumour behaviour many groups have shifted their attention to the analysis of cfDNA shed solely from tumoural cells - a component of cfDNA termed circulating tumour-derived DNA (ctDNA) (Sorenson 2000; Sorenson et al. 1994; Daniotti et al. 2007).

1.2.3 Circulating tumour derived DNA

By identifying certain cancer associated genetic or epigenetic alterations researchers have developed a way to isolate and characterise ctDNA. Multiple alterations have been targeted for this purpose including DNA methylation (Kawakami et al. 2000; Usadel et al. 2002; Grady et al. 2001; Wong et al. 1999), microsatellite alterations (Fujiwara et al. 1999; Sozzi et al. 1999), SNVs (Vasioukhin et al. 1994; Sorenson et al. 1994) and more recently somatic CNAs (Soave et al. 2017). By using this approach to negate the effects of non-tumoural derived cfDNA, researchers have been able to show that even small changes in tumour size or disease stage can be identified by tracking and characterising ctDNA in several malignancies including colorectal, breast and pancreatic cancer (Kopreski et al. 1997; Kopreski et al. 2000; Nakauchi et al. 2016; Yamada et al. 1998).

1.2.4 Advantages of ctDNA analysis

1.2.4.1 Ease of access

A tumour's genetic characteristics may have direct consequences on a patient's diagnosis, prognosis and treatment options (Heinrich et al. 2003). Currently to access this information an invasive tumour biopsy is required, which is an expensive, time consuming intervention that carries significant risks for the patient involved (Boskovic et al. 2014). The analysis of ctDNA overcomes these risks by allowing a tumour's genetic characteristics to be accessed with the use of a simple blood test alone. This also allows these characteristics to be assessed repeatedly throughout a course of treatment or follow up (when repeated tumour biopsies are unlikely to be practical) and so monitor for any dynamic changes in a tumour's genome which may reflect the development of secondary treatment resistance (Diaz et al. 2012).

1.2.4.2 Intra-tumoural Genetic Heterogeneity

During the course of a tumour's growth the process of clonal evolution results in the development of numerous discrete populations of malignant cells with distinct genetic alterations (Fisher et al. 2013). The genomic variation present between these cellular populations (known as subclones) is a phenomenon termed Intra-tumoural Genetic Heterogeneity (IGH) (Gerashchenko et al. 2013). As well as complicating the systemic treatment of malignant tumours (Fisher et al. 2013), significant IGH may result in cells containing clinically important mutations being 'missed' during the process of conventional tissue biopsy. Due to its circulatory nature, ctDNA overcomes this problem by reflecting the mutations present in all of a tumour's subclones (Perdigones & Murtaza 2017). This is key advantage, and has highlighted ctDNA as an exciting potential tool to help guide clinicians when making second line treatment choices following the development of first line treatment resistance.

1.2.4.3 Wide dynamic range

A key feature of ctDNA is the wide dynamic range through which its levels alter in response to change in disease burden. This allows even small changes in disease burden to be identified (Kopreski et al. 1997; Kopreski et al. 2000; Nakauchi et al. 2016), and is a key advantage over other circulating biomarker such as circulating tumour cells (which make up a tiny proportion of all circulating cells (Ross et al. 1993)) and certain proteins already used in a clinical setting such as CA 15-3 (Dawson et al. 2013).

1.2.4.4 Real time sensitivity

Cell free DNA is rapidly cleared from the circulation with some studies reporting a half-life of just 16 minutes (Lo et al. 1999). Although this makes strict adherence to standardised plasma collection protocols key, it also provides ctDNA with an unparalleled ability to reflect a tumour's characteristics in real-time. Despite their rapid clearance in-vivo, circulating nucleic acids stabilise rapidly once they are isolated in plasma (Barrett et al. 2014). As a result ctDNA has the additional advantages of being amenable to storage and/or transportation prior to investigation, with no detrimental effects on downstream analysis

1.2.5 Circulating biomarkers and STSs

To date no recognised circulating biomarkers of STS tissue are available for clinical use. Despite this circulating nucleic acids, tumour cells and other potential biomarkers such as exosomes have generally remained under investigated in STS patients.

1.2.5.1 Circulating tumour cells

Cells that detach from a primary malignant tumour and enter the circulating are known as circulating tumour cells (CTCs). These CTCs are thought to initiate metastases and have been highlighted as a potential source of tumoural genomic information and novel cancer biomarkers (Danila et al. 2011). The analysis of CTCs provides the potential advantage of offering up transcriptomic, proteomic and metabolomic data (Kidess & Jeffrey 2013). Despite this, although several groups have successfully identified CTCs in STS patients (Chinen et al. 2014; Chang et al. 2015), several issues surround their analysis which have generally hindered their detailed characterisation (Tellez-Gabriel et al. 2016). Firstly, analysing CTCs is technically challenging, in part as a result of their very low concentration in the circulation (Hong & Zu 2013). Secondly, CTCs are extremely diverse in terms of their morphology and genomic make up. This not only complicates the process of establishing robust assays, but also means that the small number of CTCs usually isolated from a patient's blood sample are unlikely to contain a tumour's entire complement of genetic alterations. Finally, only a small number of mesenchymal specific cell surface markers have been identified (Satelli et al. 2014; Ward et al. 2015). This greatly hampers the ability to delineate true STS CTCs from other rare cells also found in the circulation such as endothelial cells, stem cells or atypical non-tumoural cells.

1.2.5.2 Exosomes

Exosomes are extracellular vesicles that contain proteins, lipids and nucleic acids. They were initially highlighted as a potential source of circulating cancer patient biomarkers following the discovery that they are released in significantly higher numbers from tumour cells than normal cells (Rabinowits et al. 2009). Despite evidence that exosome levels may correlate with prognosis in certain malignancies including ovarian cancer (Taylor & Gercel-Taylor 2008) we are unaware of any work investigating them as potential biomarker in cases of STS. Considering that exosomes are involved in many key processes involved in sarcoma development and progression (Min et al. 2016) it is likely that this paucity of work will be addressed as the field expands.

1.2.5.3 Circulating nucleic acids

Despite the advantages of using circulating nucleic acids to monitor cancer patients the current body of literature describing cfDNA/ctDNA in STS patients consists of just 2 case reports. The first of these was performed in 2015 by Butler et al who used whole exome sequencing to analyse matched tumour and plasma samples collected from a single case of advanced metastatic intimal sarcoma (Butler et al. 2015). After noting significantly elevated cfDNA levels (63ng/ml plasma) the group successfully identified 48 exonic somatic SNVs in the tumour, of which 47 were also identified in the patient's plasma confirming the presence of ctDNA. The second case report was performed in 2017 by Namlos et al who used targeted NGS to analyse a tumour sample and 3 plasma samples collected from a patient with a spindle cell sarcoma undergoing an attempted curative surgical resection. After noting persistently elevated cfDNA levels in all 3 plasma samples collected (pre op, 110ng/ml plasma: 3/42 days post op, 76-316 ng/ml) the group went onto identify 8 somatic mutations in the patient's tumour, of which 6 were present in all 3 plasma samples collected. Six weeks following surgery the patient analysed developed widespread metastatic disease before succumbing to their disease 7 weeks later. Based on their findings Namlos et al proposed that the persistence of microscopic disease post-operatively explained this early relapse, and that ctDNA may hold a future role as a marker of such small volume disease (Namløs et al. 2017).

1.3 Telomeres and cancer

1.3.1 Protective mechanisms against cancer

Over time extrinsic factors and intrinsic processes mean that cells accumulate DNA damage, increasing their risk of developing oncogenic mutations and cancer (Gaillard et al. 2015; Best 2009). Under normal circumstances cells use several strategies to protect against this. These include mechanisms they activate in response to DNA damage itself, but also in response to the progressive shortening of their telomeres during cell division.

1.3.2 DNA damage response

To safeguard genomic integrity normal cells initiate an automated response to DNA damage known as the 'DNA damage response' mechanism (DDR) (Giglia-Mari et al. 2011). This DDR initially triggers arrest of the cell cycle in response to DNA damage. In the event of severe damage not amenable to repair, to avoid genomic instability a variety of downstream signalling pathways are subsequently activated that either trigger cell death or senescence. In the case of less severe damage one of five different DNA repair mechanisms is activated. To repair DNA single strand breakages three potential mechanisms may be activated. These are the Base Excision Repair mechanism (triggered by small lesions that do not significantly alter the DNA double helix structure), the Nucleotide Excision Repair mechanism (most active against larger DNA helix-distorting lesions) and the DNA mismatch repair mechanism (used to repair DNA replication errors and small insertions and deletions). In response to DNA double-strand breaks (DSBs) one of two other repair mechanisms are activated depending on the stage of the cell cycle when damage is detected. During the S phase the Homology directed repair mechanism is activated, whilst during the remainder of the cell cycle the less accurate, error prone non-homologous end-joining pathway (NHEJ) is most active (Mao et al. 2014). By repairing DNA damage, collectively these pathways and the DDR provide a key protective mechanism against cancer. This is highlighted by the increased incidence of malignancy seen in individuals with inherited defects in their DDR (Hoeijmakers 2001).

1.3.3 Telomere biology

In every human cell DNA is packaged as linear chromosomes. Although storing DNA in this way promotes genetic variability, it also presents two key problems. First, the inability of DNA polymerase to replicate the very terminal ends of linear DNA (a problem termed the 'end replication problem' (Olovnikov 1973)) means linear

chromosomes are susceptible to progressive shortening over time. Second, if the ends of linear chromosomes are left unprotected they run the risk of mimicking abnormal double strand DNA breakages, and activating a cell's DDR pathway as a result.

To overcome these issues specialised DNA-protein complexes called telomeres can be found capping the very terminal region of linear chromosomes. These telomeres protect coding sequences from degradation during cellular replication, whilst also differentiating chromosome ends from double strand DNA breakages preventing the formation of abnormal chromosome fusions via NHEJ (Xu et al. 2013; Palm & de Lange 2008). Telomeres also protect again genomic instability by acting as an internal lifespan for the cells they are found in. As a cell repeatedly divides the end replication problem means their telomeres progressively shorten. After a certain number of cell divisions (known as a cell's 'Hayflick limit' (Hayflick & Moorhead 1961)) this telomeric shortening triggers a cell to enter replicative senescence through the actions of the tumour suppressor *TP53* (Artandi & Attardi 2005). Although cells in this senescent state remain metabolically active they cease to divide. As a result this proliferative checkpoint (known as M1) protects against the creation of daughter cells with significantly damaged DNA at risk of malignant transformation.

1.3.4 Telomere Maintenance Mechanisms

A key hallmark of cancer cells is their ability to undergo uncontrolled cell division. To do this they must first acquire inactivating mutations in the key genes involved in regulating the M1 checkpoint (Figure 1.4). Next in order to bypass a second proliferative checkpoint known as cellular 'crisis' or M2 they must also develop the ability to maintain telomere length (Lundblad & Szostak 1989). This is facilitated by one of two 'Telomere Maintenance Mechanisms' (TMMs) - the action of the enzyme Telomerase or the less well understood Alternative Lengthening of Telomeres (ALT) mechanism.

1.3.4.1 Telomerase:

Telomerase is a reverse transcriptase enzyme with two main functional units - a ribonucleic acid (RNA) template called human telomerase RNA (hTR) and a protein subunit called human telomerase reverse transcriptase (hTERT) (Meyerson 2000). The DNA component of human telomeres consists of simple tandem repeats predominantly consisting of the sequence 5'-TTAGGG-3'. These repeats are predominantly double

stranded other than a small single strand region known as the 3' overhang found at the terminal portion of a telomere's 3' G-rich end (Henderson & Blackburn 1989). Human telomerase RNA contains a template region of 11 nucleotides that is complementary to human telomeres ($TTAGGG_n$). To begin telomerase mediated telomeric lengthening hTR partially hybridises to the 3' overhang of the telomere undergoing lengthening. Following this the hTERT component of telomerase synthesises a new DNA strand complementary to portion of hTR that remains unbound to telomeric DNA, effectively lengthening the telomere involved. Following this hTR detaches from its original binding site and translocates to the newly synthesised 3' end of telomere so the process can be repeated. After suitable telomeric extension has occurred, a DNA primase synthesises an RNA primer close to the terminal 3' end of the newly synthesised DNA strand. DNA polymerase α (the enzyme responsible for initiating DNA replication) then finally uses this primer to generate a DNA strand complementary to the telomere's new single stranded extension, leaving a small single stranded region at the telomere's most terminal end (a new 3' overhang) (Lee et al. 2014).

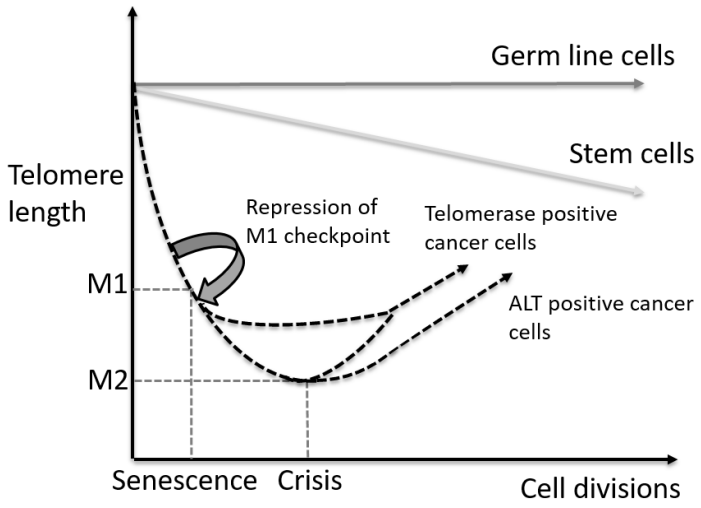


Figure 1.5: Graphic representation of telomere length following repeated cell divisions. Germ line cells, stem cells, telomerase positive and ALT positive cancer cells with incompetent M1 checkpoints are all shown.

1.3.4.2 Alternative Lengthening of Telomeres mechanism

The observation that some dividing cells can maintain their telomeres' lengths in the absence of telomerase led to the discovery of a telomerase-independent TMM called the Alternative Lengthening of Telomeres mechanism (ALT) (Bryan et al. 1995; Bryan et al. 1997). Although an exact understanding of ALT is yet to be elucidated, the high level of genome instability seen in ALT positive cells suggests it is activated at some stage after a cell has entered crisis. The rapid changes in telomere length seen in ALT positive cells also suggests that ALT is a mechanism dependant on a DNA recombination-like process (Murnane et al. 1994). Several mechanisms have been proposed for how telomeric DNA may be used as a copy template during ALT (Figure 1.5). First, strand invasion and copying may occur between telomeres on sister chromatids or different chromosomes (a proposal supported by the high levels of inter-chromosomal and sister chromatid telomere exchanges seen in ALT positive cells (Varley et al. 2002; Bailey et al. 2004).) Second, a telomere may loop back and strand invade itself allowing it to use regions of its own sequence to code from (Muntoni et al. 2009). Third, a telomere may use extrachromosomal telomeric DNA fragments as a copy template - a mechanisms supported by the high levels of extra chromosomal telomeric DNA found in ALT positive cells (Henson et al. 2009)).

1.3.4.3 Telomere maintenance mechanisms in soft tissue sarcomas

Although the incidence of telomerase activity varies in those sarcoma cohorts previously analysed, with the exception of one study (Aogi et al. 2000) telomerase appears to be activated in the minority of STSs overall (7-38%) (Lauer et al. 2002; Yoo & Robinson 2000; Johnson et al. 2005; Costa et al. 2006; Schneider-Stock et al. 1999). ALT positive cells have several well described phenotypes including heterogeneous telomere lengths (Bryan et al. 1995), high levels of extra-chromosomal circular and linear telomeric DNA (Pickett et al. 2009) and specific structures created by the co-localisation of promyelocytic leukaemia nuclear bodies and telomeric DNA known as ALT associated PML bodies (APBs) (Yeager et al. 1999). These features are rarely identified in carcinomas (breast 4%, uterine 1%, oesophageal 1%, biliary 2%, renal 10%, hepatic tumours 8%) suggesting that ALT is rarely activated in these tumours (Subhawong et al. 2009; C. M. Heaphy et al. 2011). In contrast they appears more commonly in malignant neurological tumours such as glioblastoma multiforme (25%, (Hakin-Smith et al. 2003)) and STSs. Here, the incidence of ALT varies significantly between different STS

subtypes with reported figures including 63% of undifferentiated pleomorphic sarcomas 53% of leiomyosarcomas, 33% of epithelioid sarcomas, 24%–26% of liposarcomas (Costa et al. 2006; Jeyapalan et al. 2008; Lee et al. 2015), 14% of fibrosarcomas and 11%–28% of angiosarcomas (J. Y. Liau et al. 2015; Henson et al. 2005; C. M. Heaphy et al. 2011).

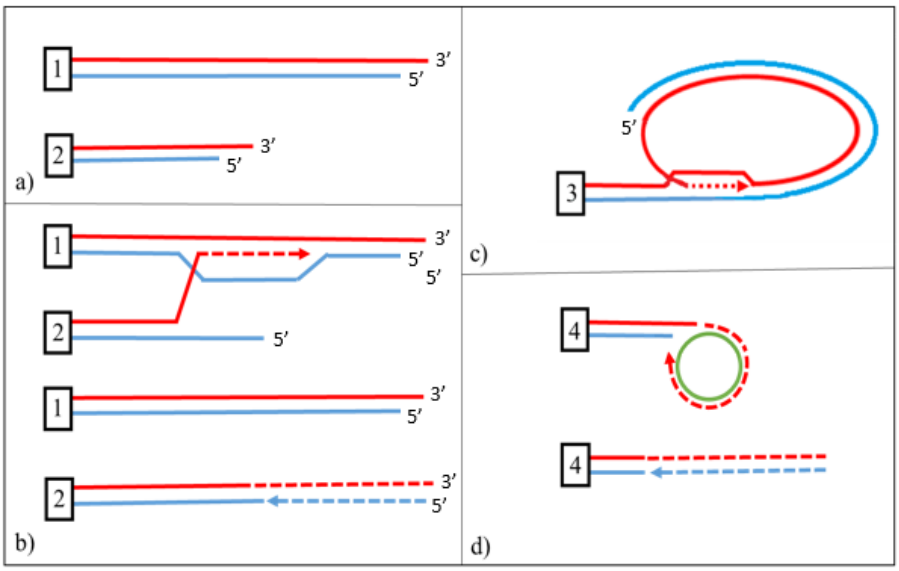


Figure 1.6: Diagrammatic representations of the possible mechanisms of telomere lengthening by ALT. 1.5a shows 2 telomeres numbered 1 (long) and 2 (short) from 2 separate chromosomes or sister chromatids. 1.5b shows telomere 2 strand invading into telomere 1 and undergoing extension using the donor telomere as a template. The completion of telomere 2's lengthening by 2nd strand synthesis is also shown. 1.5c shows another single telomere (numbered 3) undergoing lengthening by the process of intra-telomeric copying via the formation of a t-loop. 1.5d shows how an individual telomere (numbered 4) can also undergo lengthening by using a circular piece of extrachromosomal telomeric DNA (green) as a copy template followed by 2nd strand synthesis.

1.3.4.4 Genetic mutations associated with Telomere Maintenance Mechanisms

The need for cancer cells to maintain telomere length to achieve replicative immortality theoretically makes the genomic mutations that drive telomerase and ALT activity interesting candidates as potential biomarkers. In the case of telomerase positive tumours these mutations may include activating substitutions found in the promoter region of *TERT* (the gene that encodes for hTERT - the rate limiting component of telomerase) (Heidenreich et al. 2014). Although the regulators of ALT are less well understood, Alpha-thalassemia/mental retardation X-linked (*ATRX*) and Death-domain associated protein (*DAXX*) have both been identified as key inhibitors of the process (Heaphy et al. 2011; Schwartzentruber et al. 2012). As a result deactivating alterations in either gene may act as effective biomarkers in ALT positive tumours.

1.4 Aims

As outlined in section 1.1.7 several key barriers exist to improving STS patient outcome. These include an incomplete understanding key genetic alterations that drive STS development and progression and a lack of sensitive circulating biomarkers of STS tissue.

To help develop the current understanding of the genetic mutations that contribute to the development and progression of STSs the first aim of this project was to perform a detailed genomic analysis of a group of primary STSs using NGS technology and bioinformatic pipelines pursuing SNVs, small indels and CNAs.

To try to begin to address the lack of circulating STS biomarkers the second aim of this project was to investigate the circulating nucleic acid characteristics of a cohort of metastatic STS patients. To facilitate this, quantitative polymerase chain reaction (qPCR) was used to measure total cfDNA levels in these patients, whilst a custom designed targeted Iontorrent NGS panel was used to search for evidence of ctDNA in the form of circulating mutations located in well-known cancer or TMM associated genes.

Next, to further investigate the potential clinical utility of profiling circulating nucleic acid characteristics in STS patients the third aim of this project was to investigate the cfDNA and ctDNA characteristics of a group of non-metastatic STS patients undergoing attempted curative treatment. For this analysis patients' cfDNA ctDNA characteristics were investigated throughout their treatment and follow up using a combination of qPCR, tNGS and mutation/patient specific digital droplet PCR assays.

Chapter 2: Methods

2.1 Ethics and sponsorship

This project was performed according to a protocol approved by the NHS National Research Ethics Service (NRES) Committee North East - Newcastle & North Tyneside 1 (study title: Genetic profile and telomere characteristics and of high-grade soft tissue sarcomas, REC reference: 14/NE/1192, IRAS project ID: 141820) and was conducted in accordance with the Declaration of Helsinki throughout. The University Hospitals of Leicester NHS Trust sponsored the project, which was also publically registered on www.ClinicalTrials.gov (Identifier: NCT02547376). Patient recruitment performed at the University Hospitals of Leicester and Nottingham University Hospitals NHS Trusts.

2.2 Patient eligibility and enrolment

Two groups of adult patients were enrolled for analysis – groups 1 and 2.

Group 1 comprised of patients with biopsy proven STSs and metastatic disease diagnosed by the East Midlands Sarcoma Service (EMSS). No exclusion criteria were placed on the anatomical location of these patients' primary tumours, and this group included patients with retroperitoneal, trunk wall and extremity primary tumours. These patients provided tissue and plasma samples for the analysis described in chapter 4.

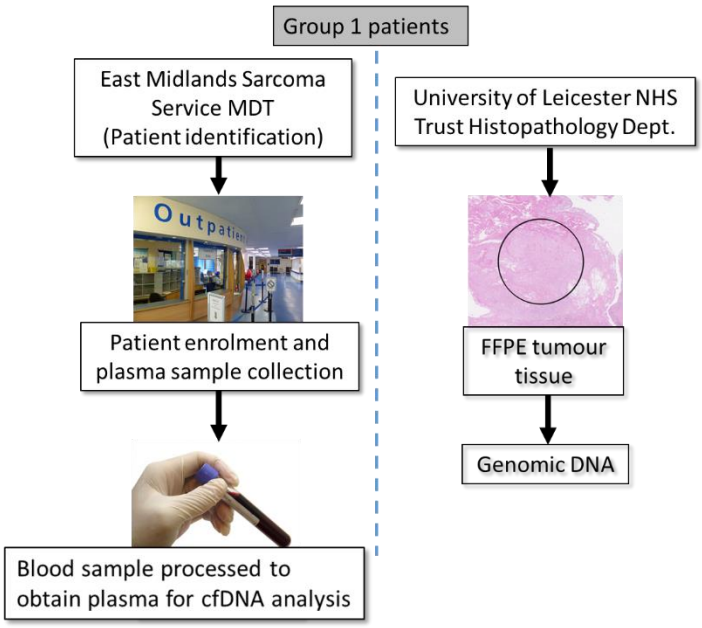
Group 2 comprised of non-metastatic STS patients scheduled to undergo attempted curative surgical resections. Patients with primary retroperitoneal STSs were excluded from this group for logistical reasons, leaving just individuals with extremity or trunk wall primary tumours. Patients in this group provided samples for the analysis in chapters, 1, 4 and 5.

All patients were initially identified at the EMSS MDT and subsequently approached at their next routine clinic outpatient appointment to be formally consented for enrolment (Figure 2.1). Table 2.1 shows the project's full inclusion and exclusion criteria.

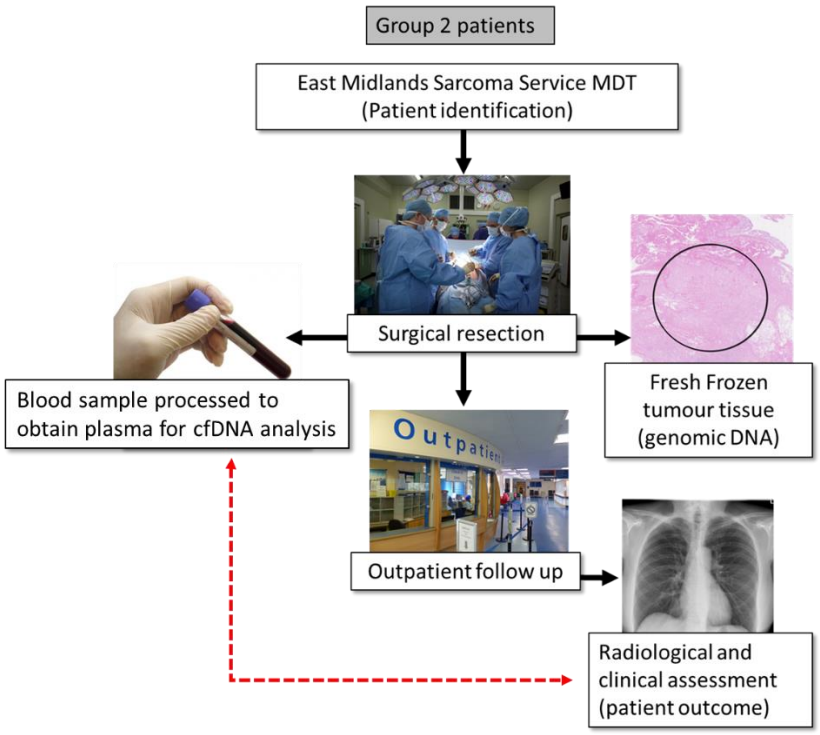
2.3 Clinical follow up

Patients in group 2 were followed up in accordance with national recommendations (Dangoor et al. 2016). In general this consisted of one early outpatient appointment following surgery to ensure satisfactory wound healing followed by serial 3 monthly

appointments. At each 3 monthly follow up appointment patients were examined looking for any evidence of local and regional recurrence and chest radiographs performed looking for metastatic recurrence in the lung.



2.1a)



2.1b)

Figure 2.1 Tissue and plasma collection from STS patients. 2.1a represents those patients in group 1 who provided consent for the analysis of one prospectively collected whole blood sample and FFPE STS tissue blocks collected at the time of diagnosis where available. 2.1b represents patients in group 2. These patients provided consent for the analysis of tumour tissue and whole blood samples collected intraoperatively, and blood

samples subsequently collected serially throughout follow up. This longitudinal design allowed any circulating nucleic acid characteristics identified to be correlated with patients' clinical outcomes (red dashed arrow).

	Inclusion Criteria	Exclusion Criteria
Group 1	Patient with biopsy proven metastatic STS	Patient unable to provide informed consent
		Patient aged under 18
Group 2	Patient with non-metastatic biopsy proven STS	Patient presenting with local or distant STS recurrence
	Patient undergoing attempted curative surgical resection (+/adjuvant therapy)	Patient with retroperitoneal STS
	Patient management and follow up planned under the EMSS	Patient unable to provide informed consent
		Patient aged under 18

Table 2.1: Enrolment criteria for STS patients. Enrolled patients in group 2 were removed from the project when they 1) developed disease recurrence, 2) were lost to follow up due to a relocation or 3) voluntarily withdrew

2.4 Sample collection

2.4.1 Group 1

2.4.1.1 Whole blood: Patients in group 1 provided one 20ml sample of whole blood for analysis collected at the time of enrolment.

2.4.1.2 Formalin fixed paraffin embedded (FFPE) STS tissue: FFPE tumour blocks were retrospectively sourced for group 1 patients from the relevant NHS histopathology departments. DNA was extracted from 1mm cores taken from these blocks from regions highlighted as representative of viable STS tissue by a specialist sarcoma histopathologist.

2.4.2 Group 2

2.4.2.1 Whole blood: Patients in group 2 provided whole blood samples intraoperatively (immediately prior to tumour removal) and at each of their subsequent outpatient follow up appointments. As outlined in section 2.3 standard STS follow up dictated that the first of these sample was collected around 2 weeks following surgery and then approximately 3 monthly thereafter.

2.4.2.2 Fresh frozen tumour tissue: Resected specimens were macroscopically reviewed by a consultant specialist sarcoma histopathologist immediately after they were removed from enrolled patients (see Figure 2.2 for example specimen). During this assessment up to 5 tissue samples felt to be representative of viable STS tissue were collected from a variety of regions within the tumour. These samples were immediately snap frozen using liquid nitrogen and stored at -80^{o}C until released for analysis. Prior to this analysis tumour samples were also microscopically assessed by a histopathologist to ensure they were neither necrotic nor misrepresentative of STS tissue. To facilitate this one 5µm Hematoxylin and Eosin (H+E) section was taken from each sample using a microtome in a cryostat to reduce the number of required freeze thaw cycles.

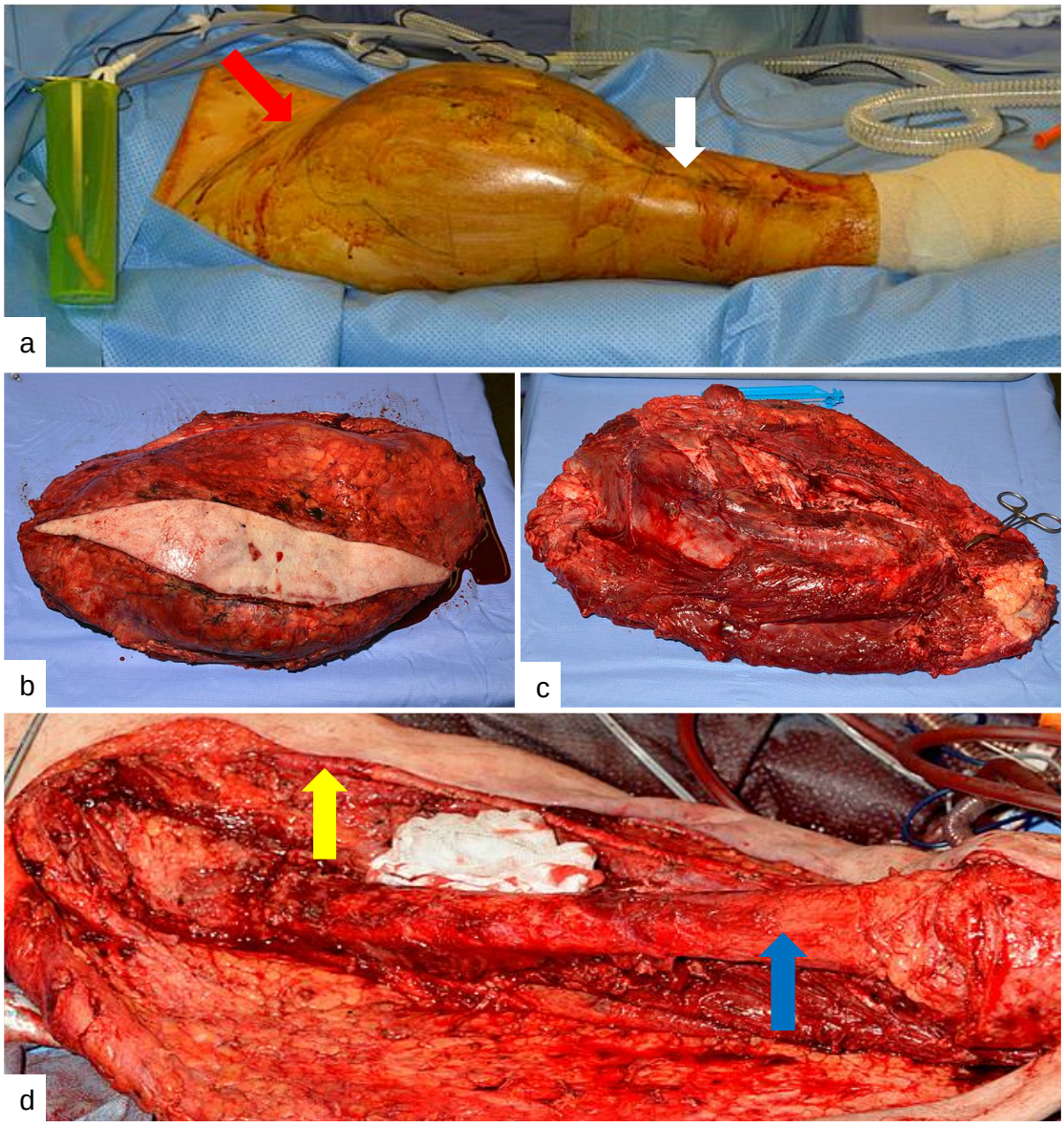


Figure 2.2: Intraoperative clinical photographs of a STS wide resection. 5.20a shows the participant 033's right leg prepared for surgery highlighting the large size of their STS. The patient is positioned supine and the patella (kneecap) (white arrow) and inguinal (groin) crease (red arrow) are highlighted for orientation. 5.20b and c show the resected tumour viewed anteriorly and posteriorly. The resected tumour was 3289cm^3. 5.20d shows the soft tissue defect left after the tumour's resection. The femur (blue arrow) and femoral artery (yellow arrow) are highlighted.

2.5 Whole blood processing

Collected whole blood samples were processed as previously described (Page et al. 2013). In short, 20ml of venous blood was collected in K2 EDTA vacutainer tubes, kept on ice, and processed within 120 mins. Blood was centrifuged (1000g 10 mins 4^{o}C) leading to the creation of 3 distinct layers - plasma, buffy coat (BC) and erythrocytes. Plasma was transferred to a new 15ml falcon tube and centrifuged for a second time (2000g 10 mins 4^{o}C) before being split into 1000 microlitre (µl) aliquots and stored at -80^{o}C. Simultaneously 400µl of BC and 1000µl of erythrocytes were collected and also stored at -80^{o}C.

2.6 Radiological patient assessment

To objectively gauge patients' disease burden and state a consultant musculoskeletal radiologist with a specialist interest in soft tissue tumours assessed their radiological investigations using the Response Evaluation Criteria in Solid Tumours (RECIST) 1.1 criteria (Eisenhauer et al. 2009). These criteria were published in 2009 to help standardise and simplify recording disease burden and state/response in cancer patients. In brief, a patient being assessed using RECIST 1.1 initially has their malignant lesions defined as 'measurable' or 'not measurable' based on the ease with which they can be sized accurately. Next a patient's 'measurable' lesions are sub-categorised as either 'target' (maximum 5/patient) or 'non-target' lesions based on their suitability for repeated measurement. To determine disease burden the sum of up to 5 of a patient's target lesions' diameters is calculated. To determine their disease state serial values for disease burden are compared. RECIST 1.1 definitions for disease state are complete response (CR), partial response (PR), progressive disease (PD) and Stable Disease (SD) (Table 2.2).

Response criteria	RECIST 1.1
Complete response (CR)	Disappearance of all lesions
Partial response (PR)	≥ 30% decrease in sum of the longest diameters of all target lesions
Stable disease (SD)	Neither PR or CR
Progressive disease (PD)	>20% increase in the sum of the longest diameters

Table 2.2: A summary of disease state/response criteria defined according to the Response Evaluation Criteria in Solid Tumours (RECIST) 1.1 criteria.

2.7 DNA extraction techniques

2.7.1 Cell free DNA: Cell-free DNA was extracted from plasma using the QIAamp Circulating Nucleic Acid kit (Qiagen, Hilden, Germany) as per manufacturer's instructions. Three 1ml plasma aliquots were combined for each extraction. First, to remove any DNAases or RNases present in plasma samples and to release any circulating nucleic acids present from bound proteins, lipids and vesicles each sample was mixed with 300µl Proteinase K and 2.4ml Buffer ACL (provided in the QIAamp Circulating Nucleic Acid Kit) and incubated at 60°C for 30mins in a water bath. Next to optimise the binding of circulating nucleic acids to a silica membrane found in a QIAamp Mini column 5.4mls of Buffer ACB (provided in the QIAamp Circulating Nucleic Acid Kit) was added to each sample and the resulting mixture vortexed for 30secs and incubated for 5mins on ice. Following this lysates were transferred onto QIAamp Mini columns and placed onto a vacuum manifold which was used to pull lysates through their columns using vacuum pressure. Next any residual contaminants still bind to the QIAamp Mini columns' silica membranes with the target circulating nucleic acids were washed away by drawing 600µl Buffer ACW1 (provided in kit), 750µl Buffer ACW2 (provided in kit) and finally 750µl of 100% ethanol through the QIAamp Mini column using the same vacuum manifold. Washed QIAamp Mini columns were then placed into clean 2ml collection tubes and centrifuged at 14,000rpm/20,000 x g for 3mins before being placed into new 2 ml collection tubes and being incubated with open lids at 56°C for 10 min to dry the silica membranes completely. Finally each QIAamp Mini column was placed into a new DNA low-bind tube and 70µl of nuclease/protease free water added. Columns' lids were then closed and each assembly incubated at room temperature for 3mins before being centrifuged at 14,000rpm/ 20,000 x g for 1 min to elute the nucleic acids.

2.7.2 Fresh frozen tumour tissue genomic DNA: Tumour tissue undergoing DNA extraction was sectioned to a width of 5µm in a cryostat using a microtome and suspended in 250µl of 1xSSC solution (see figures 2.3-2.7). 250µl of lysis solution (100mM Tris-HCl pH 7.5, 100mM NaCl, 10mM EDTA, 1% sarkosyl) and 5µl of RNase (stock concentration 10mg/ml) was added and the solution incubated on ice for 20mins. 2.5µl of Proteinase K (stock concentration 20mg/ml) was next added and the sample incubated at 55^{o}C for 6 hours inverting every hour. The solution was next added

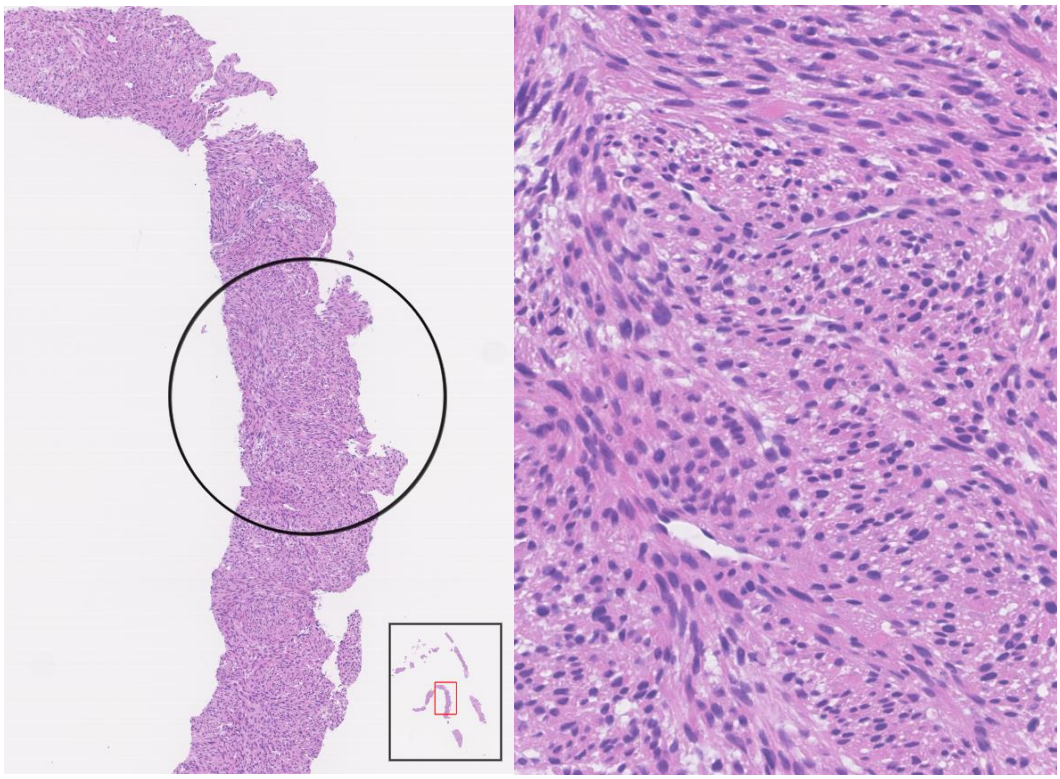


Figure 2.3a: H&E stained FFPE Leiomyosarcoma tissue from participant 1. The region highlighted by the black circle represents the area from which 1mm cores of tumour tissue were harvested for DNA extraction.

Figure 2.3b: High power magnification of region circled in figure 2.3a from which DNA was extracted. Magnification and imaging was performed using a Hamamatsu slide scanner.

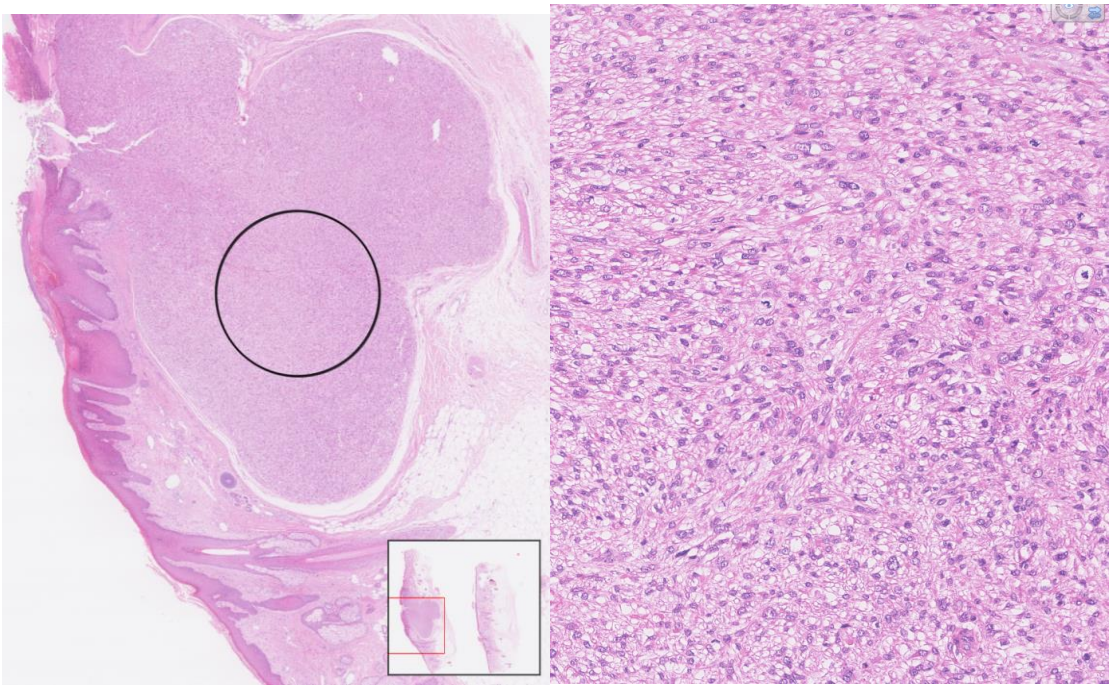


Figure 2.4a: H&E stained FFPE Leiomyosarcoma tissue from participant 2. The region highlighted by the black circle represents the area from which 1mm cores of tumour tissue were harvested for DNA extraction.

Figure 2.4b: High power magnification of region circled in figure 2.4a from which DNA was extracted. Magnification and imaging was performed using a Hamamatsu slide scanner.

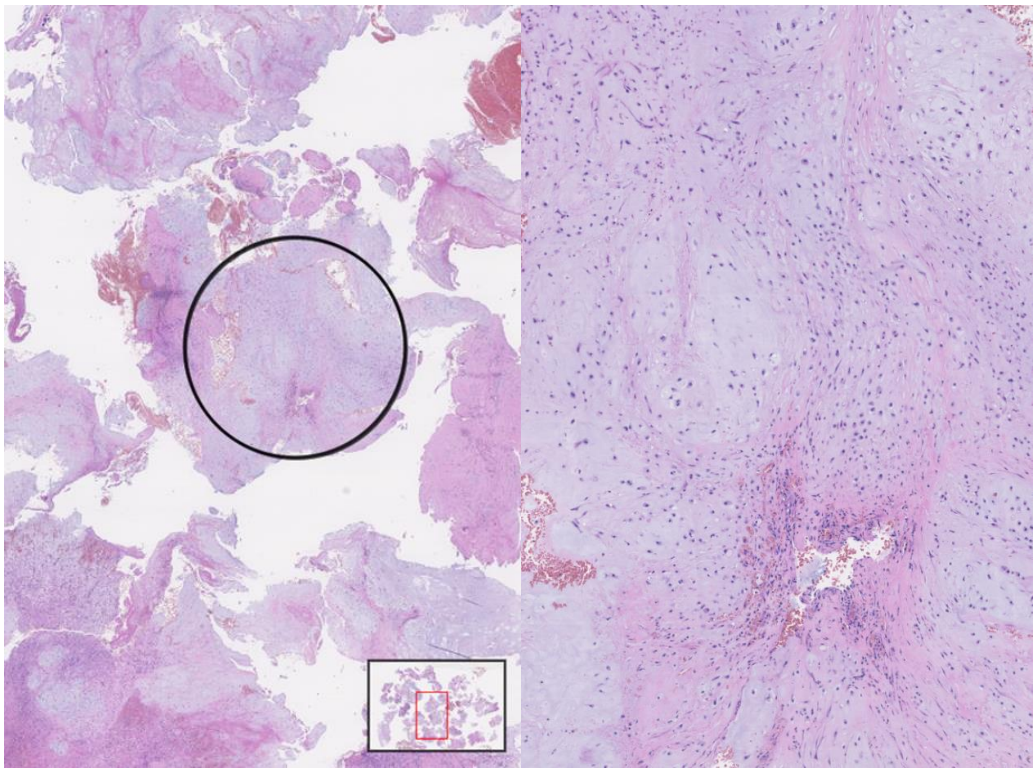


Figure 2.5a: H&E stained FFPE soft tissue Chondrosarcoma tissue from participant 4. The region highlighted by the black circle represents the area from which 1mm cores of tumour tissue were harvested for DNA extraction.

Figure 2.5b: High power magnification of region circled in figure 2.5a from which DNA was extracted. Magnification and imaging was performed using a Hamamatsu slide scanner.

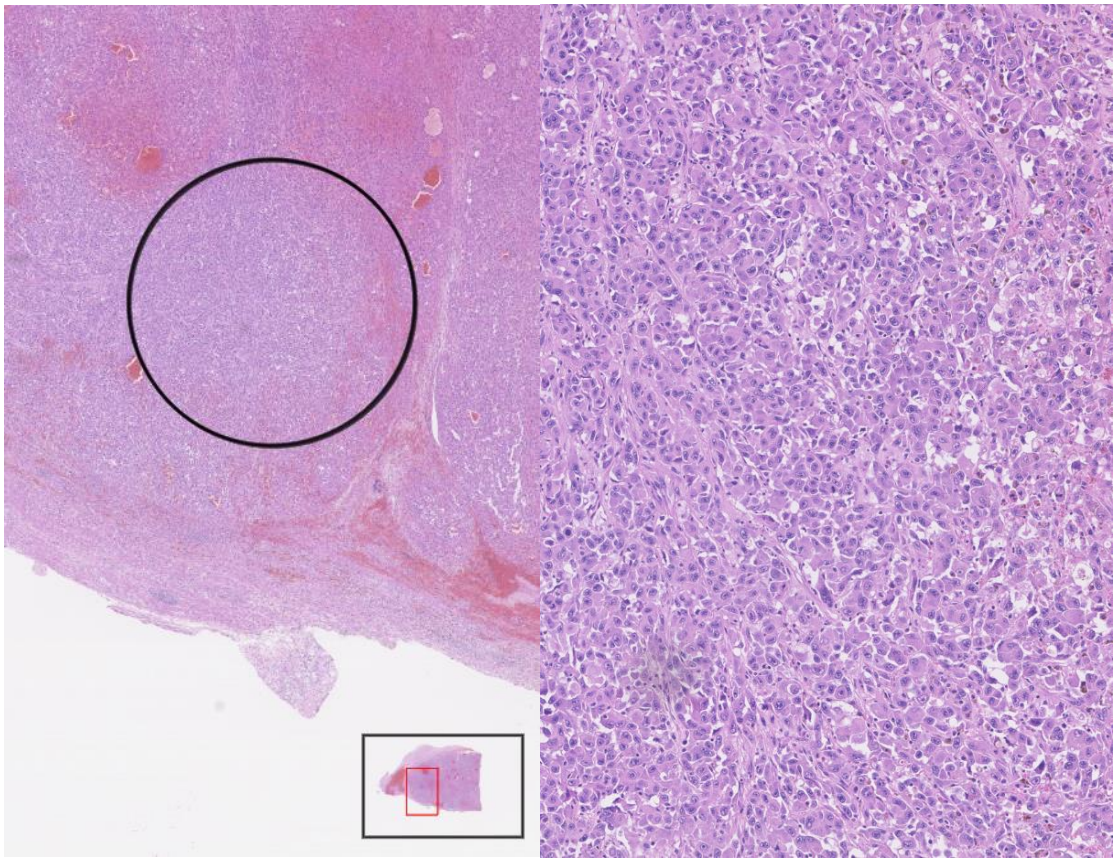


Figure 2.6a: H&E stained FFPE Epithelioid Angiosarcoma tissue from participant 5. The region highlighted by the black circle represents the area from which 1mm cores of tumour tissue were harvested for DNA extraction.

Figure 2.6b: High power magnification of region circled in figure 2.6a from which DNA was extracted. Magnification and imaging was performed using a Hamamatsu slide scanner.

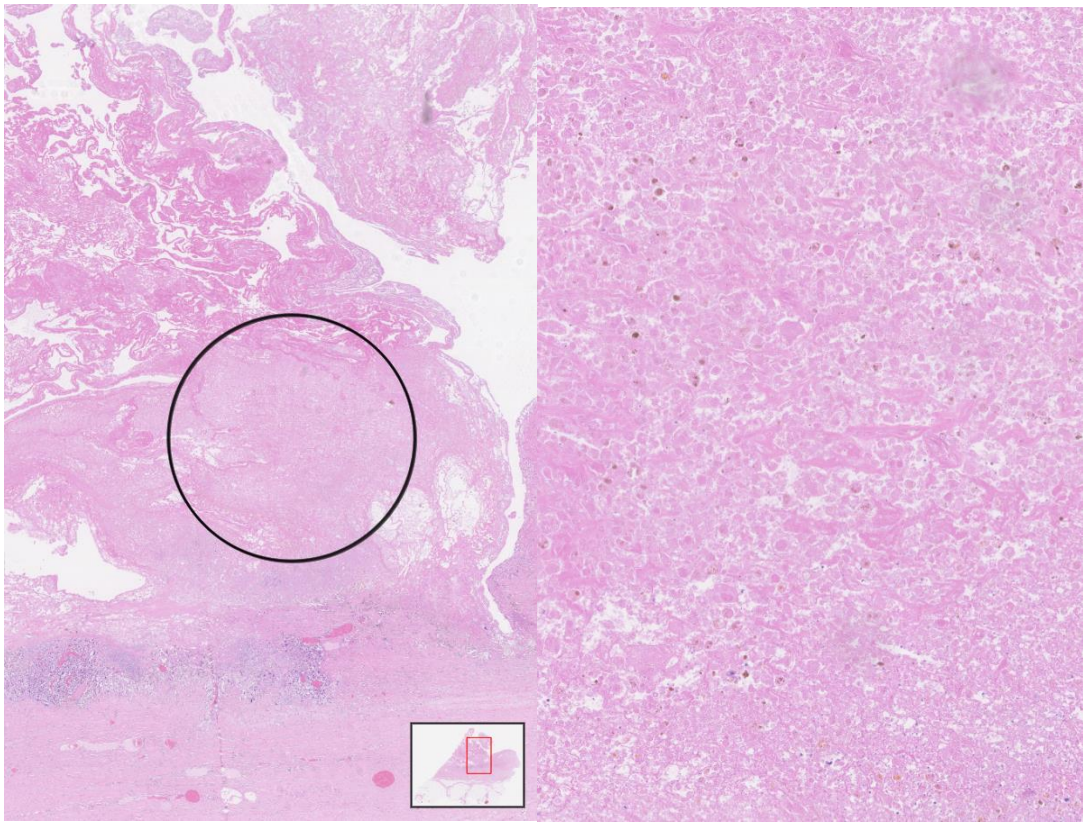


Figure 2.7a: H&E stained FFPE Undifferentiated Pleomorphic Sarcoma tissue from participant 6. The region highlighted by the black circle represents the area from which 1mm cores of tumour tissue were harvested for DNA extraction.

Figure 2.7b: High power magnification of region circled in figure 2.7a from which DNA was extracted. Magnification and imaging was performed using a Hamamatsu slide scanner.

to MaXtract High Density tubes (Qiagen, Hilden, Germany) and 500µl of Phenol:Chloroform:Isoamyl alcohol saturated with 10mM Tris, pH8.0 1mM EDTA (25:24:1) (Sigma-Aldrich, Missouri, United States) added. After mixing thoroughly the solution was centrifuged (13,000rpm/15,700 x g 7mins room temp) and the resulting aqueous component poured into a new 2ml DNA low bind Eppendorf. Next 50µl of 2M NaOAc pH5.6 and 1.375ml 100% ethanol was added and the solution left at room temp for 10 mins to allow precipitation. The resulting pellet was transferred to a new 1.5ml DNA low bind Eppendorf and washed twice with 800µl of 80% ethanol. Finally, following the removal of excess ethanol the pellet was air-dried at room temperature for 30 mins before being dissolved in nuclease/protease free water.

2.7.3 Formalin fixed paraffin embedded tissue DNA: FFPE tumour genomic DNA was extracted using the FFPE Gene Read Kit (Qiagen Hilden, Germany) according to the manufacturer's instructions. Single 1mm cores of FFPE tumour samples were first placed in 1.5ml Safelock tubes, mixed with 155µl of deparaffinization solution then incubated at 56°C for 3 min. After allowing samples to cool to room temperature 55µl of nuclease/protease free water, 25µl Buffer FTB (provided in kit) and 20µl proteinase K were added and each sample then incubated at 56°C for 1 hour, then 90°C overnight. The next day the lower, clear phase of each mixture was transferred to a new 1.5ml Safelock tube and heated to 90^{o}C for 1 hour. After this samples was mixed with 115µl of nuclease/protease free water and 35µl Uracil-N-Glycosilase (UNG) (provided in kit) and incubated at 50°C for 1 hour. Next 2µl of RNase A (provided in kit, working concentration 100 mg/ml) was added to each sample which were then incubated for 2 min at room temperature before 250µl of Buffer AL (provided in kit) and 250µl 100% ethanol were added. After this each sample was transferred to a QIAamp MinElute column suspended in a 2ml collection tube and centrifuged at 14,000rpm/ 20000 x g for 1 min. To wash away any residual contaminants 3 centrifugation steps were next performed. First 500µl of buffer AW1 (provided in kit) was added to each column before they were centrifuged at 14,000rpm/ 20000 x g for 1min. After the resulting follow through was discarded a further 500µl of buffer AW2 (provided in kit) was also added to each column which were again centrifuged at 14,000rpm/ 20000 x g for 1min. Finally, again after discarding follow through, 250µl of 100% ethanol was added to each spin column before centrifugation at 14,000rpm/ 20000 x g for 1min. Spin columns were next

transferred to new 2ml Eppendorfs and centrifuged again at 14,000rpm/ 20000 x g for 1 min to remove any residual liquid. After this columns were finally moved to new 1.5ml Eppendorfs, 30µl of buffer ATE (provided in kit) added and after a 5 minute incubation period at room temperature centrifuged at 14,000 rpm/ 20000 x g for 1min to elute DNA.

2.7.4 Buffy coat DNA: DNA was extracted from buffy coat using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) as per manufacturer's instructions. This extraction contains 4 main stages – sample lysis, adsorption of the BC DNA to QIAamp silica membranes found in QIAamp Mini columns, removal of contaminants using buffers provided and finally elution of the target DNA. Initially 20µl of proteinase K was added to 200µl of BC, followed by a further 200µl of Buffer AL (provided in kit). The resulting mixture was left to incubate (lyse) at 56°C for 10 min before 200µl of 100% ethanol was added. Next each sample was transferred directly onto the silica membrane present in each QIAamp Mini spin column. These columns were then placed onto a vacuum manifold and the lysates drawn through their respective membranes using a vacuum. Following this to wash away contaminants 750µl of Buffer AW1 and 750µl of Buffer AW2 (both provided in kit) were drawn through each QIAamp Mini column after one another (again using the vacuum manifold.) Next columns were transferred to clean 2ml collection tubes and centrifuged at 14,000 rpm/ 20,000 x g for 1 min to completely dry each silica membrane. After this 200µl of nuclease/protease free water was added to each column (directly on each silica membrane) which were then left at room temperature for 1 minute. Finally columns were then centrifuged at 8000rpm / 6000 x g to elute the target DNA.

2.8 DNA quantification / Quality control

Circulating free DNA yields were quantified using quantitative PCR (Section 2.9). Following this cfDNA samples were also analysed with an Agilent 4200Tapestation System (Agilent, California, United States). This provided data on the length of the DNA fragments present, allowing sample contamination with genomic DNA to be ruled out. Tumour DNA and BC DNA yields were determined using the Qubit® dsDNA HS (High Sensitivity) Assay Kit and a Qubit® 2.0 Fluorometer (Thermo fisher Scientific, Massachusetts, United States) according to manufacturer's instructions. For this process

a sufficient volume of Qubit working solution was first made up by diluting Qubit® dsDNA HS Reagent with Qubit® dsDNA HS Buffer at a ratio of 1:200 (both provided in kit). Next the 2 standards required for DNA quantification using the assay were prepared by mixing 190µL of this working solution with 10µl of 2 prediluted DNA standards (also provided in kit). Following this 199µL of the same working solution was added to 1µl of each DNA sample being quantified. Next all diluted standards and samples were thoroughly mixed and incubated at room temperature for 2 minutes. Following this the 2 standards were quantified using the Qubit® Fluorometer and a two point standard plot created. Using this plot the concentrations of each DNA sample being analysed was calculated using the fluorometer, and from these values the concentrations of the DNA stock solutions were extrapolated. .

2.9 Quantitative Polymerase Chain Reaction

Quantitative Polymerase Chain Reaction (qPCR) was used to quantify cfDNA levels using a GAPDH Taqman assay. To create a standard curve for this six 35µl dilutions of human genomic DNA (Roche Holding AG, Basal, Switzerland) were created ranging from 5 nanogram (ng) – 0.156ng/35µl. Next a GAPDH primer/probe mix (GAPDH FRP mix) was made up consisting of 3µl of forward and reverse GAPDH primers (Sigma-Aldrich, Missouri, United States), 200pmol/µl stock, F: 5'-GGCTAGCTGGCCCGATTT-3', R' – GGACACAAGAGGACCTCCATAAA-3', amplicon size 95bp), 2µl of Taqman VIC-MQB GAPDH probe (Applied Biosystems, California, United States / 100nM stock concentration, 5'-ATGCTTTTCCTAGATTATTC-3') and 132µl of water. Following this qPCR reaction mixes were made up in triplicates for every DNA sample to be quantified, each of the 6 human genomic DNA dilutions and one non-template control (NTC). These 10µl reaction mixes consisted of 5µl Taqman Universal Fast PCR Mastermix (Applied Biosystems, California, United States), 1.4µl of the GAPDH FRP mix (equating to 0.6µl of 10pmol/µl forward and reverse primer and 0.2µl of probe at a 1:10 dilution) and 3.6µl of the DNA to be quantified per well. For NTC reactions 3.6µl of water was substituted for this DNA. Following centrifugation for 30 seconds at 1000xg the reaction were run on a Step OnePlus Real-Time PCR system (Applied Biosystems, California, United States). Cycling conditions were 95^{o}C for 20 secs, followed by 40 cycles of 95^{o}C for 1 second and 60^{o}C for 20 seconds. The repeatability of the qPCR assay outlined above was

confirmed by quantifying total cfDNA levels in three separate plasma samples on three consecutive days (Figure 2.3). Significant variation between the cfDNA yields from different plasma aliquots collected from the same whole blood samples was also ruled out (Figure 2.4).

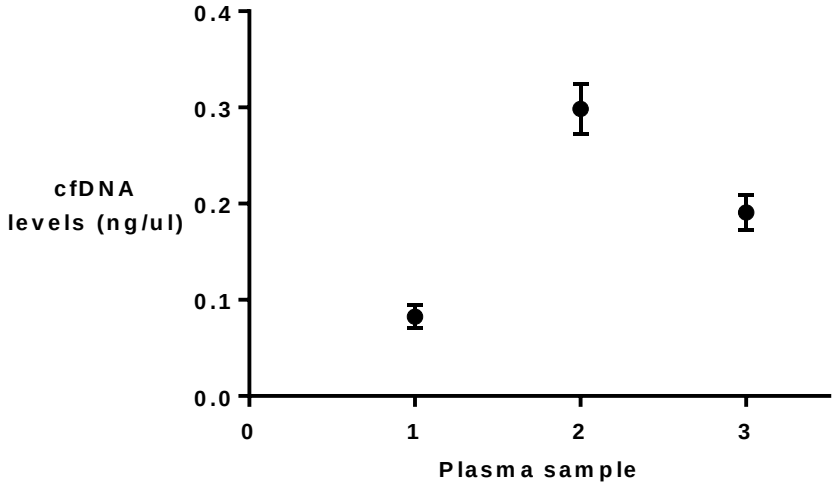


Figure 2.8: The reproducibility of the qPCR assay used to quantify plasma cfDNA levels. To check the reproducibility of the PCR assay used to calculated plasma cfDNA levels we quantified the cfDNA levels of 3 different plasma samples (numbered 1, 2 and 3) on 3 consecutive days and compared the results. The mean and standard error of the mean (SEM) of these replicates are represented by the error bars in the figure above.

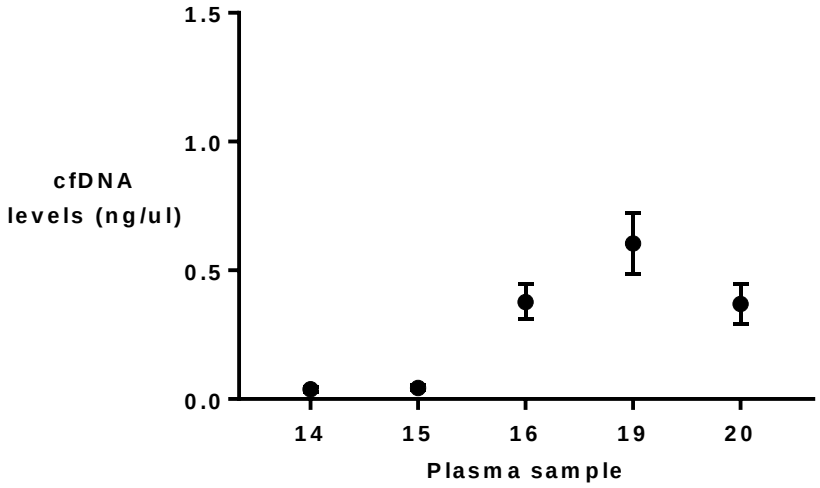


Figure 2.9: An assessment of variation in cfDNA yield of cfDNA extractions. To rule out variation in the levels of cfDNA obtained from cfDNA extractions performed on different plasma samples collected from the same whole blood sample we compared the cfDNA levels attained from every non-metastatic STS patient's blood sample which provided enough plasma to allow 3 separate cfDNA extractions. The graph shows the mean and SEM (represented by error bars) of these cfDNA levels.

2.10 Polymerase chain reaction (PCR)

The standard 10µl PCR mixes were made up of 0.3µl forward and reverse primer (stock concentration 10µM), 0.06µl of Taq Polymerase (Kappa Biosystems, Roche Holding AG, Basel, Switzerland, 0.04u/µl), 0.9µl of 11.1x buffer (45mM Tris HCL pH 8.8, 11mM Ammonium Sulphate, 4.5mM $MgCl_2$, 6.7mM 2-Mercaptoethanol, 4.4um EDTA pH 8.0, 1mM of each dNTP, 113 microgram (µg)/ml BSA), 7.44µl of nuclease free water and 1µl of DNA template (5-10ng/µl). Unless otherwise specified PCRs were performed at 96ºC for 90 seconds followed by 32 cycles at 96ºC for 20 seconds, 30 seconds at the selected annealing temperature followed by a final extension phase at 68ºC for 30seconds/kb product. Annealing temperatures were based on primers' GC content and *in silico* PCR predictions. All PCR reactions were performed in a Veriti thermocycler (Applied Biosystems, California, United States).

2.11 Agarose Gel Electrophoresis

Gel electrophoresis was performed in 2% agarose gels (NuSieve™ 3:1 Agarose Lonza or SeaKem® LE Agarose) made up with 0.5xTBE (44.5mM TRIS Base, 44.5mM Boric acid, 2mM EDTA pH 8.0) and mixed with ethidium bromide at a final concentration of 0.5µg/ml. Before being loaded DNA samples were mixed with 5xTBE loading dye. 50 bp and Lambda Hind III DNA size markers were used and also diluted with TBE loading dye prior to loading (stock concentration 50ng/µl.) Electrophoresis was performed in 0.5xTBE also containing ethidium bromide (0.5µg/ml) at a constant voltage determined by the predicted size of the DNA amplicons (range 80-140V). Gels were imaged using a Syngene Gene Genius Bio Imaging System and the GeneSnap software package.

2.12 Sanger sequencing

DNA undergoing Sanger sequencing was initially amplified using standard PCR as outlined in 2.10 but with a final reaction volume of 20µl. Following separation using electrophoresis, gels were visualised using bluelight and the band(s) of interest removed using a scalpel blade. DNA was extracted and purified from these bands using a Gel DNA Recovery Kit (Zymo Research Cat#D4002) according to the manufacturer's instructions and quantified using a Nanodrop spectrophotometer (Thermofisher, Massachusetts, United States). Next a 20µl sequencing reaction mixture was set up

consisting of the extracted template DNA (20ng/kb), 1µl of 5x sequencing buffer (Applied Biosystems, California, United States), 3.5µl of Big Dye V3.1 reaction mix (Applied Biosystems, California, United States), 1µl of reverse primer (stock concentration 3.2µM) and water. This reaction was heated for 28 cycles at 96^{o}C for 10 seconds, 55^{o}C for 5 seconds and 60^{o}C for 4 minutes in a Veriti thermocycler (Applied Biosystems, California, United States). Following this 2µl of 2.2% sodium dodecyl sulfate (SDS) was added to the mix before it was heated to 98^{o}C for 5mins then 25^{o}C for 10mins. During this process a gel filtration cartridge (EdgeBio Mayland, United States) was prepared for use by centrifugation at 3,400rpm / 850 x g for 3mins. Finally the cleaned reaction mix was pipetted onto it and the centrifugation step repeated. All samples were sent to the Protein and Nucleic Acid Chemistry Laboratory (PNACL), University of Leicester, for electrophoresis and analysis using a DNA analyser (3730 Applied Biosystems, California, United States).

2.13 Whole Exome Sequencing:

All STS samples composed of >60% viable tumour tissue were sent for whole exome sequencing with matched normal BC DNA. DNA library preparation and sequencing was performed by technical staff working for SourceBioscience or Novogene at their respective laboratories in Nottingham (https://www.sourcebioscience.com) and China (https://en.novogene.com). DNA libraries sequenced by both companies were prepared similarly using the Sureselect Human All Exon V6 library preparation kit (Agilent, California, United States). Following quality checks DNA samples were quantified using a Qubit® 2.0 fluorometer (Thermofisher, Massachusetts, United States) and libraries prepared according to manufacturer's recommendations: Initially DNA was fragmented using a hydrodynamic shearing system (Covaris, Massachusetts, United States) to generate 180-280bp fragments. Next these fragments' were blunt ended using exonuclease/polymerases, 3' adenylated and ligated to adapter oligonucleotides. Following this DNA fragments were PCR enriched and purified using biotin labelled probes, streptomycin magnetic beads and the AMPure XP system (Beckman Coulter, California, United States). Following a further quantification and sizing step using a Bioanalyzer 2100 system (Agilent) the exonic regions of interest were captured using the SureSelect Capture kit. Next captured libraries were isolated, washed and further enriched using PCR during which they are labelled with library specific index sequences

to allow them to be pooled equally in a multiplex sample. Finally this multiplex sample was loaded onto an Illumina NextSeq flow cell (v2) and sequenced using a 150bp PE kit.

2.14 Bioinformatic analysis:

Bam file production and SNV/indel calling was performed by Dr Jin-Li Luo (Bioinformatician, Department of Genetics and Genome biology, University of Leicester) as outlined below.

2.14.1 Bam file production:

Raw FASTQ files produced by sequencing were processed using Skewer v0.2.2 - a quality control tool specially designed for processing Illumina pair end sequences. During this QC stage any adaptor sequences were removed and reads trimmed based on their phred quality read scores and lengths. Parameters set during this process were a maximum allowed error rate of 0.1 (default), a mean quality threshold of 20, 3' end quality threshold 25 and a minimum read length allowed after trimming of 18. The resulting leftover reads were next aligned to the hg19 reference genome using the Burrows-Wheeler Aligner software package (0.7.12) to create .sam files. The programme Samtools (1.3.2) was then used to convert these .sam files to binary .bam files and subsequently sort and index them. De-duplication of mapped reads was performed using picardtools (2.6.0).

2.14.2 Single Nucleotide Variants (SNVs) and insertion/deletion (Indel) calling:

For SNV calling the bioinformatics tool somatic-sniper (1.0.5.0) was used to compare STS samples with BC DNA from the same patient. To identify small indels the tool Strelka (1.0.14) was used. Following variant calling the tool Annotation of Genetic Variants (ANNOVAR) was used for variant annotation.

2.14.3 Copy Number Alterations (CNAs):

Somatic tumour CNAs were characterised using the allele specific copy number package Sequenza (2.1.2). To validate this package a subset of matched tumour and normal sample were also analysed using the nanotechnology (and specifically the nCounter® v2

Cancer CN Assay) or droplet digital PCR (ddPCR). Sequenza output files were then subsequently processed using Genomic Identification of Significant Targets in Cancer (v2.0.23) (GISTIC2) to identify regions of the genome found to be significantly amplified or deleted within the cohort. It does this by assigning every amplification/deletion identified within the analysed cohort with a 'G score' (which reflects both their amplitude and frequency across the analysed samples) followed by a False Discovery Rate 'q-value' (which if lower than 0.1 indicates significance.)

2.14.3.1 Sequenza: Sequenza is software package that that consists of two distinct parts: a preprocessing python tool and an R package. The python based component of sequenza has several roles. The first is to process the sequencing data from matched normal and tumour samples to identify their base sequences and sequencing depth. Next the tool identifies the heterozygous positions in the normal specimen, any variants in the tumour sample, and the B allele frequencies (BAF) at each heterozygous position within the tumour sample (defined as the lesser of the two allelic frequencies at a heterozygous position). Finally the tool calculates the GC content of the sequenced regions from a reference genome. The next stage of the sequenza package takes the output of this analysis forward into a 'Sequenza' R package. This package consist of three components. The first (*sequenza.extract)* performs a GC-content normalization step on the data to counter for sequencing bias. The second component (*sequenza.fit*) next splits the sample's genome into regions of equal copy number (Supplementary 2.1-2.31) before factoring this data and a sample's normalised depth ratio and BAFs at SNPs into a probabilistic model that infers the analysed tumour sample's cellularity, ploidy and copy number profile. Finally, the third component of the R package (*sequenza.results*) provides graphical representations of this data for visualisation purposes.

2.14.4: Data processing and visualisation

The SNV output .VCF files (variant call format) that resulted from our somatic sniper analyses were converted into .MAF (mutation annotation format) files (Supplementary 2.32-2.62) and processed / visualised using the software package Maftools v1.4.21 (Mayakonda & Koeffler 2016). Sequenza output files were processed using GISTIC2 to identify significantly amplified / deleted genomic regions across the cohort. The files

created by this analysis were then also processed and visualised using the software package Maftools (v1.4.21).

2.15 Nanostring nCounter® v2 Cancer CN assay

The nCounter analysis system (Nanostring Technologies, Washington, United States) uses gene specific probes and digital detection technology to identify the absolute number of copies of genes in a tumoural sample. Using this approach the nCounter® v2 Cancer CN assay allows somatic CNAs to be identified in 87 cancer associated genes simultaneously. To validate the CNA analysis performed using Sequenza, samples of matched tumour and normal DNA collected from a batch of enrolled patients were simultaneously analysed using the nCounter® v2 Cancer CN assay. This analysis was performed by staff working at the University College London's Nanostring Facility. For each patient 300ng of matched tumour and normal BC DNA was analysed. First DNA was digested with Alu1 at 37ºC for 2 hours. Following a denaturisation step DNA was next processed using Nanostring Technologies' nCounter Codeset chemistry. The template was mixed with multiple pairs of probes collectively referred to as Codesets. Each codeset consists of a colour-coded fluorescent reporter probe and a capture probe, both covalently attached to oligonucleotides specific to 3 different regions within each gene of interest. Following an overnight hybridisation stage, codeset/DNA complexes were immobilised by their capture probes onto an nCounter Cartridge, before being transferred to an nCounter Digital Analyser where the numbers of each gene specific reporter probes was counted. The v2 Cancer CN assay contains multiple DNA controls added to the template prior to the process of fragmentation. These include two DNA fragments with and without Alu1 restriction sites controlling for restriction and DNA input/denaturation respectively, eight negative DNA controls to measure background probe counts, six positive DNA controls diluted in a linear fashion to measure assay efficiency, and 10 DNA controls complementary to different autosomal diploid sites in the genome to facilitate further compensation for variation in quantification or pipetting. Analysis of the raw data was performed in house as recommended by Nanostring Technologies (https://www.nanostring.com.) Following QC and data normalization the relative tumoural frequency of each gene's 2 reporter probes was determined by dividing their absolute count in each tumour by their count in the matched BC samples also analysed. Finally to determine each gene's tumoural copy number the relative frequency

of each their 3 probes was averaged and multiplied by two. In an attempt to identify only high level amplifications and homozygous deletions stringent thresholds were adopted for amplifications (copy number >4) and deletions (copy number <1). In addition for Whole exome sequencing (WES) validation purposes only those genes in which all 3 probes were amplified were considered to be truly amplified.

2.16 Targeted semiconductor Next Generation Sequencing (NGS)

Target next generation sequencing was performed using an Ion Torrent Personal Genome Machine (PGM) sequencer (Life Technologies, California, United States). This process involved 3 main steps: Ion Torrent library preparation, Ion Torrent Template preparation and Ion Torrent template sequencing.

2.16.1 IonTorrent SNV panel design:

Two custom designed panels were used for the analysis outlined in this thesis ('Sarcoma Version 1' and 'Sarcoma Version 2'). For the analysis of metastatic STS patient samples described in chapter 4 'Sarcoma V1' was used. This panel was designed following a search of the COSMIC (Catalogue of Somatic Mutations in Cancer) database (http://cancer.sanger.ac.uk/cosmic) and interface cBioportal (www.cbioportal.org) looking for the most commonly reported substitutions identified in STSs. In addition amplicons were included to span exon 9 of *ATRX* - a region with a propensity to be mutated in ALT positive tumours. In total the panel contained 58 amplicons located in 13 different genes. These average 85bp in length (range 65-100) and covered a total of 3521 COSMIC registered variants located at 1004 different hotspot locations (Supplementary 2.63). To analyse samples collected from non-metastatic patients as outlined in chapter 5 'Sarcoma V2' was used. This panel was designed to cover the most common variants found in the 3 most commonly mutated genes in STSs based on cBioportal data - *TP53*, *ATRX* and *RB1*. This panel had a total of 45 amplicons averaging 84bp in length (range 68-96) which together spanned 699 different hotspot locations associated with 1858 difference cosmic registered variants (Supplementary 2.64).

2.16.2 IonTorrent library preparation:

DNA libraries were prepared using the Ion AmpliSeq™ Library Kit v2.0 (ThermoFisher, California, United States) according to manufacturer's instructions. Ten nanograms of template DNA was amplified in a multiplex PCR using the Sarcoma V1 or 2 primer pool. Following this amplicons were blunt ended and attached to barcoded adapters by ligation. Next DNA fragments were purified using Agencourt AMPure XP beads (Beckman Coutler, California, United States) and reamplified with 5 cycles of PCR. Following a further bead purification step individual libraries were sized and quantified using a Bioanalyzer (Agilent Technologies, California, United States) and high sensitivity DNA analysis kits before being pooled, re-quantified, and diluted to 100pM in preparation for sequencing.

2.16.3 IonTorrent template preparation:

Pooled libraries were prepared for sequencing using the Ion PGM™ Hi‑Q™ OT2 Kit according to the manufacturer's instructions. Two modules were utilised during this preparation - the Ion OneTouch 2 Instrument and Ion OneTouch Enrichment System (ES). Initially the Ion OneTouch 2 was used to clonally amplify the DNA template in an emulsion PCR containing Ion PGM™ Hi‑Q™ reagent and enzyme mixes, Ion Sphere Particles™ (ISPs) and 2µl of 100pm pooled library. Following this template positive ISPs were enriched in a 2 stage process. Empty ISPs were first removed from the emulsion PCR product with the use of Dynabeads™ MyOne™ Streptavidin C1 beads. Next in an automated process on the Ion OneTouch ES positive ISPs were washed several times, before complementary template strands were 'melted' away in preparation for sequencing with 1M NaOH solution.

2.16.4 IonTorrent next generation semiconductor sequencing:

Template was sequenced on the Ion PGM™ System using the Ion PGM™ Hi‑Q™ View Sequencing Kit as per the manufacturer's instructions: Following appropriate chlorite cleaning and initialisation of the PGM™ System the enriched ISPs were mixed with control ISPs, sequencing primer and sequencing polymerase. Next template was loaded onto an Ion 314/6™ chip and a pre-planned run performed. Where cfDNA yields allowed metastatic patient plasma was sequenced in replicates.

2.16.5 IonTorrent next generation semiconductor sequencing variant calling:

During plasma analysis the presence of circulating somatic variants registered on the COSMIC database at a frequency of >0.5% was assumed to provide evidence of ctDNA. In every patient matched BC DNA was sequenced to 1) gain a measure of background sequencing noise and 2) to confirm the somatic nature of any circulating variants identified. Wherever possible matched tumour DNA was also sequenced. Sequencing data was aligned against the hg19 genome assembly. Somatic variants were identified in two ways: First sequencing data was analysed using the Variant Caller software. Next the Bam files of each amplified amplicon were manually reviewed using the Integrated Genomics Viewer (IGV) package (v2.3.25) to identify any variants missed by Variant Caller software. Variants identified by the Variant Caller software were also reviewed using IGV to rule out any sequencing artefact/errors.

2.17 Droplet digital PCR

2.17.1 Single nucleotide variant analysis

In a subset of non-metastatic patients somatic single nucleotide variants (SNVs) identified by tumour WES were sought in matched plasma samples using custom designed Taqman hydrolysis probe droplet digital PCR (ddPCR) assays.

2.17.1.1 SNV selection:

The SNVs chosen for profiling in patient plasma samples were selected on the basis of their predicted effect (non-synonymous), location (exonic) and the quality of the raw sequencing data they were identified from (somatic sniper somatic score >20, tumour read depth >50) (Table 2.3). Prior to ddPCR analysis the presence of each target SNV in tumour tissue was confirmed using Sanger sequencing.

2.17.1.2 SNV ddPCR assay development:

Commercially available, wet lab validated rare event detection ddPCR assays were used wherever possible to identify circulating SNVs (Bio-Rad Laboratories, California,

United States). In the absence of such assays, Taqman hydrolysis probe ddPCR assays were developed and optimised in house (Table 2.4) using a combination of online tools.

Patient number	Gene	SNV position (Chr:Loci)	Coding strand (+/-)	Base change	Predicted effect	Mutation Frequency	SIFT prediction	Detected in matched plasma
003	*VWDE*	7:12384078	-	T>C	Cys1302Arg	42%	D	N
006	*TP53*	17:7577022	-	C>T	Arg306Ter	56%	.	Y
	BRIP1	17:59761496	-	C>G	Pro971Ala	20%	T	N
009	*PTCH1*	9: 98239884	-	C>A	Ala332Glu	23%	D	N
	LPP	3:188327063	+	C>A	Pro182Thr	46%	D	N
018	*FLT4*	5: 180046092	-	G>A	Val927Met	18%	D	Y
022	*DACH1*	13: 72053389	-	A>C	Glu594Asp	21%	.	N
	EPHB6	7: 142563798	+	G>A	Gly397Arg	44%	.	Y
024	*MMS22L*	6: 97634424	-	C>T	Gln728Ter	25%	.	N
025	*ITIH2*	10: 7769692	+	C>T	Arg394Trp	37%	D	N
	KDM5B	1: 202777369	-	C>T	Pro22Leu	88%	D	N
026	*PTPRB*	12:70970320	-	C>T	Thr677Ile	73%	T	N

Table 2.3: Single nucleotide variants targeted in non-metastatic STS patients' plasma using droplet digital PCR. Mutation frequencies are based on Illumina WES data. **D –** Damaging, **T -** Tolerated.

Target mutation (with coding strand)	Target strand (+/-)	Amplicon size (bp)	Assay primer sequences (5'-3') (Forward/Reverse)	Probe sequence (5'-3') (Mutant)	Probe sequence (5'-3') (Wildtype)
VWDE (7:12384078, T>C) (-)	(+)	101	CAATATGTATGTTCTGTTTTAGaCAA / GGGGCAACACACTCCCTACT	TTAGCCATTTGTAAATATCC[A][c]GT	TAGCCATTTGTAAATATCCA[t]GTGGAA
*TP53 (*17:7577022, C>T) (-)	(-)	Unknown			
*BRIP1 (*17:59761496, C>G) (-)	(+)	127	TGTTAGCTAGGAGCAGAAAGTTA/ TGTTGAAAGTTGGGCTTGT	TCTTAGATGA[T][g]CAGTATTC	TCTTAGATGA[T][c]CAGTATTC
*PTCH1 (*9:98239884, C>A) (-)	(-)	114	CTCCAAGTCCCAGGGTGC/ GTTGTTGCAGCGTTAAAGGAA	CC[T][t]CAGCCACTGACAGT	CAGTCCT[g]CAGCCACTGA
LPP (3:188327063, C>A) (+)	(-)	98	CCAACCCCCTCTAACAGCAA / TGTTCCGATTGGAGCCACAG	TACATTGAA[A][a]CACAGCCTGCAC	TACATTGAAA[c]CACAGCCTGCAC
*FLT4 (*5:180046092, G>A) (-)	(-)	91	GAAGTTGGAGAGGTTGCC/ GGTTCCGAACGCACG	C[A][t][G][A]TCACCATGAGG	C[A][c][G]ATCACCATGAGG
*DACH1 (*13: 72053389, A>C) (-)	(+)	123	GGCTGTTGAAAGTTGCCAT / ACTGCTTCTCAAGTGTTTCCC	CCAACTGG[A][c]AAAACTG	CCAACTGG[A][a]AAAACTG
*EPHB6 (*7:142563798, G>A) (+)	(+)	73	GCAAGGCTCAGCACTCATG / ATTGAAGAGCAGGTCCCCTC	ACCC[C][t]CAGCTCC	ACCC[C][c]CAGCTCC
MMS22L (6:97634424, C>T) (-)	(+)	126	GTACAGTCTTCATTATCGGCTA / CGCAAGTTGTGAGAAAGG	AAGAG[T][t]AGAGAATGTCAC	AAGAGT[c]AGAGAATGTCAC
*ITIH2 (*10:7769692, C>T) (+)	(-)	143	ATGGGATTTTCTGCGTGTG / GAGTTGGGGTCTAACAGTCC	CACTCCT[A][t][G]GGCAATCTTC	CACTCCT[A][c]GGGCAATCTTC
KDM5B (1:202777369, C>T) (-)	(-)	140	GTGGATGAAAGCGAAGGGG / CCACCACACTGCACCCAG	AACTCGCCCAGC[a]G	AACTCGCCCAGC[g]G
PTPRB (12:70970320, C>T) (-)	(+)	86	CGGTGAACAATTCCGGTCGT / ACTGAACCACCTTGCCGTCA	ATAACTATGAG[G]T[A][A][t][A]T	TATGAGGTA[A][c][A]TTGTCTC

Table 2.4: Droplet digital PCR assay design. Probe bases enclosed in square brackets represent locked nucleic acids (LNAs). Small case characters in the probes represent the sites of the target mutations.

2.17.1.2.1 SNV ddPCR assay primer design: To obtain the genomic sequences surrounding each target SNV the National Centre for Biotechnology Information (NCBI) map viewer tool was used (www.ncbi.nlm.nih.gov/mapview/) with the hg19 assembly adopted as a reference. Next each region was checked for homology with other areas of the genome using the NCBI Blast tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Following this primers for each assay were designed using the Primer3 tool (http://primer3.ut.ee/) with defined parameters including predicted annealing temperature (60-66^{o}C), GC content (40-60%), product size (50-120) and difference in forward and reverse primer annealing temperature (<2^{o}C). Potential primer pairs were next analysed using the University of California, Santa Cruz's (UCSC) genome browser's in silico tool (http://genome.ucsc.edu/cgi-bin/hgPcr) to confirm specificity, following by the Integrated DNA Technologies Oligoanalyzer tool (www.idtdna.com/calc/analyzer) to rule out a high chance of self-dimer or heterodimer formation. Once ordered from Sigma-Aldrich (Missouri, United States) the specificity of ddPCR assay primers was confirmed in two ways – 1) standard PCR and gel electrophoresis and 2) Evagreen PCR performed on the QX200™ droplet digital™ PCR platform (section 2.17.1.3).

2.17.1.2.2 SNV ddPCR assay Taqman hydrolysis probe design: Two Taqman hydrolysis probes (Eurogentec, Liège, Belgium) were designed for each assay - one complementary to the mutant genotype labelled with a FAM reporter and another to the wild type genotype labelled with a HEX reporter. Black hole quenchers were attached to both probes with LNAs as required to optimise annealing temperatures. Probes were designed with a target annealing temperature of 70^{o}C and similarly to primers checked using the Integrated DNA Technologies Oligoanalyzer tool for unacceptable self-dimer or primer heterodimer risk. Prior to the analysis of cfDNA extracted from patients' plasma samples, the optimal cycling conditions and sensitivities of each complete ddPCR assay were checked using patient tumour and BC DNA as template.

2.17.1.3 Evagreen ddPCR: Evagreen ddPCR was performed using the QX200™ Droplet Digital™ PCR System as per the manufacturer's instructions. This system consists of three instruments - the QX200 Droplet Generator, the C1000 Touch™ Thermal Cycler and the QX200 Droplet Reader. Twenty µl reaction mixes were made up

consisting of 2xQX200 ddPCR EvaGreen Supermix, Forward and Reverse primers (final concentration 250nm) and 5ng of human genomic DNA. Using DG8^{TM} gaskets and cartridges this reaction mix was mixed with 70µl of QX200 droplet generation oil for Evagreen and then partitioned by the process of droplet generation by the QX200 Droplet Generator. Following this droplets were immediately loaded onto a 96 well plate which was sealed with a foil lid by heating to 180^{o}C for 5 seconds using a PX1 PCR plate sealer. Next plates were loaded onto a C1000 Touch™ Thermal Cycler where amplification was performed with the temperature ramping rate set to 2^{o}C/second. Cycling conditions were 1) 95^{o}C for 5 mins 2) 40 cycles of 30 seconds at 95^{o}C and 60 secs at 95^{o}C 3) 5 mins at 4^{o}C and 4) 5 mins at 90^{o}C. Following amplification reaction plates were then loaded onto a QX200 Droplet Reader in which droplets are streamed in a single file past an optical detection system allowing them to be counted. Finally raw data was uploaded into the Quantasoft^{TM} 1.7.4 software package (Bio-Rad, CA) for analysis.

2.17.1.4 SNV ddPCR reaction conditions:

Rare event detection ddPCR using Taqman hydrolysis probes was also performed using the QX200™ Droplet Digital™ PCR System as per the manufacturer's instructions. Twenty µl ddPCR reaction mixes were made up with 2x ddPCR™ Supermix for Probes, Forward and Reverse primers (final concentration 900nM each), mutant/wildtype probes (final concentration 250nM each) and 5ng of template DNA (unless otherwise specified). This reaction was mixed with 70µl droplet reader oil and again partitioned using a QX200 Droplet Generator, loaded onto a PCR plate, foil sealed and processed on a C1000 Touch™ Thermal Cycler with the temperature ramping rate 2^{o}C/second. The cycling conditions used for amplification were 1) 95^{o}C for 10 mins 2) 40 cycles of 30 seconds at 94^{o}C and 60 secs at each assay's optimum annealing temperature and 3) 98^{o}C for 10mins. Following amplification droplets were again immediately analysed using a QX200 Droplet Reader where the number of empty, mutant and wildtype positive droplets were counted. Raw data was again analysed using the Quantasoft^{TM} software package where all droplets thresholds were set in 2D prior to final data interpretation.

2.17.2 Copy number analysis:

Droplet digital PCR *MDM2* copy number analysis was also performed using the QX200™ Droplet Digital™ PCR System. For the analysis performed commercially available wet lab validated *MDM2* and RNaseP TaqMan® Copy Number ddPCR assays were run in a duplex reaction as per the manufacturer's instructions. Twenty µl ddPCR reaction mixes were made up with 2x ddPCR™ Supermix for Probes, 20x Taqman *MDM2* copy number assay, 20x Taqman RNaseP copy number reference assay and 10ng of tumour DNA, BC DNA or H_20 for NTC reactions. This reaction was mixed with 70µl droplet reader oil and again partitioned using a QX200 Droplet Generator, loaded onto a PCR plate, foil sealed and processed on a C1000 Touch™ Thermal Cycler with the temperature ramping rate 2^{o}C/second. The cycling conditions used for amplification were 1) 95^{o}C for 10 mins 2) 40 cycles of 30 seconds at 94^{o}C and 60 secs at 60^{o}C and 3) 98^{o}C for 10mins. Following amplification droplets were again analysed using a QX200 Droplet Reader where the number of empty, mutant and wildtype positive droplets were counted. Raw data was then analysed using the Quantasoft^{TM} software package where droplets thresholds were set in 2D prior to final data interpretation and comparison of droplet counts to gather a ratio of *MDM2* and RNaseP copy number in samples.

Chapter 3: The Genetic Characteristics of Soft Tissue Sarcomas

3.1. Background and aim

A malignant tumour's development and progression is the direct consequence of the accumulation of mutations to oncogenes and tumour suppressor genes within its cells. Historically STSs have broadly been categorised into two groups based their karyotype complexity and their mutational profiles (Mertens et al. 2010).

Despite this there remains a need to further investigate the STS genotype to 1) develop a better understanding of the molecular pathways that contribute to sarcomagenesis 2) identify new potential therapeutic targets and 3) define new ways to stratify/classify STSs to facilitate a more personalised approach to management. The aim of the work outlined in this chapter was to use NGS to characterise the somatic exonic SNVs, indels and CNAs in a cohort of 31 primary STSs.

3.2 Overview experimental approach

For the work described in this chapter samples of fresh frozen STS tissue and buffy coat cells were collected from a cohort of 31 STS patients. DNA was extracted from these tissues and following appropriate QC sequenced using NGS Illumina WES. Next the raw data produced by this analysis were processed using bioinformatic pipelines validated by the use of droplet digital PCR and Nanostring technology to characterise any somatic exonic SNVs, indels, or CNAs present.

3.3 Patient and tumour characteristics

Matched tissue samples from 31 patients (19M: 12F) with a median age of 70 years (range 22-89) were sent for WES (Table 3.1). The STSs analysed included a wide range of histological subtypes which overall had a mean Trojani grade of 2.6 (range 2-3) and a median volume of 190cm^3 (8-4263) when resected.

The management and oncological outcome of patients is reported in Table 3.2. The median duration of patient follow up was 48 weeks (range 6-291). During this period 11 patients (35%) recurred (10 metastatic: 1 local recurrence) at the median of 30 weeks post operatively (range 6-80).

Five patients died during follow up. One of these died of a cause unrelated to their disease 73 weeks post resection. The remaining four patients died of metastatic disease with a median survival of 52 weeks (range 24-97).

3.4 Single nucleotide variants / Indels:

3.4.1 Mutation load

The median number of non-synonymous exonic SNVs identified per tumour was 26 (range 5-55) which equates a mean of 0.43 SNVs /Mb of sequenced DNA. Overall 89% of these were missense substitutions, 7% were nonsense substitutions and 4% were located at splice sites (Table 3.3 / Figure 3.1). Sixty eight percent of the total SNVs identified were transitions (max 72%, min 57%) and 31% were transversions (max 42%, min 27%) (Supplementary 3.1 and 3.2).

The total number of exonic deletions identified in the cohort was 43 which were overall a median length of 4bp (1 - 35). The median number of exonic deletions identified per tumour was 1 (range 0-6) of which the majority caused a frameshift of some kind (67% frames shift versus 33% in frame).

The total number of exonic insertions identified in the analysed tumours was 21 which were a median length of just 1bp (range 1-21). Although one or more insertion(s) were present in 14 of the analysed tumours the median number of exonic insertions/tumour in the whole cohort was 0 (range 0-4). The majority of insertions identified were frameshift in nature (90% frameshift versus 10% non-frameshift) (Table 3.3 / Figure 3.1).

The median number of SNVs and indels combined per tumour was 30 (range 5-58). This mutational load is lower than the majority of other malignancies and similar to the figure previously published on TCGA (Figure 3.2).

Tumour ID (age/sex)	STS Subtype	Tumour Trojani grade	Resected STS volume (cm^3)	Analysed using Nanostring Cancer CN assay
T1 (63/M)	Extra-skeletal myxoid chondrosarcoma	Unknown	588	N
T2 (55/M)	Undifferentiated Pleomorphic Sarcoma	2	8.3	N
T3 (62/F)	Leiomyosarcoma	3	Unknown	Y
T4 (76/F)	Myxofibrosarcoma	3	663	Y
T5 (74/M)	Leiomyosarcoma	2	74	N
T6 (65/M)	Synovial sarcoma	2	65	N
T7 (80/F)	Haemangiosarcoma	Unknown	Unknown	Y
T8 (65/F)	Undifferentiated Pleomorphic Sarcoma	3	364	Y
T9 (69/M)	Myxofibrosarcoma	2	144	Y
T10 (37/F)	Undifferentiated Pleomorphic Sarcoma	3	630	Y
T11 (63/M)	Dedifferentiated Liposarcoma	2	759	Y
T13 (50/M)	Undifferentiated Pleomorphic Sarcoma	2	67	N
T14 (89/F)	Undifferentiated Pleomorphic Sarcoma	2	190	N
T15 (76/F)	Undifferentiated Pleomorphic Sarcoma	3	497	N
T16 (72/M)	Undifferentiated Pleomorphic Sarcoma	3	4263	N
T18 (83/M)	Undifferentiated Pleomorphic Sarcoma	3	2940	N
T19 (67/F)	Undifferentiated Pleomorphic Sarcoma	2	129	N
T20 (74/M)	Leiomyosarcoma	3	525	N
T21 (22/M)	Soft tissue Ewing's sarcoma	3	151	N
T22 (80/M)	Leiomyosarcoma	3	3289	Y
T23 (46/M)	Undifferentiated Pleomorphic Sarcoma	3	2947	Y
T24 (69/M)	Undifferentiated Pleomorphic Sarcoma	2	27	N
T25 (87/F)	Myxofibrosarcoma	3	38	N
T26 (74/F)	Myxofibrosarcoma	2	9	N
T27 (74/M)	Dedifferentiated Liposarcoma	2	576	N
T28 (49/F)	Leiomyosarcoma	2	61	N
T29 (70/M)	Myxofibrosarcoma	3	135	N
T30 (70/M)	Myxofibrosarcoma	3	68	N
T31 (81/F)	Myxofibrosarcoma	3	70	N
T32 (77/M)	Myxofibrosarcoma	2	2160	N
T33 (74/M)	Leiomyosarcoma	3	506	N

Table 3.1: The clinical characteristics of STSs analysed using Illumina Whole Exome Sequencing / the nCounter^{R} V2 Cancer CN assay. **UPS** - undifferentiated pleomorphic sarcoma; **MFS** - myxofibrosarcoma; **LMS** - leiomyosarcoma.

Tumour	(Neo)adjuvant treatment	Resection margins (R0/R1/R2)	Outcome	Follow up (weeks)	Disease free survival (weeks)	Overall survival (weeks)
T1	Adjuvant radiotherapy	R1 (planned)	NED	150	n/a	n/a
T2	Neoadjuvant radiotherapy	R0	AWD	41	41	n/a
T3	Neoadjuvant radiotherapy	R1 (planned)	AWD	6	6	n/a
T4	Neoadjuvant radiotherapy	R1 (planned)	AWD	81	81	n/a
T5	Neoadjuvant radiotherapy	R0	NED	291	n/a	n/a
T6	Adjuvant radiotherapy	R1 (planned)	DOD	15	15	24
T7	Nil	R0	AWD	14	14	n/a
T8	Neo. chemotherapy / Adjuvant radiotherapy	R0	AWD	54	54	n/a
T9	Adjuvant radiotherapy	R1 (unplanned)	DOD	30	30	44
T10	Nil	R0	NED	51	n/a	n/a
T11	Neoadjuvant radiotherapy	R0	NED	86	n/a	n/a
T13	Adjuvant radiotherapy	R0	NED	210	n/a	n/a
T14	Adjuvant radiotherapy	R0	DOD	66	66	73
T15	Adjuvant radiotherapy	R2 (planned)	NED	178	178	n/a
T16	Adjuvant radiotherapy	R0	DOC	97	n/a	97
T18	Neoadjuvant radiotherapy	R0	NED	130	n/a	n/a
T19	Neoadjuvant radiotherapy	R0	NED	69	n/a	n/a
T20	Neoadjuvant radiotherapy	R1 (planned)	NED	61	n/a	n/a
T21	Neo. chemotherapy / Adjuvant radiotherapy	R0	NED	53	n/a	n/a
T22	Neoadjuvant radiotherapy	R0	AWD	32	33	n/a
T23	Neoadjuvant radiotherapy	R1 (planned)	DOD	20	20	23
T24	Nil	R0	AWD	30	30	n/a
T25	Nil	R0	NED	49	n/a	n/a
T26	Nil	R0	NED	49	n/a	n/a
T27	Neoadjuvant radiotherapy	R0	NED	42	n/a	n/a
T28	Adjuvant radiotherapy	R0	NED	40	n/a	n/a
T29	Neoadjuvant radiotherapy	R1 (unplanned)	NED	30	n/a	n/a
T30	Adjuvant radiotherapy	R1 (unplanned)	NED	38	n/a	n/a
T31	Adjuvant radiotherapy	R0	NED	40	n/a	n/a
T32	Neoadjuvant radiotherapy	R0	NED	31	n/a	n/a
T33	Neoadjuvant radiotherapy	R0	NED	34	n/a	n/a

Table 3.2: The management and oncological outcome of STS cases genotyped using WES. Every included patient underwent attempted curative surgery to remove their STS. The resection margins shown are those obtained at the completion of patients' surgical management. R0-wide; R1-marginal; R2-intracapsular. AWD-alive with disease; DOC-died from other causes; DOD-died from disease; NED-no evidence of disease.

	Deletions		Insertions		Substitutions			
Tumour ID	Frame Shift	In Frame	Frame Shift	In Frame	Missense	Nonsense	Splice Site	Total
T30	1	0	2	0	46	6	3	58
T16	3	3	1	0	41	3	1	52
T33	0	0	0	0	45	2	4	51
T22	3	1	2	2	34	1	2	45
T18	1	2	1	0	34	4	2	44
T25	1	0	0	0	38	2	1	42
T3	1	0	1	0	35	1	0	38
T11	0	0	0	0	35	2	0	37
T27	3	1	0	0	30	2	0	36
T8	2	0	2	0	29	3	0	36
T26	1	2	0	0	29	1	1	34
T32	0	0	0	0	29	3	2	34
T9	2	3	0	0	22	3	3	33
T13	3	0	1	0	24	1	2	31
T31	2	0	1	0	27	1	0	31
T19	0	0	0	0	25	4	1	30
T14	0	1	1	0	22	1	0	25
T23	0	1	1	0	19	2	2	25
T7	1	0	1	0	19	2	2	25
T10	2	0	0	0	17	1	3	23
T2	1	0	0	0	20	1	1	23
T15	2	0	0	0	19	1	0	22
T5	0	0	0	0	16	2	0	18
T21	0	0	2	0	12	1	0	15
T24	0	0	2	0	11	1	1	15
T1	0	0	0	0	10	1	0	11
T6	0	0	0	0	9	1	1	11
T20	0	0	0	0	8	0	0	8
T28	0	0	1	0	6	0	1	8
T29	0	0	0	0	8	0	0	8
T4	0	0	0	0	5	0	0	5

Table 3.3: The Single Nucleotide Variant and Insertion/deletion characteristics of tumours analysed using WES. Tumours are ranked in descending order by the total number of mutations they contained.

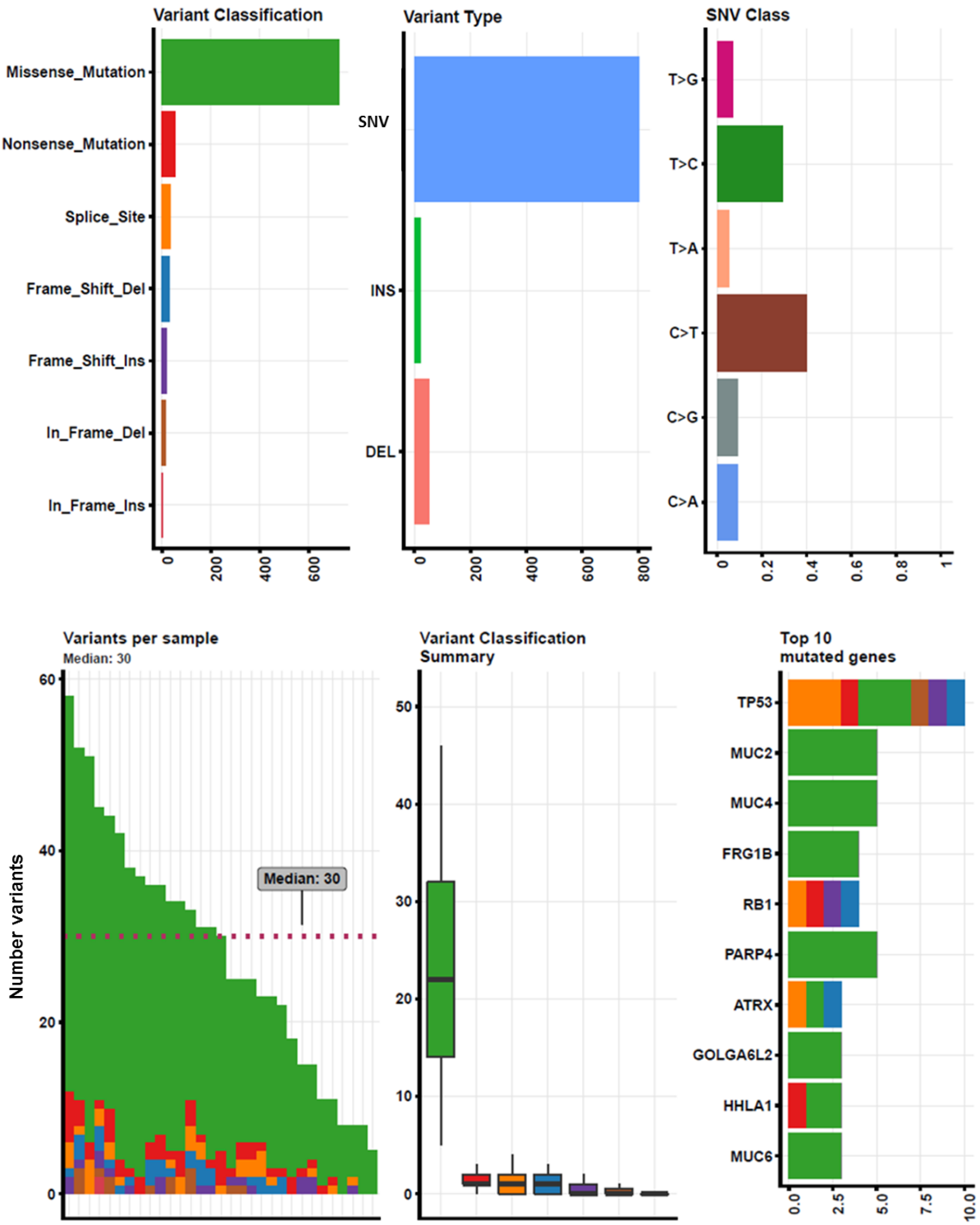


Figure 3.1: Graphical summary of the Single Nucleotide Variant and Insertion/deletion characteristics of the STSs analysed using WES. The variants shown in the graphs 'Variants per sample' 'Variant Classification summary' and 'Top 10 mutated genes' are shown in the same colours as they are designated in the graph entitled 'Variant classification'.

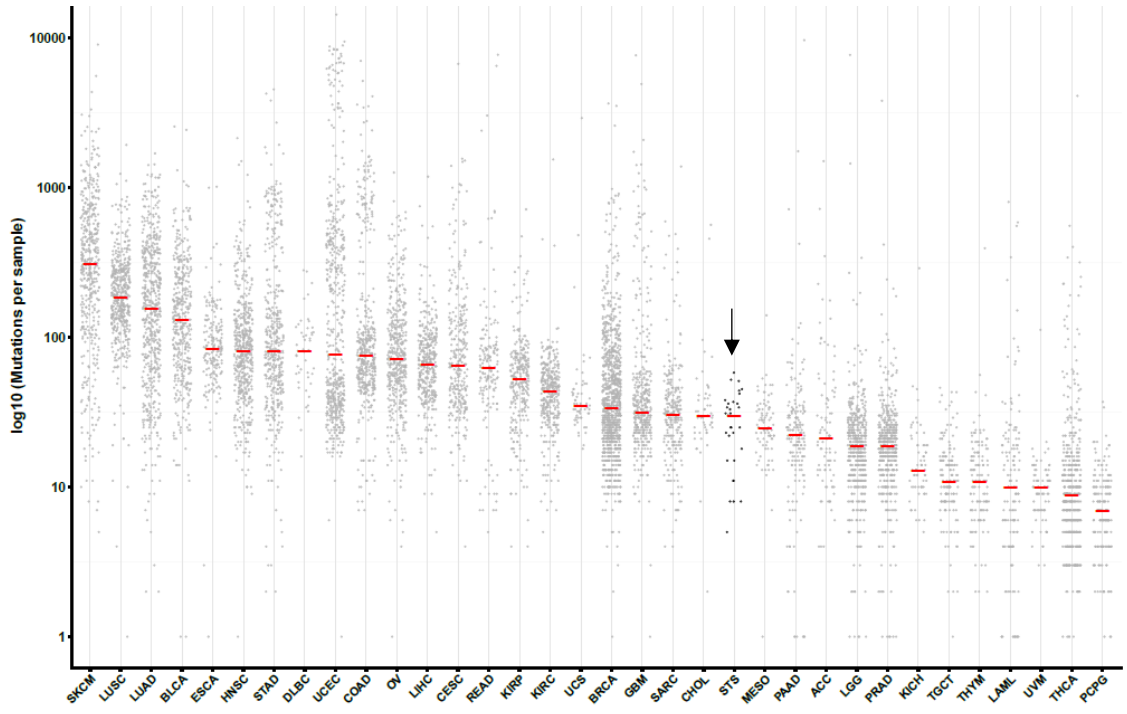


Figure 3.2: A scatter plot showing the SNV/indel mutational load of the STSs analysed compared with other malignancies included on The Cancer Genome Atlas. The x-axis represent the different tumour types and y-axis the number of exonic SNVs/indels in a logarithmic scale. **STSs**: The analysed soft tissue sarcomas (black arrow): **LAML**-Acute Myeloid Leukaemia: **ACC**-Adrenocortical carcinoma: **BLCA**-Bladder Urothelial Carcinoma: **LGG**-Brain Lower Grade Glioma: **BRCA**-Breast invasive carcinoma: **CESC**-Cervical squamous cell carcinoma and endocervical adenocarcinoma: **CHOL**-Cholangiocarcinoma: **COAD**-Colon adenocarcinoma: **ESCA**-Esophageal carcinoma: **GBM**-Glioblastoma multiforme: **HNSC**-Head and Neck squamous cell carcinoma: **KICH**-Kidney Chromophobe: **KIRC**-Kidney renal clear cell carcinoma: **KIRP**-Kidney renal papillary cell carcinoma: **LIHC**-Liver hepatocellular carcinoma: **LUAD**-Lung adenocarcinoma: **LUSC**-Lung squamous cell carcinoma: **DLBC**-Lymphoid Neoplasm Diffuse Large B-cell Lymphoma: **MESO**-Mesothelioma: **OV**-Ovarian serous cystadenocarcinoma: **PAAD**-Pancreatic adenocarcinoma: **PCPG**-Pheochromocytoma and Paraganglioma: **PRAD**-Prostate adenocarcinoma: **READ**-Rectum adenocarcinoma: **SARC**-Sarcoma: **SKCM**-Skin Cutaneous Melanoma: **STAD**-Stomach adenocarcinoma: **TGCT**-Testicular Germ Cell Tumours: **THYM**-Thymoma: **THCA**-Thyroid carcinoma: **UCS**-Uterine Carcinosarcoma: **UCEC**-Uterine Corpus Endometrial Carcinoma: **UVM**-Uveal Melanoma.

3.4.2 Mutated Genes

783 genes were found to contain at least 1 SNV/indel. The most commonly mutated of these were *TP53* (10 different mutations, 10 different tumours), *MUC2* (5, 4), *MUC4* (5, 4), *FRG1B* (4,4), *RB1* (4,4), *PARP4* (5,3) and *ATRX* (3,3) (Figure 3.3 and Supplementary 3.3).

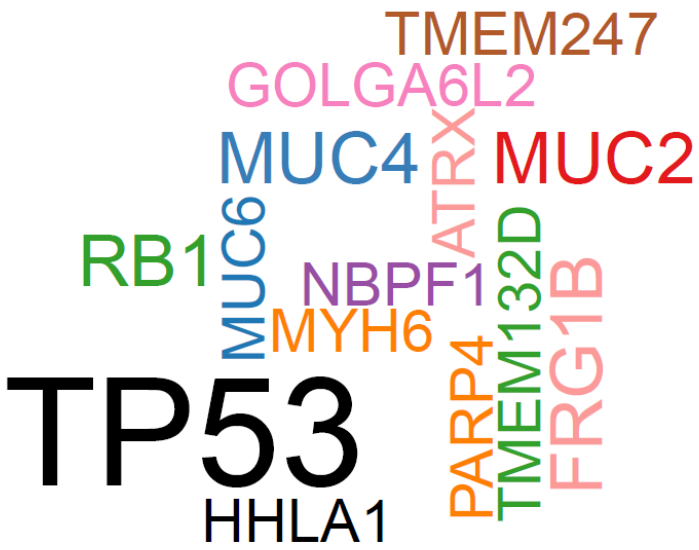


Figure 3.3: Genecloud showing the genes most commonly found to contain SNVs/indels in the analysed STSs. All genes found to contain ≥ 3 SNVs/indels are shown. The size of each gene's name is proportional to the number of variants they contained.

3.4.3 Driver mutations

The only gene identified in the analysed cohort as having SNV/indel characteristics statistically indicative of a driver gene was *PARP4* (P<0.0001, OncodriveCLUST). Two different *PARP4* missense substitutions were identified a total of 5 times in 3 different tumours (Gln1059Arg , n=3 / Ile1039Thr, n=2) (Figure 3.4).

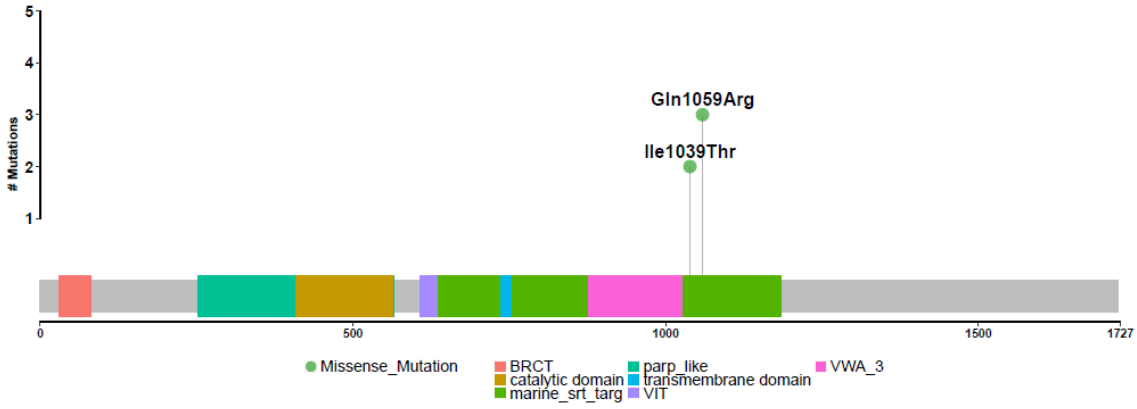


Figure 3.4: SNV characteristics of Poly(ADP-Ribose) Polymerase Family Member 4 (*PARP4*). A lollipop plot is shown highlighting the hotspot locations of the mutations identified during our analysis. The two different missense substitutions identified are shown with their frequencies and predicted effects. **BRCT** –BRCT domain, **PARP_like** – Poly (ADP-ribose) polymerase catalytic domain, VWA**_3 -** Von Willebrand factor type A domain, Marine**_srt_target -** Marine proteobacterial sortase target protein, **VIT** - Vault protein inter-alpha-trypsin domain.

3.4.4 Mutation co-occurrence/ exclusivity

An analysis of mutation that are likely to interaction identified 8 pairs of genes that were mutated in the same tumours in statistically significant manner ($P<0.05$) (Table 3.4 and Figure 3.5).

		Tumour number				
Gene1	Gene2	Neither Gene mutated	Genes 1 and 2 mutated	Gene 2 mutated	Gene 1 mutated	P Value
TMEM132D	*CACNG4*	28	2	0	1	0.006
CACNG4	*FRG1B*	27	2	2	0	0.013
TMEM247	*HHLA1*	27	2	1	1	0.019
NBPF1	*MUC6*	27	2	1	1	0.019
HHLA1	*TP53*	21	3	7	0	0.027
TMEM247	*TP53*	21	3	7	0	0.027
MUC4	*MYH6*	26	2	1	2	0.037
FRG1B	*TMEM132D*	26	2	1	2	0.037

Table 3.4: Pairs of genes found to contain SNV/indels in a co-existing manner in the analysed tumours. Analysis was performed using a pair-wise Fisher's Exact test to look for significant pairs of genes and the cometExactTest to look for significantly altered gene sets involving >2 genes, none of which were identified. No gene pairs/sets were identified as containing mutations in mutually exclusive way.

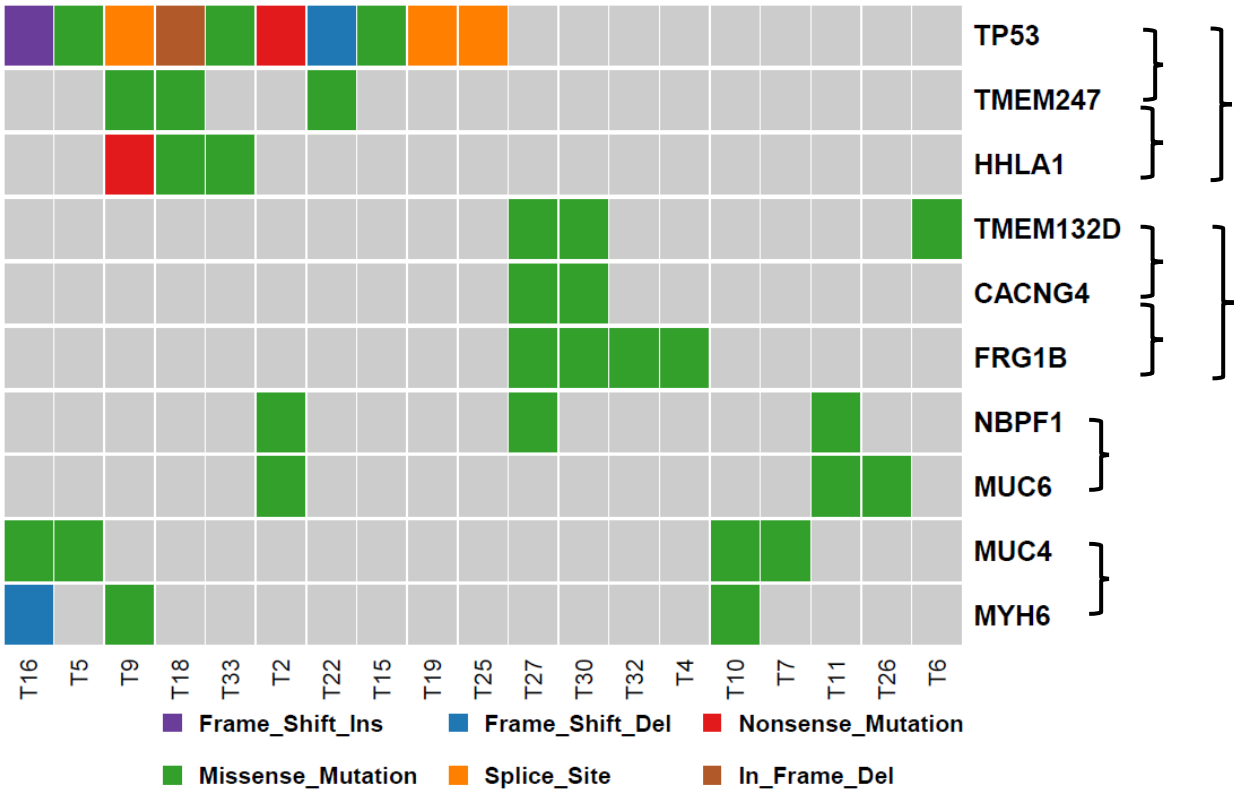


Figure 3.5: An oncostrip highlighting any gene pairs found to contain mutations in a co-existing manner. The analysed tumours are shown on the X axis with pairs of genes highlighted with braces on the right of the figure.

3.4.5 Mutational signature

Mutational analysis identified 3 discrete SNV signatures in the analysed STSs. These were numbered signatures 1, 2 and 3 (Figure 3.6.) When these signatures were compared to other signatures previously reported in cancer (Alexandrov et al. 2013) similarities were seen with 2 previously reported signatures (numbered 1 and 5.)

3.4.6 Subtype comparisons

The average number of SNV/indels per tumour varied between STS subtypes although not significantly (Figure 3.7, Supplementary 3.4-3.7). The highest mean number of SNVs/indels was seen in the dedifferentiated liposarcomas (mean number SNVs/indels 37, range 15-52, number of cases 11), followed by myxofibrosarcomas (31, 5-58, 8), undifferentiated pleomorphic sarcomas (30, 15-52, 11) (in which no indels were identified), leiomyosarcomas (28, 8 -51, 6), haemangiosarcomas (25, n/a, 1), soft tissue Ewing's sarcoma (15, n/a, 1) then synovial sarcoma (11, n/a, 1). In those subtypes in which >1 tumour was analysed a wide range of genes were found to contain mutations (Supplementary 3.8-3.11). None of these showed a clear propensity to mutate in a particular subtype(s) (Figure 3.8).

No additional mutational signatures or differences between signatures were identified when tumours were subcategorised and re-analysed according to their histological subtype.

3.4.7 Neoadjuvant treatment

No significant difference was seen in the mutational load of those tumours that received neoadjuvant radiotherapy (n=14), chemotherapy (n=2) or no pre-operative treatment (n=15) (Figure 3.9 Supplementary 3.12/3.13). No genes were found to be statically more commonly mutated following the use of neoadjuvant treatment. Mutational signature analysis revealed no difference in the signatures of those tumours that did and did not received neoadjuvant treatment, and no new signatures in either group.

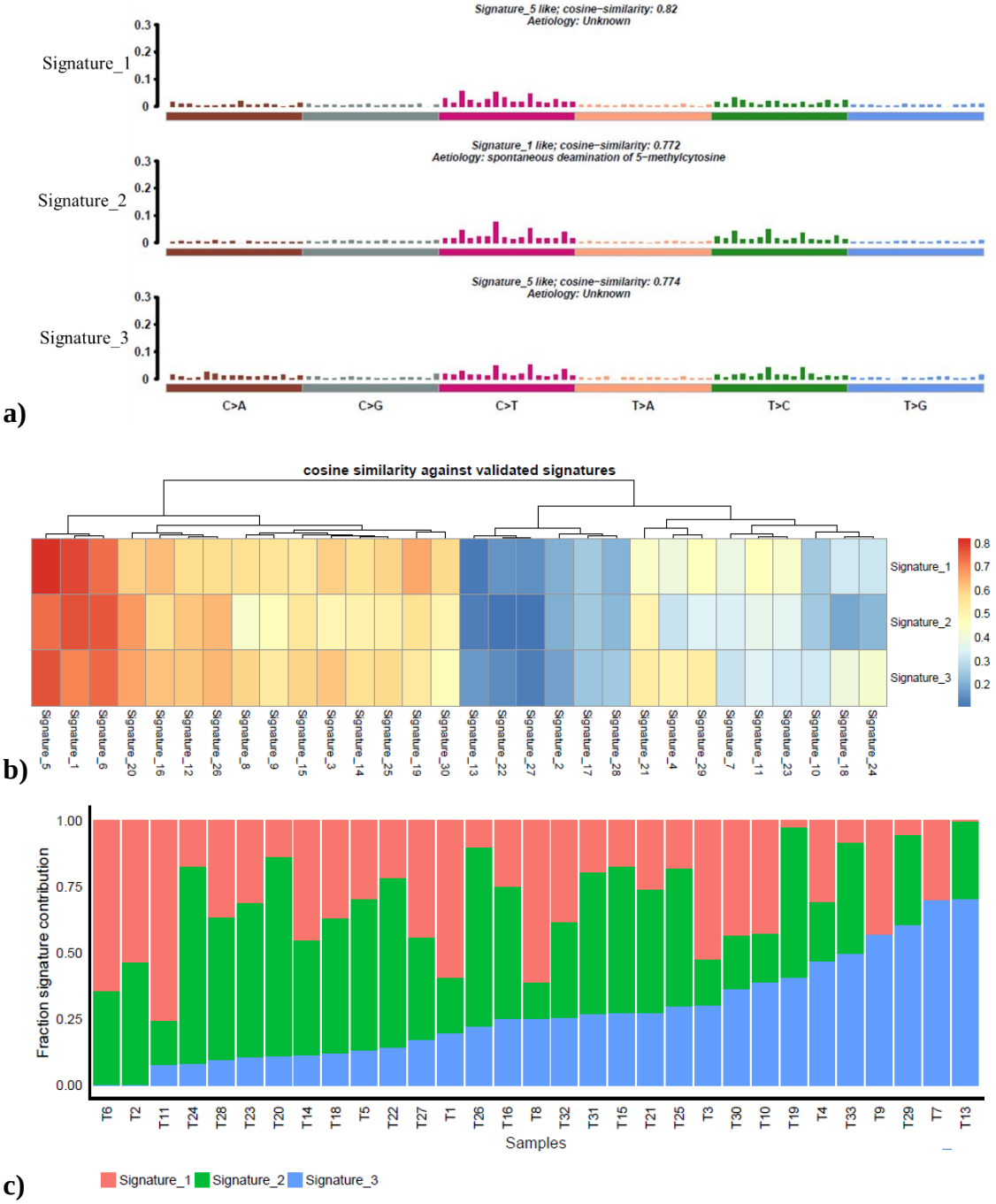


Figure 3.6: Mutational signatures identified in the analysed tumours. 3.6a shows the SNV characteristics of the 3 signatures identified in the STS cohort analysed numbered 1, 2 and 3. 3.6b is a heatmap which compares these signatures with the 30 validated different signatures previously identified in cancer, The strength of the similarities between these are represented by cosine similarity values ranging 0 (weakest, blue) to 1 (strongest, red). 3c is a barplot showing the contributions that each of the 3 signatures identified made in each analysed tumour.

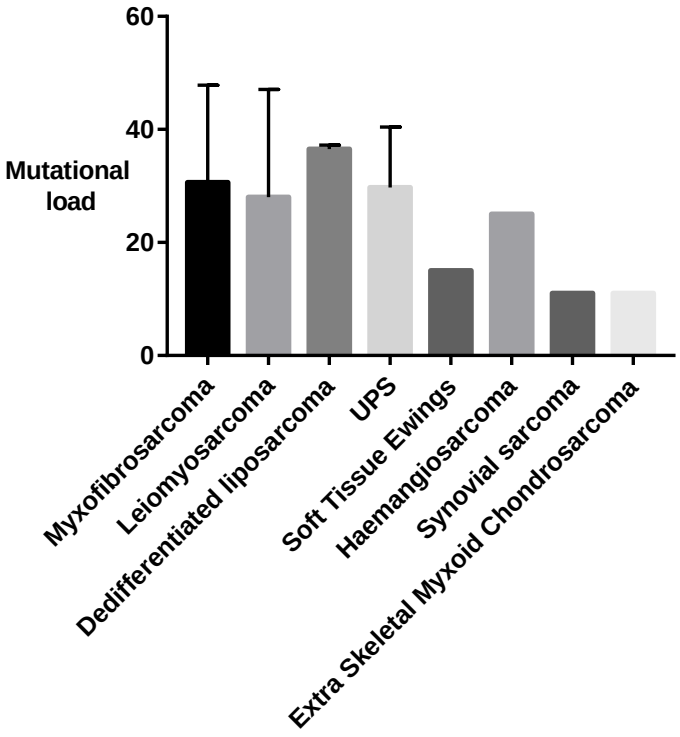


Figure 3.7: Mutational load of STSs analysed. Tumours are categorised according to their histological subtype. The y axis represent the mean number of SNV/indels per tumour and error bars show the SEM. Multiple comparisons using ANOVA revealed no significant differences. Myxofibrosarcomas (n=8), leiomyosarcomas (n=6), dedifferentiated liposarcomas (n=2), undifferentiated pleomorphic sarcoma (n=11). **UPS**- undifferentiated pleomorphic sarcoma.

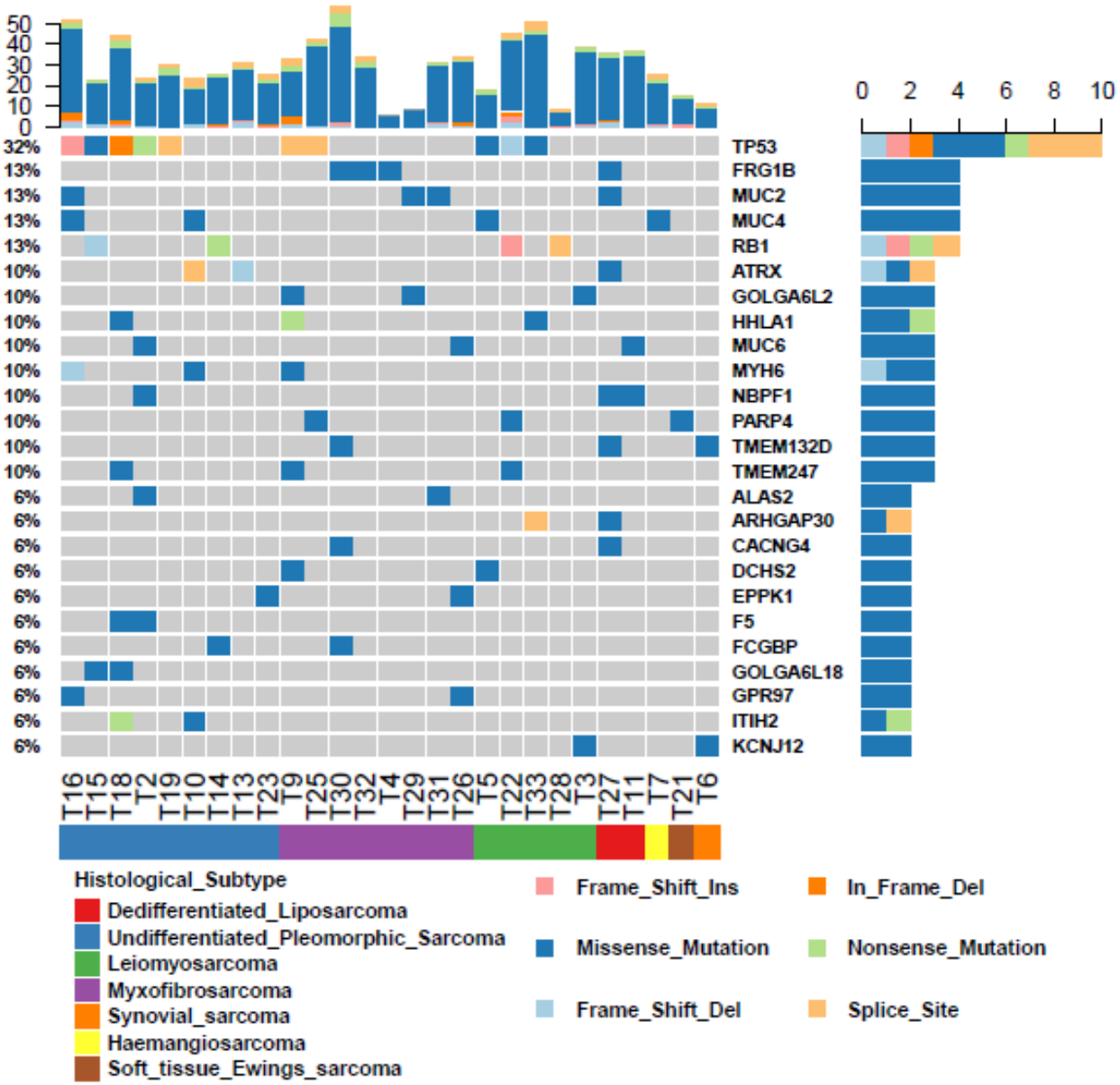


Figure 3.8 Oncoprint showing the key genes altered in the analysed STS samples. The top 25 most commonly mutated genes are shown with tumours grouped by their histological subtype. Four tumours are not shown as they did not contain any mutations in the top 25 most commonly altered genes (T1, T8, T20 and T24).

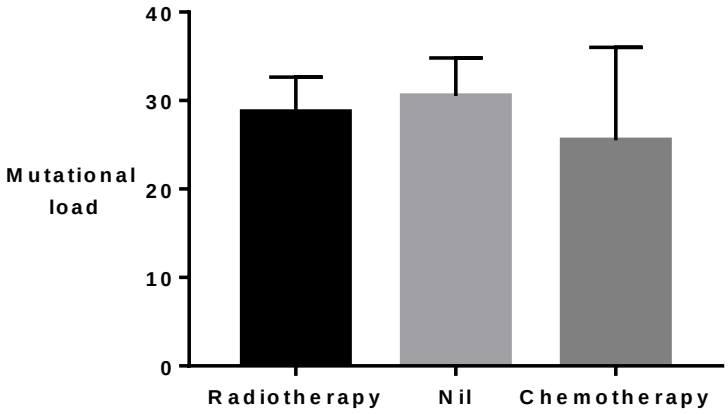


Figure 3.9: Mutational load of STSs analysed categorised by neoadjuvant treatment. The y axis represent the mean number of SNV/indels per tumour with error bars showing the SEM. Fourteen patient received neoadjuvant radiotherapy, two neoadjuvant chemotherapy and 15 received no pre-operative treatment. Multiple comparisons using ANOVA revealed no significant differences.

3.4.8 Oncological outcome comparison

There was no significant difference in the mutational SNV/indel load of those tumours that did and did not go onto recur during follow up (Figure 3.10, Supplementary 3.14/3.15). Statistical analysis failed to identify any genes that were predictive of recurrence i.e. significantly more mutated in those tumours that recurred.

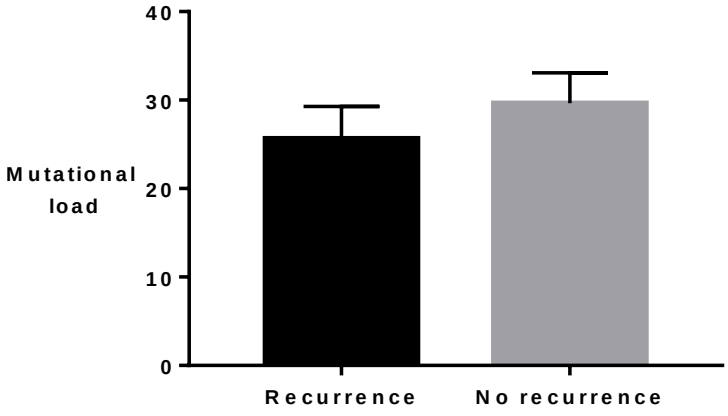


Figure 3.10: Mutational load of STSs analysed categorised by outcome. The y axis represent the mean number of SNV/indels per tumour with error bars showing the SEM. No significant difference was seen in the mutational load of each group (P=0.46, paired t-test). A retrospective analysis indicates that the power for this comparison was just 12.4% ($\alpha = 0.05$).

3.5 Genome complexity

A visual assessment of each tumour's exomes showed that 4/31 tumours (13%) had simple, almost diploid genomes (defined as containing 5 or less exonic regions with an abnormal copy number.) These cases included one extra-skeletal myxoid chondrosarcoma, one myxofibrosarcoma, one synovial sarcoma and one haemangiosarcoma (Table 3.5 and Supplementary 3.16). The median Sequenza estimation of ploidy in the cohort was 3.5 (range 1.9-6.8).

In two tumours the witnessed genome complexity did not match the predicted complexity based on their histopathological subtype. These included one case of myxofibrosarcoma (predicted complex, witnessed simple) and one case of soft tissue Ewing's sarcoma (predicted simple, witnessed complex).

Three of the four tumours with simple karyotypes (75%) recurred during follow up (one of which died of their disease during follow up). Eight of the 27 tumours (30%) with complex karyotypes recurred during follow up (of which 3 died of their disease). The remaining 19 tumours with complex karyotypes remained disease free during follow up.

Tumour	Subtype	Ploidy	Predicted genome complexity	Witnessed genome complexity
T1	ESMC	2	Simple	Simple
T2	UPS	5.6	Complex	Complex
T3	Leiomyosarcoma	6.8	Complex	Complex
T4	**Myxofibrosarcoma**	**2.2**	**Complex**	**Simple**
T5	Leiomyosarcoma	2.2	Complex	Complex
T6	Synovial sarcoma	2.2	Simple	Simple
T7	Haemangiosarcoma	2.1	Simple	Simple
T8	UPS	5.1	Complex	Complex
T9	Myxofibrosarcoma	6.6	Complex	Complex
T10	UPS	3.8	Complex	Complex
T11	Dedifferentiated Liposarcoma	5.2	Complex	Complex
T13	UPS	2.3	Complex	Complex
T14	UPS	5.3	Complex	Complex
T15	UPS	3.5	Complex	Complex
T16	UPS	2.4	Complex	Complex
T18	UPS	5.7	Complex	Complex
T19	UPS	3.4	Complex	Complex
T20	Leiomyosarcoma	2	Complex	Complex
T21	**Soft tissue Ewing's sarcoma**	**2.6**	**Simple**	**Complex**
T22	Leiomyosarcoma	5.4	Complex	Complex
T23	UPS	1.9	Complex	Complex
T24	UPS	6	Complex	Complex
T25	Myxofibrosarcoma	5.2	Complex	Complex
T26	Myxofibrosarcoma	3	Complex	Complex
T27	Dedifferentiated Liposarcoma	2.5	Complex	Complex
T28	Leiomyosarcoma	2.2	Complex	Complex
T29	Myxofibrosarcoma	5.9	Complex	Complex
T30	Myxofibrosarcoma	4.9	Complex	Complex
T31	Myxofibrosarcoma	6	Complex	Complex
T32	Myxofibrosarcoma	2	Complex	Complex
T33	Leiomyosarcoma	5	Complex	Complex

Table 3.5: The predicted and witnessed genome complexities of the analysed tumours. Two tumours' genome complexities did not match those predicted by the current literature based on their histological subtype (highlighted in bold). The predicted ploidy of each tumour (Sequenza analysis) is also shown in the table. ESMC - Extra-skeletal myxoid chondrosarcoma, UPS - Undifferentiated Pleomorphic Sarcoma.

3.6 Copy Number Alterations

3.6.1 Sequenza analysis validation

A subset of patients' tumour and BC DNA was analysed using Nanostring technology and/or ddPCR to validate the Sequenza analysis performed.

3.6.1.1 Nanostring nCounter® v2 Cancer Copy Number (CN) Assay

Matched tumour and BC DNA was sent from 9 patients for analysis using the nCounter^{R} V2 Cancer CN assay (see Table 3.1 for clinical details).

No amplified or deleted genes were identified in 4 of these tumours (T4, T7, T9 and T23) (Figure 3.11). A review of WES data for these tumours showed excellent correlation, with none of the genes investigated on the Cancer CN assay identified as amplified or deleted.

One tumour (T3) was found to have an isolated deletion of *TP53* by the Cancer CN assay. This was also identified in the WES data (CN=0) (Figure 3.12).

The remaining 4 tumours had ≥1 amplified gene(s) identified by the Cancer CN assay (Table 3.6). Three of these tumours had a single amplified gene identified (T8:*CCNE1*; T10:*MAPK7*; T22:*MCL1*) and one had 2 amplified genes detected (T11;*CCNE1 MDM2*). All of these amplifications were also identified by Sequenza (Figures 3.13-3.16).

3.6.1.2 Droplet digital PCR

To investigate discrepancies between the actual copy number values of certain genes produced by Sequenza and the Nanostring Cancer CN assay the copy number of *MDM2* in T11 was investigated using ddPCR (Figure 3.17). The *MDM2* copy number based on this analysis was 24, compared with 10 based on the Cancer CN assay and 20 based on Sequenza. The *MDM2:RPPH1* (control) copy number ratio based on the ddPCR analysis performed was 6:1. This compared to a ratio of 5:1 based on our Sequenza analysis.

Tumour	Gene	Ploidy	Nanostring (copy number)	Sequenza (copy number)
T8	*CCNE1*	5.1	10	20
T10	*MAPK7*	3.8	4	7
T11	*CCNE1*	5.2	5	4
	MDM2		10	20
T22	*MCL1*	5.4	4	11

Table 3.6: Amplified genes identified by Sequenza and the Nanostring Cancer CN assay but with discordant absolute copy number values. *MDM2* was located between 2 segments in T11 with copy numbers of 4 and 20. The latter was assumed to be accurate based on subsequent ddPCR analysis of T11 shown in Figure 3.17.

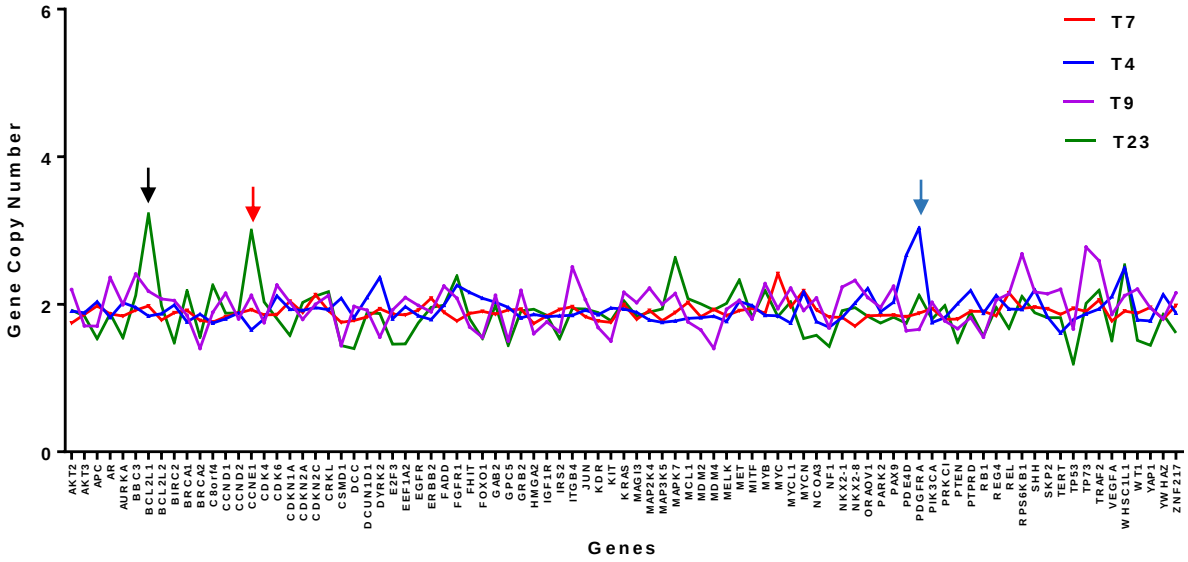


Figure 3.11: Scatter plot showing the Nanostring Cancer CN assay data for those tumours with no significant CNAs identified. The 87 genes investigated by the Cancer CN assay are shown on the x-axis with gene copy number shown on the y axis. The mean copy number of *BCL2L1* was 3.2 (black arrow) and *CCNE1* was 3.0 (red arrow) in T23. The mean copy number of *PDGFRA* in T4 was 3.0 (blue arrow).

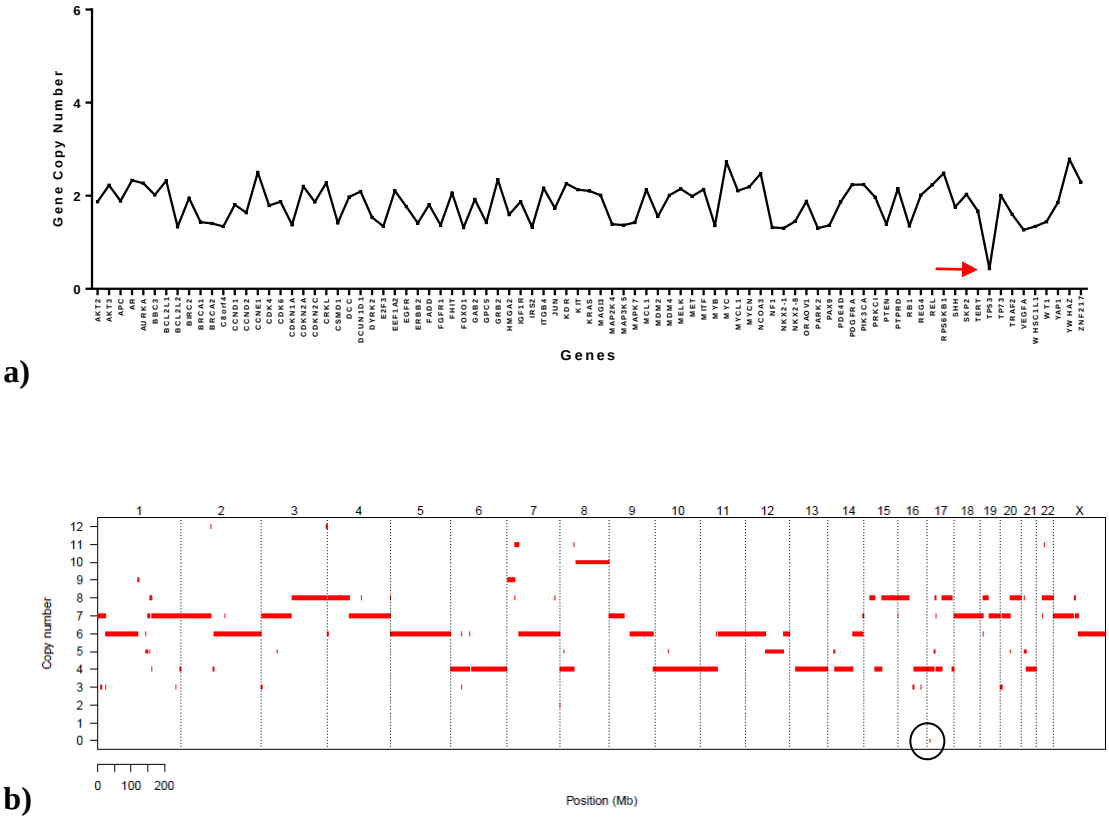
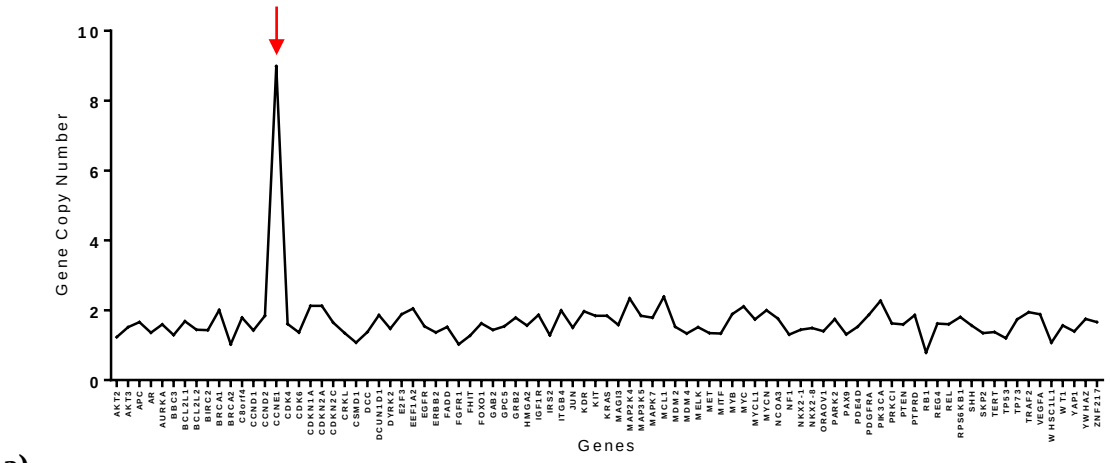
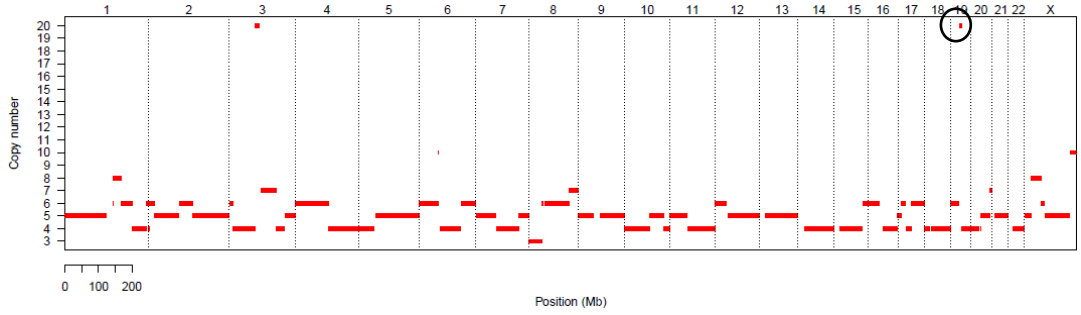


Figure 3.12: Validation of the Sequenza analysis of T3 copy number using the nCounter v2 Cancer CN assay. 3.15a is a scatter plot that shows the results of the Cancer CN assay analysis of T3. The plot's x axis shows the 87 genes included on the assay and the y-axis represents gene copy number. 3.15b is an exome wide plot produced during the analysis of NGS data by Sequenza showing copy number throughout T3's genome. *TP53* (located at chr17:7,661,779-7,687,550) is shown in a segment with a CN of 0 (black circle). The boundaries of this 170kb deleted segment were chr17:7557419-7728246.



a)



b)

Figure 3.13: Validation of the Sequenza analysis of T8 copy number using the nCounter v2 Cancer CN assay. 3.16a is a scatter plot that shows the results of the Cancer CN assay analysis of T8 highlighting the *CCNE1* amplification seen (red arrow). The plot's x axis shows the 87 genes analysed and the y axis the copy number of these genes. 3.16b is an exome wide plot of copy number produced during the analysis of T8's WES data by Sequenza. *CCNE1* (located at chr19:29,811,898-29,824,312) is present within a 12.4kb genomic segment with a CN of 20 (black circle). The boundaries of this segment were chr19: 27732056-31930388.

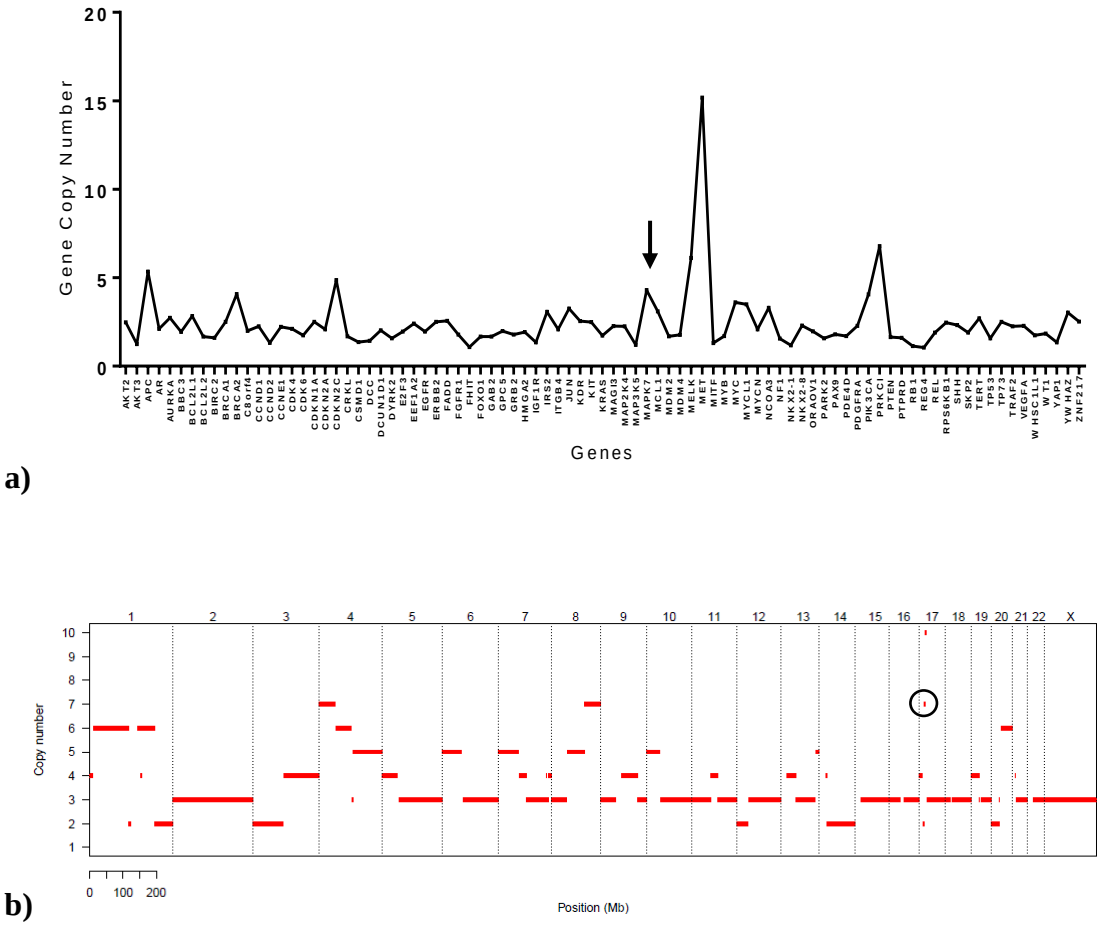
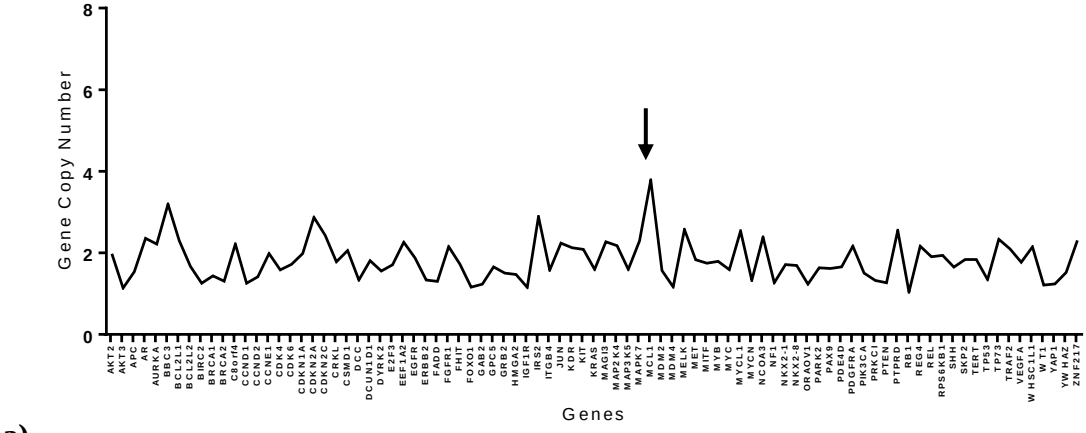
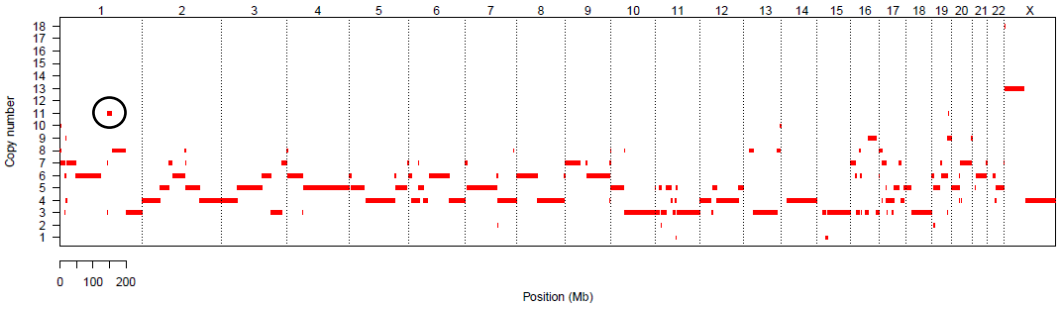


Figure 3.14: Validation of the Sequenza analysis of T10 copy number using the nCounter v2 Cancer CN assay. 3.17a is a scatter plot that shows the results of the nanostring CN assay analysis of T10. The plot's x axis shows the 87 genes analysed and the y axis the copy number of these genes. Although 7 genes had a mean copy number >4 only one (*MAPK7*) had amplification of all three probes included in the assay (probe copy numbers 3.9, 5.3, 3.7, black arrow). The mean copy numbers of the remaining 6 genes were skewed significantly by a single amplified probe (*APC*: 12.2, 1.2, 1.0 / *BRCA2:* 8.7, 2.3, 1.2 /*CDKN2C:* 2.7, 8.8, 3.0 / *MELK:* 14.6, 2.0, 1.7 / *MET:* 2.8, 41.6, 1.2 / *PRKCL:* 1.8, 16.0, 2.6l). 3.17b is an exome wide plot produced during the analysis of NGS data by Sequenza showing copy number throughout T10's exome. *MAPK7* (located at chr17:19,377,721-19,383,544) is present within a genomic segment with a CN of 7 (black circle). The boundaries of this 4460kb segment were chr17: 16229117-20688819.

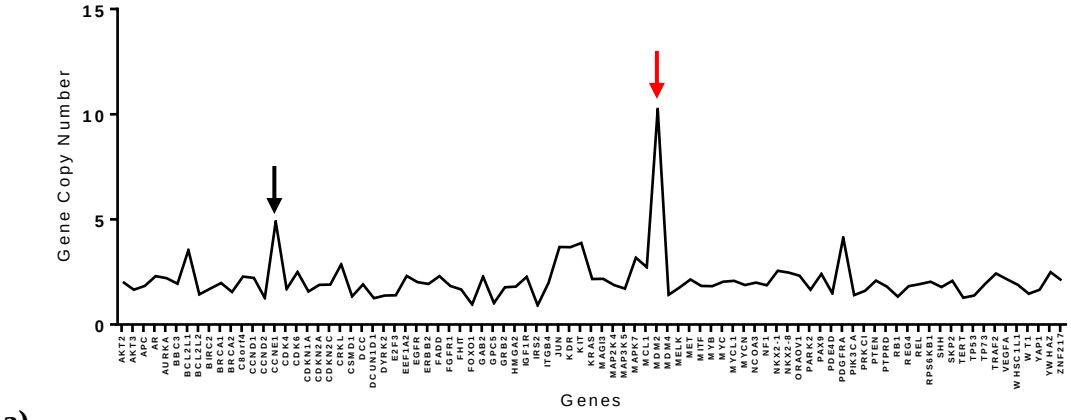


a)

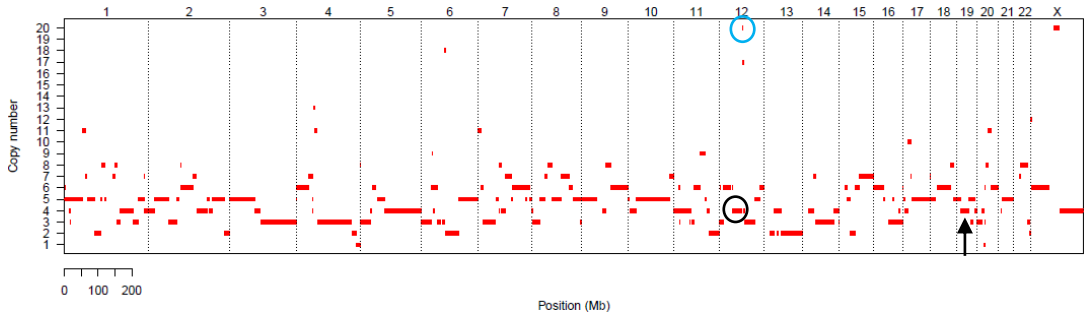


b)

Figure 3.15: Validation of the Sequenza analysis of T22 copy number using the nCounter v2 Cancer CN assay. 3.18a is a scatter plot that shows the results of the Cancer CN assay analysis of T22. The plot's x axis shows the 87 genes analysed and the y axis the copy number of these genes. The only amplified gene highlighted during this analysis was *MCL1* (mean probe copy number 4.0, black arrow). 3.18b is an exome wide plot of copy number produced during the analysis of T22's WES data by Sequenza. *MCL1* (located at chr1:150,574,551-150,579,738) is present within a 12123kb genomic segment with a CN of 11 and boundaries of chr1: 145,366,463-157,489,516 (black circle).

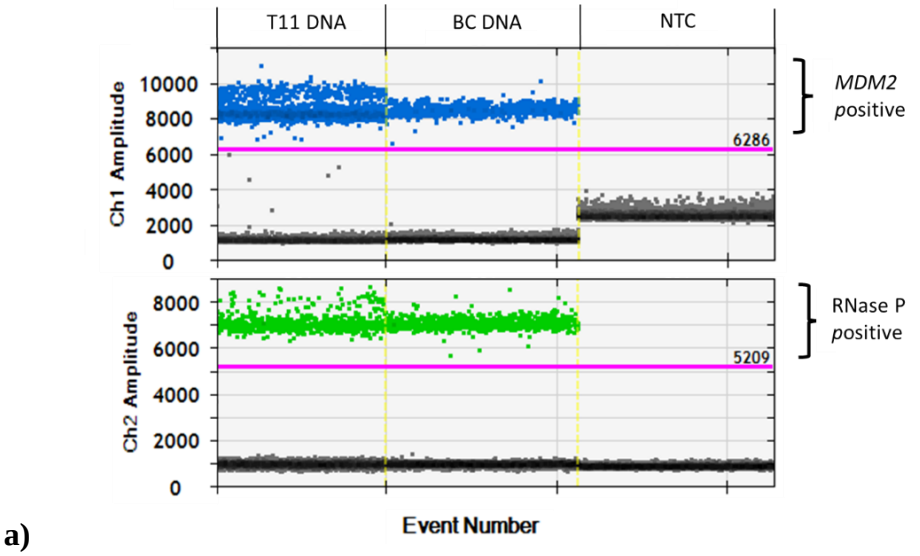


a)

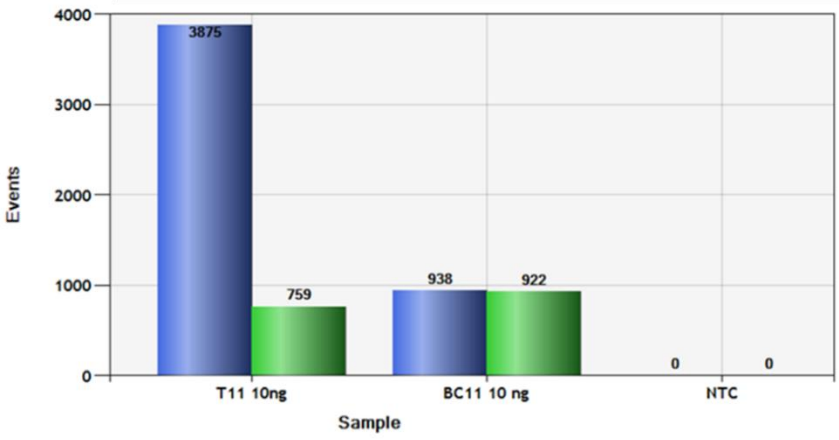


b)

Figure 3.16: Validation of the Sequenza analysis of T11 copy number using the nCounter v2 Cancer CN assay. 3.19a is a scatter plot that shows the results of the Cancer CN assay analysis of T22. The plot's x axis shows the 87 genes analysed and the y axis the copy number of these genes. Two genes were identified as amplified during this analysis - *CCNE1* (mean probe copy number 4.9, black arrow) and *MDM2* (mean probe copy number 10.2, red arrow). 3.19b is an exome wide plot of copy number produced during the analysis of T11's WES data by Sequenza. *CCNE1* (located at chr19:29,811,898-29,824,312) is present within a 9273kb genomic segment with a CN of 4 and boundaries of chr19: 27731991-37005298 (black arrow). *MDM2* (located at chr12:68,808,168-68,850,686) was present between 2 adjacent segments with copy numbers of 4 (chr12: 53647373-68720335, black circle) and 20 (chr12: 68946732-69645864, blue circle).



a)



b)

Figure 3.17: Droplet digital PCR copy number analysis of *MDM2* in T11. 3.20a shows the droplet counts of *MDM2* and *RPPH1* in samples of T11, matched BC DNA and a non-template control (NTC) reaction. *RPPH1* encodes for RNaseP and was used as a control to allow calculation of the copy number of *MDM2*. Cumulative droplet count (event number) is shown on the X axes with each interval representing 1000 droplets. The fluorescent amplitudes of the 2 channels used are shown on the Y axes (channel 1/FAM/MDM2 probe, channel 2/HEX/RPPH1 probe.) 3.20b is a bar chart comparing MDM2 and RPPH1 droplet counts in each reaction. The droplet count of each probe in each sample is shown above each relevant bar. The ratio of *MDM2:RPPH1* positive droplets in T11 was 6.2:1 and 1:1 in the BC DNA. Based on WES data the copy number of 87bp amplicon targeted in RPPH1 in T11 was 4 (located at chr.14:20811565). If this is assumed accurate then the copy number of MDM2 in T11 is 24. This compares to a copy number of 10 based on the Cancer CN assay analysis and 20 based on WES data analysis using Sequenza.

3.6.2: Mutated genomic segments

Numerous exonic segments were found to be amplified or deleted in multiple tumours in the cohort (Figures 3.18/3.19). A complete list of these segments are shown in supplementary files 3.17 and 3.18 including the cytobands and boundaries of each segment and the genes found within them.

Analysis with GISTIC2 identified 17 exonic segments commonly amplified in our cohort (present in 6-26 tumours) and 5 exonic segments that were regularly deleted (lost in 13-20 tumours). Although 9 of these amplifications and 2 of the deletions were present in ≥50% of tumours (Table 3.6), q values revealed that only 5 amplifications and 3 deletion were present to a significant degree within the cohort.

Of these, those CNAs identified in >50% of cases involved 7p22.1 (amplified in 21/31cases) 17p11.2 (amplified in 20/31 cases), 17q25.3 (amplified in 18/31 cases), 1q21.3 (amplified in 16/31 cases), 1q42.2 (deleted in 20/31 cases) and 6q14.1 (deleted in 19/31 cases).

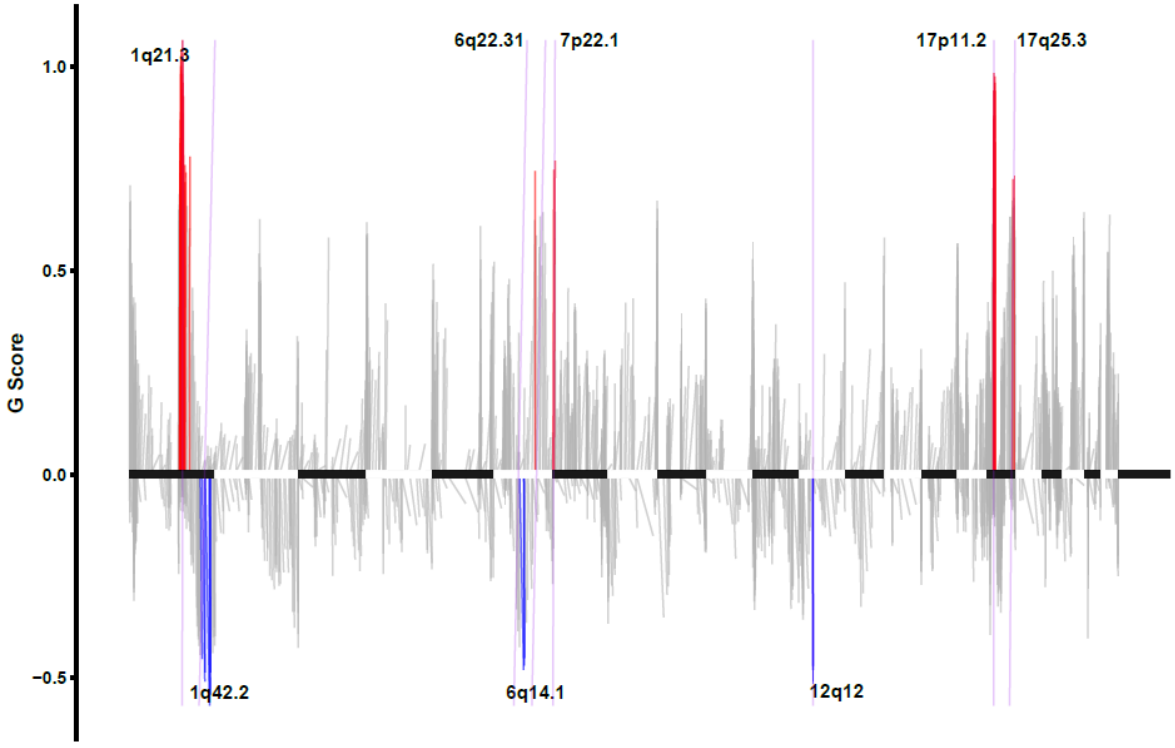


Figure 3.18: Exonic plot showing the commonly amplified/deleted segments identified in the analysed tumours. The segments G scores are shown on the y axis and segment location within the genome on the X axis. All altered segments with q values < 0.1 are highlighted in red (amplification) or blue (deletion).

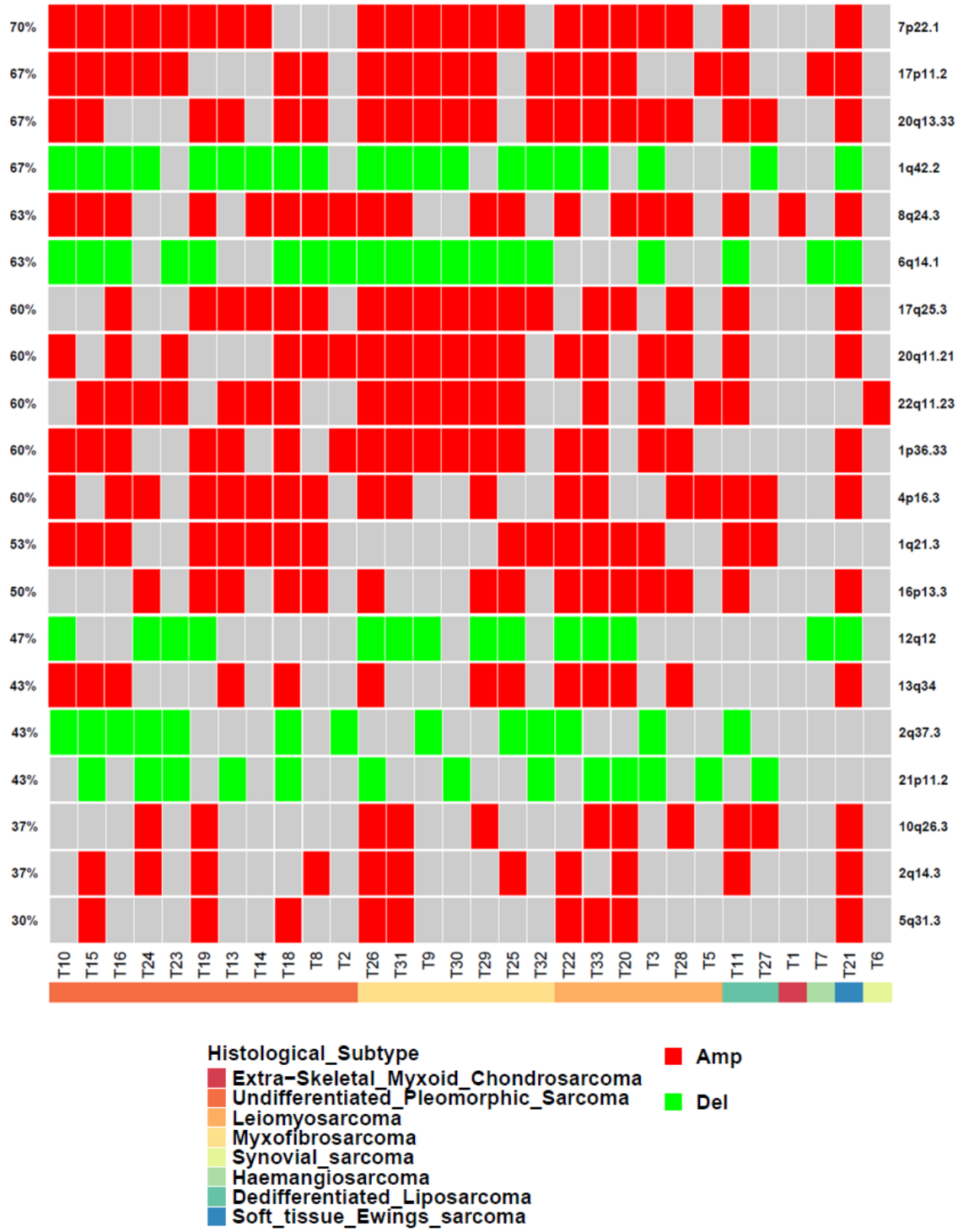


Figure 3.19: Oncoprint showing the commonly amplified and deleted exonic segments identified in the analysed tumours (n=30). Amplified regions were defined by GISTIC 2 as being having a log2 ratio above 0.1. Conversely deleted regions were defined as having a log2 ratio below -0.1. Tumours are ranked according to the most commonly altered genomic regions with the top 20 most commonly altered segments shown.

Genomic region (number genes in region)	Copy number alteration	Region limits (wide peak)	Genes of interest in region
1p36.33 (80)	Amp	chr1:1-2393331	
1q21.3 (70)	**Amp**	**chr1:153723055-155668645**	**MUC1, TPM3**
2q14.3 (6)	Amp	chr2:128929890-130842194	
3p12.1 (19)	Amp	chr3:81688158-97550998	
4p16.3 (58)	Amp	chr4:1-2965694	WHSC1, SPON2
5q31.3 (29)	Amp	chr5:140616746-140940005	PCDHGC3
6q22.31 (209)	**Amp**	**chr6:109875295-149517318**	**MYB, PTPRK, ROS1, TNFAIP3, CDK19, GOPC, FAM162B, ECT2L**
7p22.1 (17)	**Amp**	**chr7:1-6451668**	**RAC1, CARD11**
8q24.3 (86)	Amp	chr8:144155127-146364022	RECQL4
10q26.3 (1)	Amp	chr10:135438952-135440319	
13q34 (29)	Amp	chr13:112268546-115169878	
16p13.3 (98)	Amp	chr16:1263714-2995709	TSC2, MAPK8IP3, TRAF7, C16ORF42
17p11.2 (11)	**Amp**	**chr17:17938534-18303040**	**TOP3A**
17q25.3 (245)	**Amp**	**chr17:64215537-79977293**	**H3F3B, ITGB4, LLGL2, SRSF2, RNF213, ASPSCR1**
20q11.21 (11)	Amp	chr20:29647500-30163748	
20q13.33 (82)	Amp	chr20:60919127-63025520	PTK6
22q11.23 (22)	Amp	chr22:24729391-26166941	
1q42.2 (410)	**Del**	**chr1:202904412-249250621**	**FH, H3F3A, MDM4, TRAF5, AKT3, SLC45A3**
2q37.3 (177)	Del	chr2:225771687-243199373	COL4A3, SPP2, TRIP12
6q14.1 (114)	**Del**	**chr6:57503219-96044796**	**LCA5, SMAP1**
12q12 (1)	**Del**	**chr12:40740730-40904094**	**LRRK2**
21p11.2 (0)	Del	chr21:10782384-10811535	

Table 3.7: Amplified and deleted exonic regions identified in the analysed tumours. Amplified / deleted regions were defined by GISTIC2 as having a log2 ratio above 0.1 or below -0.1 respectively. The region limits represent the boundaries of the entire amplified/ deleted genomic regions whilst the peak limits represent the region of maximum amplification /deletion. All regions with q values <0.1 highlighted in bold.

3.6.3 Mutated genes

Amplified +/- deleted genes were identified in 30/31 tumours analysed (97%). In these tumours the median number of amplified genes was 714 (range 12-942) and deleted genes was 525 (range 0-704) (Table 3.6)

There was no significant difference in the mean number of amplified/deleted genes in those tumours that and did not go onto recur during follow up (847 vs 1104 P=0.15).

A wide range of genes (n=1850) were located in the amplified/deleted regions identified by GISTIC2 including multiple cancer associated genes (Supplementary 3.19 and Table 3.7).

A moderate positive correlation was seen between the between the number amplified/deleted genes and the number of non-synonymous exonic SNVs/indels present in the analysed tumours (R^2 = 0.41 / P=0.0001) (Figure 3.20).

The only tumour with no amplified or deleted genes identified was T4.

Tumour	Histological subtype	Genes with gain in copy number	Genes with loss in copy number	Total genes with altered copy number
T25	Myxofibrosarcoma	942	702	1644
T18	Undifferentiated pleomorphic sarcoma	833	702	1535
T22	Leiomyosarcoma	857	588	1445
T9	Myxofibrosarcoma	743	702	1445
T15	Undifferentiated pleomorphic sarcoma	710	702	1412
T16	Undifferentiated pleomorphic sarcoma	697	701	1398
T26	Myxofibrosarcoma	846	526	1372
T19	Undifferentiated pleomorphic sarcoma	840	525	1365
T21	Soft tissue Ewing's sarcoma	823	525	1348
T32	Myxofibrosarcoma	642	702	1344
T30	Myxofibrosarcoma	743	525	1268
T3	Leiomyosarcoma	552	702	1254
T31	Myxofibrosarcoma	717	525	1242
T33	Leiomyosarcoma	823	412	1235
T8	Undifferentiated pleomorphic sarcoma	696	524	1220
T10	Undifferentiated pleomorphic sarcoma	511	702	1213
T13	Undifferentiated pleomorphic sarcoma	768	411	1179
T11	Dedifferentiated liposarcoma	776	291	1067
T14	Undifferentiated pleomorphic sarcoma	562	410	972
T29	Myxofibrosarcoma	809	115	924
T24	Undifferentiated pleomorphic sarcoma	297	589	886
T28	Leiomyosarcoma	774	0	774
T20	Leiomyosarcoma	742	2	744
T2	Undifferentiated pleomorphic sarcoma	390	291	681
T27	Dedifferentiated liposarcoma	214	411	625
T23	Undifferentiated pleomorphic sarcoma	123	293	416
T5	Leiomyosarcoma	304	1	305
T7	Haemangiosarcoma	12	115	127
T1	Extra-skeletal myxoid chondrosarcoma	87	0	87
T6	Synovial sarcoma	23	0	23

Table 3.8: Mutational load of amplified/deleted genes in analysed tumours. Tumours are shown in descending order ranked by the total number of genes they contained with an abnormal copy number. Amplified genes were defined by GISTIC 2.0 as being located within a regions with a log2 ratio above 0.1. Conversely deleted region / genes were defined as those with a log2 ratio below -0.1.

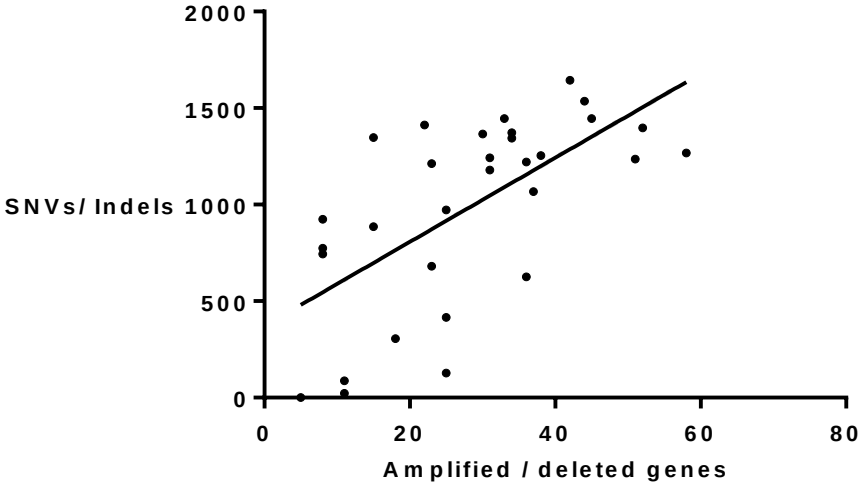


Figure 3.20: Scatter graph comparing the CNA and SNV/indel mutational loads of analysed tumours. The x-axis shows the number of genes with an abnormal copy number whilst the y-axis shows the number of exonic non-synonymous SNVs/indels identified. A moderate positive correlation was seen between the number of amplified/deleted genes and the number of non-synonymous exonic SNVs/indels identified ($R^2 = 0.41$ / P= 0.0001).

3.7 Discussion

In recent years increasing attempts to characterise the genomes of rare cancers has been performed to identifying those mutations central to tumourigenesis, identify novel therapeutic targets and to allow for better STS stratification. Despite an increasing availability of NGS and recent attempts to include rare tumour types in this process (i.e. the 100,000 Genome Project) sequencing data for STSs remains limited. To help address this have performed a detailed genomic characterisation of 31 primary STSs reporting on the somatic genetic mutations that most commonly alter gene expression – SNVs, indels, amplifications and deletions.

3.7.1 Experimental approach

The 31 analysed STSs were genotyped using Illumina NGS based on the technology's high throughput capabilities and sequencing accuracy (van Dijk et al. 2014; Quail et al. 2012). The raw data produced during our analysis was generally of excellent quality based on the low number of reads discarded during data filtration, the low number of sequencing errors reported by Novogene and Source Bioscience (<0.02%) and the high proportion of bases with phred quality scores of >Q20 (95%)/>Q30 (89%).

To identify somatic SNVs and indels we subsequently used the bioinformatic tools SomaticSniper and Strelka, based on their high sensitivities in samples with high and low purity (99.8%/96.0% respectively) (Larson et al. 2012; Saunders et al. 2012). To identify somatic CNAs the software package Sequenza was used. Although Sequenza has been shown to accurately estimate sample cellularity, ploidy and copy number down to a purity of 30% (Favero et al. 2015), comparisons between other read depth based CNA packages have revealed a poor consensuses when calling CNAs (Zare et al. 2017). Considering this we also validated our Sequenza data using a combination of Nanostring technology (previously shown to identify >97% of amplifications present (Suzuki et al. 2017; McIntyre et al. 2017; Ahn et al. 2016)) and ddPCR.

3.7.2 Sequenza analysis validation

To validate our Sequenza data 9 tumours were analysed using the nCounter® v2 Cancer Copy Number Assay. Overall this supported our Sequenza data with the CN profiles produced by the nCounter CN assay and Sequenza showing complete concordance in 5/9 STSs analysed. In the remaining 4 STSs a total of 5 genes were identified as amplified by the nCounter CN assay. The copy number of each of these genes was also elevated according to Sequenza, although the actual copy numbers predicted by the two analyses were different. This discordance is likely a reflection of the challenges posed by intra-tumoural heterogeneity, regional sequencing bias or the 'contamination' of STS tissue with healthy tissue (although the purity of all 4 tumour samples with discordant CN data was >60% - well above the 30% threshold under which Sequenza's sensitivity has been shown to drop.) To investigate the discrepancies in copy number above we also validated our Sequenza data using ddPCR. During this analysis *MDM2* levels were quantified in T11 using *RRPH1* as a reference gene (as recommended by Thermofisher). This revealed an *MDM2*:*RPPH1* CN ratio of 6:1, which was a reassuringly similar ratio to that estimated by Sequenza (5:1) supporting the validity of both experimental approaches.

3.7.3 Single Nucleotide Variants/ Indels

Errors during DNA replication and DNA damaging agents result in the accumulation of somatic SNVs/indels in a cell's genome over time (Katsonis et al. 2014; Lin et al. 2017). If these are located in the open reading frame and are able to overcome the protective effects of DNA degeneracy and the non-sense-mediated mRNA decay pathway transcription will be altered, potentially inducing an oncogenic effect.

3.7.3.1 Mutational load and frequency: The median number of non-synonymous exonic SNVs/indels detected per STS in our cohort was similar to the figure previously reported by TCGA. Although differences in patient and tumour characteristics between studies will invariably result in differences in mutational load, generally this is a reassuring measure of the stringency parameters adopted during our analysis.

The mean mutational load identified in our cohort is relatively low compared to those seen in many other malignancies. Considering that a cancer cell's mutational load increases with patient age and tumour cell renewal rate (Tomasetti et al. 2013), this may

be explained by the low turnover rate of mesenchymal cells (4%/year vs >15%/year in endothelial cells (Bergmann et al. 2015)) and the young age of many STS patients.

Overall deletions were twice as common as insertions in the analysed STSs (43 vs 21). This has previously been reported in multiple cancers including those involving bone, lung, skin, thyroid, breast, adrenal and salivary glands, the liver, central nervous system, genitourinary and gastrointestinal tracts and haematopoetic and lymphoid systems (Iengar 2012).

Given that a shift in 2 of the 3 codon positions results in a change in the open reading frame, frameshift indels should theoretically occur twice as often as in-frame indels in the cancer genome. Despite this, frameshift indels were 3 times as common as in-frame indels in our cohort. This higher than expected frameshift:in-frame ratio is the result of a high number of frameshift insertions in the analysed STSs (n=19, 89% of total insertions). This finding may be explained in two ways. Firstly, those frameshift insertions present may have had strong pro-proliferative effects, resulting in their selection during tumour evolution. Alternatively, those non-frameshift insertions present may have produced either no selective advantage, or particularly deleterious effects, resulting in their removal from their respective tumours over time.

The prevalence of different base substitutions varied significant in the analysed STSs from 5% (T>A) to 40% (C>T). The high number of C>T substitutions seen is consistent with other cancers (Iengar 2012) and reflects the spontaneous deamination of cytosine known to occur at methylated CpG sequences found in abundance in the cancer genome (Illingworth et al. 2010). Although there are twice as many possible transversions than transitions, 68% of the SNVs identified in the analysed STSs were transitions. This imbalance has also previously been reported in cancers (Rubin & Green 2009) and suggest either that 1) transitions are generated at an faster rate than transversions (by processes such as tautomeric shift and cytosine deamination) or 2) that transversions are repaired at a faster rate than transitions are (potentially due to the relatively strong protective effects that DNA degeneracy elicits on them).

3.7.3.2 Neoadjuvant treatment: Although the genotoxic effects of ionising radiation are used to induce cell death in cancer patients (Ahmad et al. 2012) we identified no relationship between neoadjuvant treatment and mutational load/profile in the STSs analysed. Assuming alterations were induced by radiotherapy this suggests that 1) any cells that survived radiotherapy only underwent a small number of divisions before surgery (and so only made a small contribution to the total DNA sequenced), 2) any SNV/indels induced by radiotherapy created no proliferative advantages and so were at too lower frequency in to be detected or 3) the radiation induced alterations were predominantly made up of double strand DNA breakages, not characterised during our analysis. An alternate explanation is that very high levels of cell death in those tumours which received neoadjuvant treatment resulted in complete degradation of any DNA present that contained radiation induced alterations prior to sequencing. This is supported by the significantly higher levels of macroscopic tumour necrosis seen in those tumours that received neoadjuvant radiotherapy compared to those that did not ($p=0.05$) (≥90% of tumour volume in some cases).

3.7.3.3 Genes: The 5 genes most commonly found to contain SNVs/indels in the analysed STSs were *TP53* (32%) *MUC2/4* (13%), *FRG1B* (13%) and *RB1* (13%). Many of the alterations identified in these genes have been reported on COSMIC, potentially indicating a cancer associated role (Supplementary 3.20/3.21).

For *TP53* and *RB1* a role is sarcomagenesis is also supported by the high incidence of SNVs/indels also reported in TCGA (present in 33% and 15% of reported STSs respectively).

TP53 is the most commonly mutated human tumour suppressor gene and contains somatic alterations in almost every type of cancer in between 5 and 50% of cases (Olivier et al. 2010). *TP53* plays a key role in 1) inhibiting cellular proliferation in the presence of DNA damage (Zilfou & Lowe 2009) and 2) disease progression in late stage malignancy (Adorno et al. 2009). This explains how somatic *TP53* inactivation promotes oncogenesis, as well as the high rates of cancer seen in individuals with inherited germline *TP53* mutations (Li Fraumeni syndrome (Correa 2016)).

RB1 is another of the most frequently mutated tumour suppressor genes in human cancer (Chinnam & Goodrich 2011). Functioning *RB1* is central to the regulation of the cell cycle. As a cell moves towards the G1/S checkpoint pRb (the protein encoded by *RB1*)

becomes phosphorylated activating cyclin-dependent kinases that allow progression into S phase. This explains the susceptibility to cancer seen following germline or somatic alterations in *RB1* (Goodrich 2006) and how damaging somatic *RB1* alterations may promote sarcomagenesis.

Inevitably some differences are seen in the SNV/indel characteristics of our cohort and those reported by Barretina et al (Barretina et al. 2010), Movva et al (Movva et al. 2015) and TCGA. Potential explanations for this include variation in sample sizes (as large as 31 vs 591), enrolled patient disease stages, experimental techniques and STS subtypes in each cohort. Three examples of the latter include 1) GISTs (11% of Barretina et al's cohort but excluded from ours) 2) dedifferentiated liposarcomas (22% of TCGA patients compared with 3% of ours - a reflection of the propensity of this subtype to involve the retroperitoneum) and 3) rhabdomyosarcomas (14% of Movva et al's cohort but none of ours - a result of our exclusion of patients aged <18). Recognising these potential explanations for the variance seen, the relatively high frequency of *FRG1B* and *MUC2/MUC4* SNV/indels in our cohort (13%) should still be recognised.

FRG1B (FSHD region gene 1 family, member B) expression has been detected in a variety of tissues, at its highest levels in the brain and gonads. *FRG1B* mutations have been identified in multiple malignancies including glioma, prostate, bladder/urothelial, lung, head and neck and thyroid cancer (Mészáros et al. 2016; Kim et al. 2013). Despite this, little is currently known of the in vivo function of *FRG1B*, necessitating further work before any explanations can be offered for potential oncogenic/tumour suppressive roles/mechanisms.

The Mucins are a family of proteins that line epithelial surfaces and are classified as either 'secreted' (MUC2) or 'transmembrane' (MUC4) (Kufe 2009). Both types of Mucins have been linked with cancer. Secreted mucins provide a tumour suppressor effect by acting as a protective barrier against epithelial cell inflammation (Velcich et al. 2002). In contrast, transmembrane mucins promote onocogenesis by relaying proliferative signals and disrupting the epithelial layer (Kufe 2009). Although these links with cancer should be noted, the high frequency of Mucin mutations seen in our cohort may in fact not reflect an oncogenic phenotype, but instead one of the structural properties of the Mucin genes themselves. These properties include their large size (MUC2 540300 Da and MUC4 231518Da) and their association with multiple tandem

repeat segments which are prone to both SNVs (Tian et al. 2008) and indels (Batzer & Deininger 2002)). Another potential explanation for the high frequency of MUC2/4 variants seen is that the genes' repetitive genomic regions may have led to an increased number of false positive SNVs being called during analysis due to sequencing and/or mapping errors. This final potential explanation highlights the importance of analysing our data with a second variant caller package in the future to further clarify the biological significance of our findings.

The third most commonly altered gene is STSs on TCGA is *ATRX (*containing SNVs/indels in 15% of the cases presented. This was mirrored in our cohort with 10% of cases containing an *ATRX* SNVs/indels. A clear association between *ATRX* loss and the activation of ALT has previously been reported (Liau et al. 2015). Considering this the high frequency of *ATRX* alterations seen in STSs can be explained by the high proportion of STSs that activate ALT to maintain their telomere length. Loss of *ATRX* expression is most commonly reported in the STS subtypes UPSs, pleomorphic liposarcomas and dedifferentiated liposarcomas where there is a significant association seen with a complex karyotype (P>0.001) (Koelsche et al. 2016; Liau et al. 2015). If it is assumed that the *ATRX* SNV/indels present resulted in loss of *ATRX* expression this is consistent with our cohort, in those cases of UPS and dedifferentiated liposarcoma with *ATRX* mutations all had complex karyotypes.

In breast and ovarian cancer *TP53* mutations have been linked with aggressive tumour behaviour (Langerød et al. 2007; Wang et al. 2004). Using animal models *ATRX* deficiency has also be shown to reduce survival in glioblastoma (Koschmann et al. 2016). Recognising the small size of our cohort, these prognostic relationships were not immediately apparent in our cohort, with fewer mutant *TP53* patients recurring (30% vs. 38%) and dying (2 vs. 3) during follow up than patients with *TP53* wildtype tumours, and none of the 3 tumours containing *ATRX* mutations recurring during follow up. However, it should be highlighted that p53 and *ATRX* down regulation by mechanisms other than the acquisition of somatic SNVs/indels has not been excluded by our analysis, which may involve epigenetic regulation of *TP53/ATRX* transcription or the overexpression of negative regulators of p53 (such as *MDM2*). *ATRX* loss has been shown to increase SNV mutational load, impair non-homologous end joining and increase tumour sensitivity to DNA-damaging agents in glioblastoma (Koschmann et al.

2016). Although those tumours in our cohort with *ATRX* mutations did not have particularly high SNV/indel mutational loads, the apparent sensitivity of *ATRX* void tumours to DNA-damaging agents should investigated for in for STSs.

In some cancers, mutations in certain genes and/or their under or over expression have been shown to correlate with particularly aggressive or indolent disease pathways. Although this has been investigated in STSs (Beck et al. 2010) only very few genomic characteristics have been linked with prognosis (Davicioni et al. 2006). The development of a concise panel of genetic markers predictive of clinical outcome would help clinicians to both counsel patients and make more informed treatment decisions. In our cohort no genes were more commonly mutated in those tumours that recurred compared to those that remained disease free, precluding the identification of any potential markers for such a STS panel. Our failure to identify any gene(s) with these characteristics may be a result of the small size of our cohort and/or participants' relatively short follow up periods. Furthermore, it should be recognised that we made no formal assessment of actual gene expression, meaning that gene up or down regulation by epigenetic mechanisms or promoter region abnormalities would not be recognised by our current analysis

3.7.3.4 Driver genes: A mutation that confers a selective growth advantage to its cell promoting the process of tumour development is known as a driver mutation (Vogelstein et al. 2013). The genes in which these driver mutations occur are known as driver genes. Only one potential driver gene was identified in our analysis based on its SNV/indel characteristics – Poly (ADP-Ribose) Polymerase Family Member 4 (*PARP4*). Two different *PARP4* missense SNVs were identified in the analysed STSs – Gln1059Arg (n=3) and Ile1039Thr (n=2). These were spread across three different tumours including 1 MFS, 1 leiomyosarcoma and 1 soft tissue Ewing's sarcoma. Although the incidence of *PARP4* mutations in our cohort was considerably higher than previously reported in STSs on TCGA (10% vs. 0.2%), a high number of germline *PARP4* mutations have been identified in other malignancies including breast and thyroid cancer (Ikeda et al. 2016). Furthermore the in vivo knock down of *PARP4* expression has been shown to increase cellular proliferation indicating a role as a tumour suppressor (Ikeda et al. 2016). *PARP4* contains a BRCA1 carboxy-terminal domain (BRCT) previously linked with control of the cell cycle. This suggests that *PARP4* may illicit its tumour suppressive role via the

process of DNA repair, although other functions of other members of the PARP family such as the regulation of chromatin and transcription may also contribute (Bork et al. 1997). Paradoxically inhibiting the protective DNA repair role of *PARP4* has been exploited to accelerate cancer cell death in other cancers (Morales et al. 2014). If the relatively high number of *PARP4* mutations identified in our cohort is confirmed to accurately reflect the loss of *PARP4* expression, these are approaches which may be suitable for use in selected STS patients.

3.7.3.5 Mutational signature: Different genotoxic processes generate different base substitutions in different sequence contexts. These characteristics are known as a tumour's 'mutational signatures' and through their analysis and comparison hypotheses can be made on a tumour's aetiology. A detailed analysis of 30 malignancies (not including STSs) has previously identified over 20 recurrent 'mutational signatures' (Alexandrov et al. 2013).

Three discrete mutational signatures were identified in our analysed STSs, which each had a significant similarity to 1 of 2 mutational signatures previously identified in cancer genomes (mutational signatures 1 and 5).

Mutational signature number 1 is the commonest mutational signature which is dominated by C>T mutations thought to result from the spontaneous deamination of methyl-cytosine seen at relatively high levels in the cancer genome. Signature number 1 has also been shown to correlate strongly with patient age, suggesting a significant proportion of the somatic mutations involved are acquired throughout a patient's lifetime. Mutational signature number 5 is rarer than number 1 although has been identified in multiple malignancies including B cell lymphoma, medulloblastoma, myeloma, kidney, lung and thyroid cancers. Signature number 5 is characterised by T>C mutations, which in contrast to signature number 1 do not correlate with age, suggesting their development at different rates in different individuals. This is a characteristic which may be explained by exposure to exogenous carcinogens, although these are yet to be identified.

3.7.3.6 Clonal evolution: Intra-tumoural heterogeneity secondary to clonal evolution has previously been identified in certain malignant tumours using multiregional sequencing (Abbosh et al. 2017). Several STSs in our cohort contained SNVs in cancer associated genes at different frequencies of read depth. One example was seen in T2

which contained single SNVs in *TP53* (allelic frequency 57%), *MYOC5* (27%) and *BRIP1* (20%).

The frequency of the *TP53* SNV could be explained in 2 ways – either the alteration is homozygous and present in 57% of the sampled cells, or heterozygous but subject to loss of the wildtype allele (LOH) in a proportion of the cells sampled. Regardless of which explanation is accurate, the variant's absence in some the sampled cells may also be explained in 2 ways - 1) contamination of the analysed tumour sample with healthy cells (such as white blood cells or stromal cells) or 2) the presence of multiple subclonal tumour cell populations in the analysed STS, not all of which contained the *TP53* variant. Similarly the *BRIP1*/ *MY05C* SNVs' frequencies may also either reflect the presence of normal cells within the analysed tumour sample or multiple subclonal tumour cell populations. If these subclonal populations were present in T2, the differences in the *BRIP1*, *M05C* and *TP53* SNVs' frequencies may also be explained in two ways: Firstly the *BRIP1* / *MY05C* mutations may have developed after the *TP53* mutation - an explanation supported by the previous identification of the *TP53* SNV in the early stages of malignancy (Giannakis et al. 2016) and *MYO5C/BRIP* mutations in more advanced stages of ovarian cancer (Pennington et al. 2014). Secondly, the *BRIP1* and *MY05C* SNVs may have provided less of a proliferative advantage than the *TP53* mutation, resulting in them being 'phased out' of T2 over time (again due to clonal evolution.) Despite the *TP53*, *MYO5C* and *BRIP* SNV characteristics outlined above, to reliably confirm the presence of intratumoural heterogeneity in the analysed STSs multi-regional sequencing would be required.

3.7.3.7 Gender comparison: No significant difference was identified in the SNV/indel mutational load of the male and female patients analysed (P=0.44). The small size of our cohort precluded any statistical comparison between the SNVs/indel mutational load of those genes located on the X chromosome in men and women. In fact only two of the 25 most commonly mutated genes identified in the cohort were found on the X chromosome - *ATRX* and *ALAS2*. *ATRX* was mutated in 2 male patients' tumours (1 frameshift deletion / 1 missense mutation) and one female patient's tumour (1 splice site SNV). *ALAS2* was mutated in one male patient's tumour (1 missense SNV) and 1 female patient's tumour (1 missense SNV).

3.7.4 Copy Number Alterations (CNAs)

CNAs play a key role in many malignancies, particularly those with complex genomes such as many STS subtypes (Guillou & Aurias 2010). Several amplified/deleted exonic segments were identified as significantly altered in the analysed STSs. Four of these amplifications and 2 deletions were identified in >50% of STSs analysed. All of these CNAs have also previously been identified in STSs by Barretina et al (Barretina et al. 2010) and TCGA (albeit at varied frequencies), and contain a variety of cancer associated genes.

3.7.4.1 Amplified genomic regions: The most common significantly amplified region identified in the analysed STSs was 7p22.1 (21/31 tumours, 68%). Amplification of 7p22.1 has previously been identified in 37.5% of prostate cancer patients (Hughes et al. 2006) which may reflect the action of several oncogenes located in the segment. These include *RAC1* (a cytoskeleton modulator with oncogenic effects on angiogenesis and metastasis (Bid et al. 2013)) and Platelet-derived growth factors alpha (*PDGFA*). Platelet-derived growth factors alpha is a member of the platelet-derived growth factor (PDGF) family which have transmembrane receptor protein tyrosine kinase activity. When PDGFA binds to its specific platelet-derived growth factor receptor (PDGFR) the PI3K-AKT-mTOR and RAS-MAPK signalling pathways are initiated which have a well-recognised association with tumour biology (Demoulin et al. 2012). This interaction has been shown to play a key role in many biological processes which falter in cancer including cellular proliferation, survival, differentiation, migration and angiogenesis (Heldin et al. 2013). Oncogenic mutations and overexpression of PDGFs and PDGFRs has previously been identified in multiple cancers where it has been associated with tumour invasion, drug resistance, lymphatic dissemination, metastatic disease and poor survival (Wehler et al. 2008; Manzat Saplacan et al. 2017). The activation of the kinase receptor PDGFR by PDGFA has also shown to induce proliferation in osteosarcoma cells (Farooqi & Siddik 2015), and through it over-expression offers itself as a potential diagnostic tool and/or therapeutic target.

The next two most abundant significantly amplified regions identified were 17p11.2 (20/31 STSs, 65%) and 17q25.3 (18/31 tumours, 58%). The 17p11.2 amplification identified was relatively small consisting of only 11 genes of which only one has

previously been associated with cancer - *TOP3A. TOP3A* codes for a type 1 DNA topoisomerase. These are a family of enzymes which control the topology of DNA by transiently inducing single strand DNA breakages - a process central to DNA replication, protein transcription and protein DNA interactions (Champoux 2001). The presence of certain *TOP3A* SNVs has been linked with an increased risk of breast cancer (Broberg et al. 2009) and *TOP3A* overexpression has also been shown to increase cell proliferation in malignant osteosarcoma (where 17p11.2 amplification has been identified in 50% of cases (Both et al. 2012; Both et al. 2017)). Although *TOP3A* containing complexes are key for ALT mediated telomere lengthening (Sobinoff et al. 2017), the same complexes have also been shown to reduce sister chromatid exchanges (a recognised source of LOH in cancer cells (Broberg et al. 2009)). This highlights the need for further work to further investigate the role of *TOP3A* in cancer, especially given the high frequency of *TOP3A* amplification seen in our cohort and elsewhere (11% in Barretina et al /TCGA datasets).

The 17q25.3 amplification identified was much larger (245 genes) and contained several oncogenes including *GRB2* (known to stimulate cellular proliferation, metastases and invasiveness (Giubellino et al. 2008)) and *ITGB4* (a transmembrane receptor with oncogenic effects on cellular invasion, progression and metastasis (Li et al. 2017)). A third gene of interest in 17q25.3 is *H3F3B.* Together with *H3F3A, H3F3B* encodes for H3.3 – a histone that maintains stability in transcriptionally active regions of the genome (Yuen & Knoepfler 2013). SNVs in *H3F3B* have been identified in several tumours including chondroblastomas(Behjati et al. 2013), glioblastoma (Lewis et al. 2013) and ovarian adenocarcinoma (Presneau et al. 2005). Although *H3F3B* amplification has previously only been identified in 4% of STSs, significant *H3F3B* overexpression has been reported in colorectal cancer where a potential role as a biomarker has been proposed (Ayoubi et al. 2017). Considering this, further work is needed to both determine if the higher rate of *H3F3B* amplification we identified is mirrored in larger STSs cohorts, and to confirm the oncogenic effects of *H3F3B* amplification. *H3F3B* hypomethylation has previously been *s*hown to downregulate tumour suppressors including p16Ink4A (Chan et al. 2013). If this hypomethylation results in *H3F3B* upregulation (based on the down-regulatory effect of methylation elsewhere in the genome (Moore et al. 2013)) this offers one potential mechanistic explanation for these oncogenic effects.

The last of the most commonly identified significant amplifications discussed here involved 1q21.3 (16/31 STSs, 52%). Amplification of 1q21.3 has been identified in 10-

30% of breast cancers, particularly in cases of recurrent disease (70%) (Goh et al. 2017). Although this propensity to occur in recurrent tumours was not seen in our cohort, several potential oncogenic effects of 1q21.3 amplification have been proposed. These include the overexpression of *MUC1* (previously shown to activate the PI3K/AKT/mTOR pathway (Nath & Mukherjee 2014; Kufe 2013)) and *TPM3* (previously associated with inhibition of cell to cell adhesion and increased cell migration/invasion (Jeanes et al. 2008)).

3.7.4.2 Deleted genomic regions: The 2 most common significant deletions identified in our cohort involved 1q42.2 and 6q14.1.

Loss of 1q42.2 was identified in 20/31 STSs (65%). This region contains *H3F3A* in which recurrent SNVs have been identified in mesenchymal tumours including osteosarcoma (particularly in patients aged >30) (Koelsche et al. 2017). As outlined above abnormal H3.3 incorporation into DNA leads to genomic instability and abnormalities with chromosomal segregation (Lin et al. 2013). This is supported by the instability and segregation errors seen in animal models following the loss of H3.3 (Bush et al. 2013; Lin et al. 2013), and provides an explanation for how *H3F3A* deletion may lead to oncogenesis. The presence of other tumour suppressors in 1q42.2 should also be noted. Cancer cells grow rapidly and as a consequence suffer from a lack of oxygen despite increased angiogenesis. This leads to the re-programming of the normal metabolic pathways, and it is increasingly evident that many genes involved in metabolic pathways play direct roles in tumour progression. One such gene is Fumarate hydratase (*FH*) which is another tumour suppressor located in 1q42.2 whose deletion has previously been linked the development of multiple tumour including leiomyomas and renal cell carcinoma (Vocke et al. 2017). Under normal circumstances *FH* encodes for the enzyme fumarase which is central to the Krebs cycle. Loss of *FH* or de-functioning mutations in *FH* result in a build-up of fumarate which in turn leads to an increased production of reactive oxygen species and a state of pseudohypoxia. This results in an increased dependence of the affected cell to glycolysis (Furuta et al. 2010), which is a necessity for any tumour cell to survive in their relatively hypoxic environments. This explains *FH*'s tumour suppressive function, and highlights how its deletion may play a key role in the process of sarcomagenesis.

Loss of 6q14.1 was identified in 19/31 STSs analysed (61%). A similar deletion involving 6q12-22 has previously been reported in 50% of the prostate cancers, with follow up analyses identifying several candidate tumour suppressors in this region (also located in 6q14.1) (Verhagen et al. 2002). These included *SMAP1* (a GTPase-activating protein with a role in DNA repair commonly mutated in DNA mismatch repair cancers (Sangar et al. 2014)) and *HMGN3* (a nucleosome bound protein known to modulate chromatin structure, regulate epigenetic modifications and play a role in differentiation (Furusawa & Cherukuri 2010; Barkess et al. 2012)). Interestingly despite these potential explanations for an oncogenic effect of 6q14.1 deletion, neither *SMAP1* nor *HMGN3* were consistently deleted in the TCGA cohort (1.1% and 0.8% of cases respectively).

3.7.4.3 *MDM2*: *MDM2* amplification is a characteristic commonly used to diagnose cases of dedifferentiated liposarcoma in a clinical setting. Considering this, it is pleasing that significant *MDM2* amplification was identified in both cases of dedifferentiated liposarcoma analysed in our cohort (T11 CN=20, T27 CN =20) providing further validation of our experimental approach.

3.7.5 Genome complexity

Like previous studies we categorised the analysed STSs as having simple or complex karyotypes. For this we defined a 'simple' karyotype as one containing <5 exonic segments with an abnormal copy number. In 29 of 31 STSs karyotype complexity was as expected based on their histological subtype. The 2 remaining cases included 1 myxofibrosarcoma (T4, predicted complex, witnessed simple) and 1 soft tissue Ewing's sarcoma (T21, predicted simple, witnessed complex). In T4 the discordance seen may reflect significant sample contamination with normal cells, which is further indicated by the T4's low SNV/indel burden and low cellularity predicted by Sequenza (0.29). The majority (27/31) of STSs had complex karyotypes. The elevated (>2.7) ploidy of 18 of these tumours confirms the presence of significant chromosomal abnormalities likely to include polyploidy and aneuploidy (Mosieniak & Sikora 2010). Aneuploidy occurs following the dysfunctional segregation of chromosomes during anaphase. This may occur for a variety of reasons including spindle defects, abnormal centrosome number or telomere attrition resulting in the breakage/fusion/bridge cycle or abnormal chromosome

fusions. For telomere attrition to trigger these responses a cell must have errors in their cell cycle checkpoints, which may explain link previously reported between *TP53/ATRX* mutations and aneuploidy (Baumann et al. 2010) (which was also evident in our cohort).

Several of the STSs analysed contained exonic regions with multiple segments with various different copy numbers. One potential explanation for this is the presence of chromothripsis - a process during which a chromosomal region is 'shattered' into multiple segments simultaneously (Stephens et al. 2011). Several causes for chromothripsis have been proposed including premature chromosome condensation, the fragmentation of dicentric chromosomes formed during telomeric crisis and ionising radiation (Maciejowski et al. 2015; Ernst et al. 2016; Stephens et al. 2011). Despite this differentiating between the standard accumulation of discrete alterations in the same chromosome over time and chromothripsis can be difficult. To address this issue certain hallmarks of chromothripsis have been outlined including the presence of multiple clustered rearrangements confined to one or a few chromosome(s), imperfectly repaired DNA fragments and CNAs typically oscillating between two or three states. Although our analysis was insufficient to identify all of these, the presence of multiple short exonic segments with abnormal copy numbers confined to a small number of chromosomes was identified.

As discussed previously many STSs activate the ALT to maintain their telomeres. The phenotypes of ALT have previously been associated with complex genomic characteristics (Lovejoy et al. 2012) suggesting that ALT may be instigated following cellular crisis (a period characterised by severe genomic instability and large chromosomal rearrangements). Although no formal assessment of TMMs was made during our analysis, the high level of genomic complexity seen in the cohort combined with relatively high number of *ATRX* mutations identified suggests a high frequency of ALT positive tumours. However to confirm this further work to identify the known phenotypes of ALT is required.

Although complex karyotypes have been linked with poor prognosis in leukaemia (Stölzel et al. 2016) our cohort size and the presence of many other prognostic factors prevent any statement on the prognostic impact of karyotype complexity in STSs based on our analysis alone.

3.7.6 Tumours with low mutational load

Two of the STSs analysed had particularly low mutational loads - T4 (myxofibrosarcoma) and T6 (synovial sarcoma). There are 4 potential explanations for this:

Firstly the phenotypic effects of the few exonic mutations found in those cancer associated genes in T4/T6 may have been significant enough to induce malignant transformation. Examples of these genes include BRD9 (Huang et al. 2014), PRB1 (Zhang et al. 2015) and FRG1B in T4, and ERAP2 (Gadalla et al. 2013), FERMT3 (Djaafri et al. 2014) and RASAL3 (Zenonos 2013) in T6, and although this scenario would indicate that certain STSs only contain a relatively small number of driver mutations, similar findings have been reported in other malignancies including medulloblastomas (0-2 driver mutations, (Vogelstein et al. 2013)). Secondly, it may be that key exonic driver mutations present in T4/T6 were missed during our analysis, potentially due to contamination of the sequenced tumour samples with normal cells, or too stringent QC thresholds being adopted during the bioinformatic analyses. Thirdly, our experimental approach (WES NGS) may have missed the oncogenic variants that drove tumour development in T4/T6. Potential examples of these variants would include intergenic, intronic or untranslated regions alterations, or more complex genomic abnormalities such as chromosomal translocations. Finally, it may be that epigenetic mechanisms drove the process of sarcomagenesis in T4/T6. Potential examples of such mechanisms include changes in DNA methylation, nucleosome location, histone modification or the post transcriptional effects of non-coding RNAs such as microRNAs (Sharma et al. 2009)

3.7.7 Actionable mutations

Several mutation identified in our cohort have previously been shown to predict treatment response. Examples included point mutations in *PTCH1* (linked with sensitivity to hedgehog inhibitors (Sharpe et al. 2015)) and *TP53* (predictive of response to many therapies including cisplatin, epirubicin, 5-fluorouracil and methotrexate (Hientz et al. 2017)), deletions of *BRCA2* (predictive of sensitivity to *PARP* inhibitors (Iyevleva & Imyanitov 2016)) and *STK11* (predictive of sensitivity to MAPK/p38 inhibitors (Grossi et al. 2015)) and amplifications of *CCNE1* (predictive of sensitivity to CDK2 inhibitors

(Etemadmoghadam et al. 2013)) and *MDM2* (predictive of MDM2 inhibitor sensitivity (Saiki et al. 2015)). This highlights how the detailed sequencing of STSs may facilitate a more personalised approach to treatment, although obviously significant cost analyses are required before such an approach is offered in a clinical setting.

3.7.8 Inter-tumoural heterogeneity

3.7.8.1 Mutational load: No difference was seen in the SNVs/indel mutational loads of different subtypes in our cohort. Furthermore no mutational signatures unique to individual subtypes were identified. This is in contrast to findings in other malignancies such as melanoma (Hayward et al. 2017) and argues against the presence of STS subtype specific aetiologies.

Large variation was seen between the SNV/indel loads of the individual tumours analysed. This variation was present both between tumours of the same, and different subtypes, with the most pronounced examples of the former seen in myxofibrosarcomas (which contained between 5-58 variants) and leiomyosarcomas (8-52). No correlation was seen between patient age and SNV/indel mutational load ($R^2 = 0.09$ / $P=0.11$).

Our cohort's size prevented any comparison between the commonly amplified/deleted regions in different subtypes. However, large variation was seen in the overall number of amplified/deleted genes present in individual tumours. This variation was again evident when tumours of both the same, and different subtypes were compared, with the most striking examples of the former seen in cases of myxofibrosarcoma (with individual tumours containing between 0 – 1644 amplified/altered genes) and leiomyosarcoma (305-1445).

Finally large variation was also seen in the size of the individual CNAs present in the STSs analysed, indicating that a variety of mechanisms led to these alterations. Examples of this variety are highlighted by the whole chromosome arm (8q) amplifications seen in T1 (likely the result of an error in segregation) and the relatively small 29kb deletion seen in T5 (potentially the result of non-allelic non homologous recombination).

3.7.8.2 Mutated genes: Substantial inter-tumoural SNV/indel heterogeneity was evident in the analysed cohort. This is highlighted by 1) the high number of altered genes in the cohort 2) the presence of just 2 recurrent SNVs in the entire cohort and 3) the relatively low proportion of tumours that contained mutations in even the most commonly altered genes (<13% with the exception of *TP53*). The high number of mutated genes in the analysed cohort suggests the involvement of multiple oncogenic signalling pathways including the MAPK/ERK (activated by *PDGFA/GRB2*) and PI3K pathways (activated by *ITGB4*). These provide targets for several molecular therapies, highlighting the potential benefits of performing detailed STS sequencing. However, the low number of recurrent SNVs/indels in our cohort suggests that offering treatment based on these characteristics alone is unlikely to be cost effective. Although many different genomic regions were amplified/deleted in the analysed STSs, many were present in multiple tumours. Considering this, developing therapies to target these recurrent CNAs may offer the best chance of providing a significant proportion of STS patients effective treatment. Furthermore, as a tumour's mutations will define many of its characteristics including its treatment sensitivities, these recurrent CNAs may provide the best way moving forwards to stratify STSs in a clinically useful way.

3.7.9 Summary

We have successfully characterised the SNVs, indel and CNA characteristics of a 31 STSs using Illumina WES. This analysis revealed a relatively low SNV/Indel burden although a high number of recurrent CNAs. This highlights the importance of somatic CNAs for sarcomagenesis, but also suggests that amplifications/deletions may play a useful role as targets for future STS therapeutics. As expected the majority of STSs analysed had complex karyotypes, likely to result from polyploidy, aneuploidy and potentially chromothripsis. A wide range of altered genes were identified during our analysis confirming 1) the presence of multiple oncogenic pathways and 2) significant heterogeneity in the genomic signatures of STS of both the same, and different histological subtypes. This heterogeneity suggests that stratifying STSs by their mutations (rather than their morphology) may provide the most accurate way to predict STS patients' outcome, and moving forward a more personalised approach to treatment. The recurrent nature of many of the amplifications/deletions identified during our analysis suggests that CNAs may serve as the most useful type of mutation for such stratification.

3.7.10 Future work

Our analysis is limited by its size. Moving forwards collaborations such as the 100,000 genome project will create larger datasets of STS NGS data. These should be used to perform a more meaningful analyses of different STS subtype's mutational signatures (and so their aetiology) and more specifically the amplified and/or deleted segments they contain - a comparison that is underpowered in our analysis due to our cohort's size. Larger cohorts will also allow those mutations with the strongest effect on patient outcome to be identified (assuming appropriate clinical information is collected) which should be encouraged. Ideally whole genome NGS data should be viewed as the gold standard for the provision of STS sequencing data in any future studies, although analysis of any co-existing epigenetic mechanisms will also be required for a complete understanding of sarcomagenesis.

A second limitation of our current analysis is the use of just one somatic variant caller programme and one copy number package to analyse the WES data presented. Although the tools selected have been shown to be highly sensitive, variation between different bioinformatic tools is an increasingly recognised issue (Zare et al. 2017). Although time constraints meant the use and comparison of data produced by multiple packages was not possible, this work is scheduled for the immediate future.

Another limitation of our work is the absence of any kind of pathway analysis. This work takes the distribution of somatic SNVs across a group of tumours and groups them based on their involvement in certain cellular mechanisms. This allows the importance of different oncogenic pathways and the potential efficacy of different treatments for these tumours to be predicted, and is work that we plan on performing imminently.

Our analysis highlighted several specific genetic characteristics that need further investigation. Certain amplified/deleted segments were seen consistently (but not unanimously) across the cohort, and work to confirm the key driver genes in these regions is needed. Our SNV analysis revealed a mutational profile similar to the previously reported mutational signature number 5, and relatively high levels of FRG1B and PARP4 alterations (of which the latter was identified as a potential driver). Analysis of the mutational profiles of cells exposed to known oncogenic genotoxins is required to determine the exact aetiology of signature number 5, and further in vivo work to

determine the function of FRG1B and the phenotypic effects of PARP4 suppression is also warranted.

Chapter 4: The circulating nucleic acid characteristics of metastatic STS patients

4.1 Background and aim

At present no circulating biomarkers of STS tissue exist. If successfully developed these may help clinicians to identify STS recurrence earlier than current means (allowing more patients to be offered curative treatment) and also allow disease burden to be monitored longitudinally in patients with known metastatic disease. In recent years cfDNA and ctDNA characteristics have been highlighted as an exciting potential source of prognostic, predictive and therapeutic cancer biomarkers. These markers would have many attractive characteristics although despite these, and the obvious need to develop new STS biomarkers, the circulating nucleic acid characteristics of only 2 STS patients have been reported to date. To start to investigate the potential of using circulating nucleic acids as biomarkers in STS patients, the work outlined in this chapter aimed to characterise the cfDNA and ctDNA in a group of metastatic soft tissue sarcoma patients.

4.2 Overview experimental approach

Two groups of participants were enrolled to provide samples for the analysis described in this chapter - one containing patients with a biopsy proven metastatic STS and another containing healthy adult controls. One whole blood sample was collected from every participant and processed to isolate plasma and lymphocytes from which cfDNA and BC DNA was extracted.

First total cfDNA levels were measured in every participants' plasma using quantitative PCR. Next an Iontorrent AmpliSeq™ panel custom designed for use in STS patients was used to sequence patients' cfDNA looking for evidence of ctDNA. The same panel was also used to sequence participants' BC DNA (to confirm the somatic origin of any circulating variants identified) and in those cases in which ctDNA was identified matched FFPE STS DNA (to confirm the tumoural origin of any circulating variants identified.) Next the cfDNA and ctDNA characteristics of every patient was correlated with their demographics and clinical characteristics (including tumour burden and disease state) to identify any interesting correlations.

4.3 Participant and tumour characteristics

4.3.1 Patients: Eleven patients with metastatic STSs were enrolled for analysis (5F:6M) with a mean age of 68.8 years (range 52.2-84.9) and a diverse range of STS subtypes (Table 4.1). The mean patient RECIST 1.1 score of the enrolled patients was 161 although individual scores varied widely highlighting the broad range of disease burdens seen in the cohort (range 25.2-341.9). Tumour Trojani grade was 3 in 6 cases, 2 in 2 cases, 1 in 1 case and unavailable in 2 cases. Disease state could be calculated from serial imaging in seven cases of which six had progressive disease and one had stable disease. None of the enrolled patients received any systemic oncological treatment or radiotherapy prior to sample collection.

4.3.2 Healthy controls*:* Six healthy adult controls (4F:2M) with a mean age of 45.4 years (range 33.0-59.6) were enrolled to provide control plasma samples during cfDNA analysis. These patients were significantly younger than the metastatic STS patient group ($P<0.002$) (Figure 4.1).

4.4 Cell free DNA levels

Cell free DNA levels were significantly higher in the metastatic STS patient group compared to the healthy control group (48.37ng/ml (range 9.0-106.0) vs 3.9 ng/ml (range 1.9-7.4), P=0.006, unpaired t-test) (Figure 4.2). No correlation was seen between cfDNA levels and participant age in either the patient or control groups (R^2=0.002 / P=0.89 and R^2=0.001 / P=0.85 respectively) (Figure 4.3). A weak positive correlation was seen between cfDNA levels and disease burden in the STS patient group (R^2=0.26 / P=0.11) (Figure 4.4). No obvious relationship was seen between patient cfDNA levels and either the Trojani grade of their tumour (Figure 4.5), their disease state or their STS subtype (Figure 4.6).

No significant difference was seen between cfDNA levels of those patients that died during follow up and those that survived (P=0.34) (Figure 4.7) and no correlation was seen between survival length and cfDNA levels in those patients that died (R^2=0.07 / P=0.61) (Figure 4.8).

Patient	Sex	Age	Histological subtype	STS Trojani grade	cfDNA levels (ng/ml plasma)	Disease burden (RECIST 1.1 Score)	Disease state	Evidence ctDNA	Overall survival (months)	Date FFPE STS tissue collection	Date cfDNA collection
1	F	74	Leiomyosarcoma	1	105.9	329	-	Y	5.6	16/01/2015	09/03/2015
2	F	79	Leiomyosarcoma	2	26.6	70	PD	Y	n/a	12/06/2014	09/03/2015
3	M	46	UPS	3	62.6	194	-	Y	11.5	20/01/2016	29/06/2015
4	M	46	Soft Tissue Chondrosarcoma	3	9	44	PD	Y	n/a	03/10/2013	14/09/2015
5	M	83	Epithelioid Angiosarcoma	3	89.4	98	-	Y	n/a	18/11/2015	30/11/2015
6	F	73	UPS	3	90.7	218	PD	N	2.8		
7	M	52	Synovial sarcoma	3	37.6	187	-	N	14.9		
8	F	72	Extra-skeletal myxoid chondrosarcoma	2	40.6	97	PD	N	n/a		
9	M	47	Spindle Cell Sarcoma	-	38.1	342	PD	N	5.9		
10	M	80	Liposarcoma	-	10.5	164	PD	N	4.9		
11	F	57	Synovial sarcoma	3	21.1	25	SD	N	n/a		

Table 4.1: Metastatic STS patients' demographics, clinical and cfDNA characteristics: RECIST 1.1 scores represent the sum diameter of all measureable lesions (mm). Disease burden and states (where serial comparable imaging was available) was calculated according to RECIST 1.1 criteria. **PD** – progressive disease, **SD** – stable disease. **UPS** - Undifferentiated pleomorphic sarcoma. Overall survival is shown in months for those patients that died during the period of analysis.

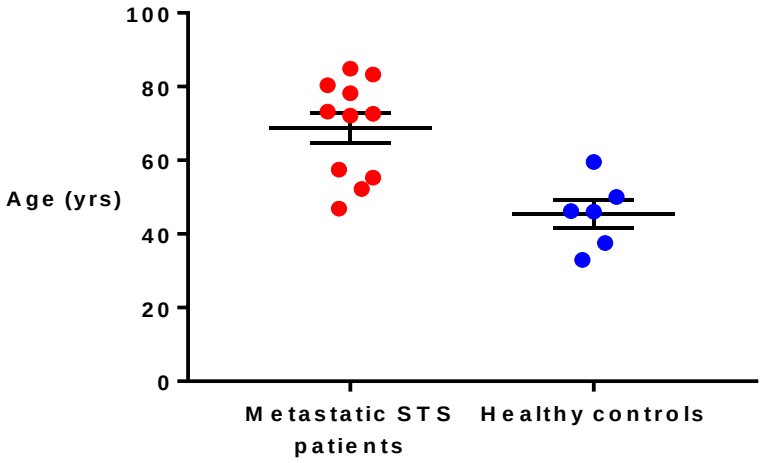


Figure 4.1: Comparison of metastatic STS patient and healthy control groups' ages. The mean and standard error of the mean are represented by horizontal bars. Mean STS patient age was 68.8 years (± 4.1 SEM, n=11) compared with 45.4 years (± 3.8 SEM, n=6) in the healthy controls. This difference was significant (P<0.002, unpaired t-test).

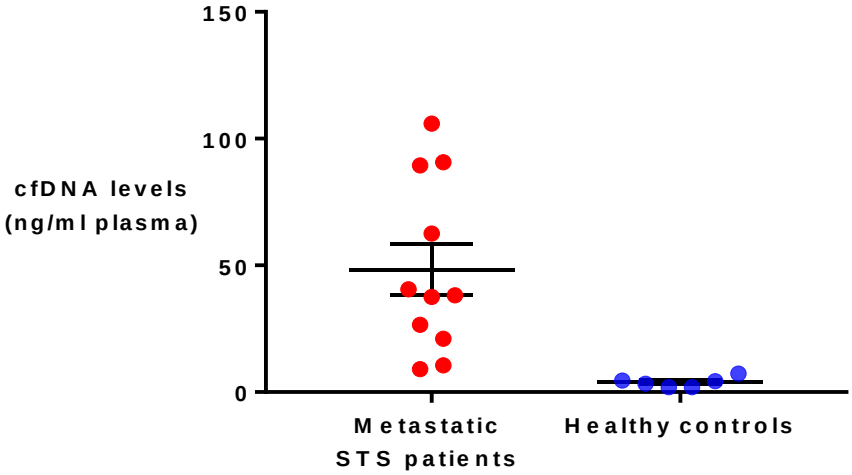
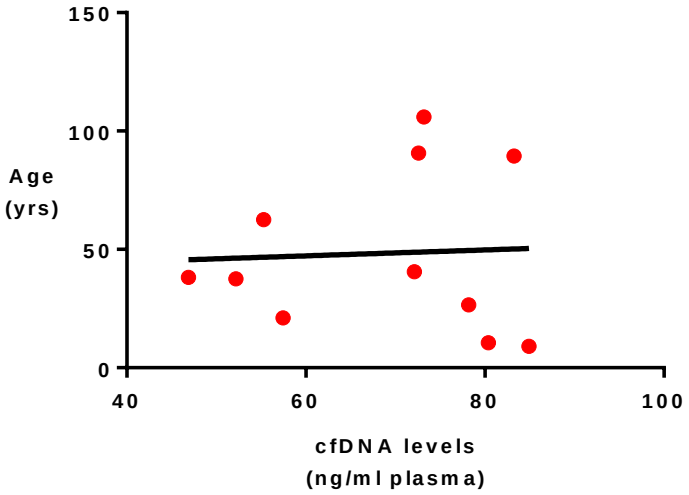
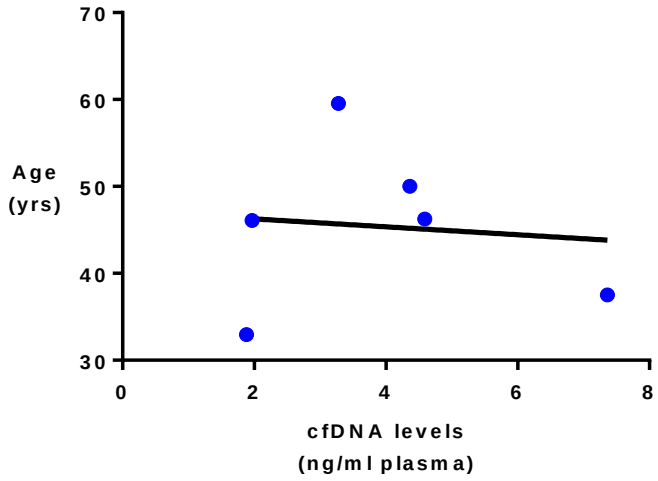


Figure 4.2: Cell free DNA levels in metastatic STS patients and healthy controls. The mean and standard error of the mean are represented by horizontal bars. Mean STS patient cfDNA concentration was 48.37ng/ml (± 10.21 SEM, n=11) compared with 3.9 ng/ml (± 0.8 SEM, n=6) in the healthy controls. This difference was significant (P=0.006, unpaired t-test). A retrospective analysis indicates that the power for this comparison is 99% ($\alpha = 0.05$).



a)



b)

Figure 4.3: Correlation between participants' age and cfDNA levels. Figures 4.3a and 4.3b represents the metastatic STS patient group and control groups respectively. No significant correlation is present between age and cfDNA levels in either group (R^2=0.002 / P=0.89 and 0.001 / P=0.85).

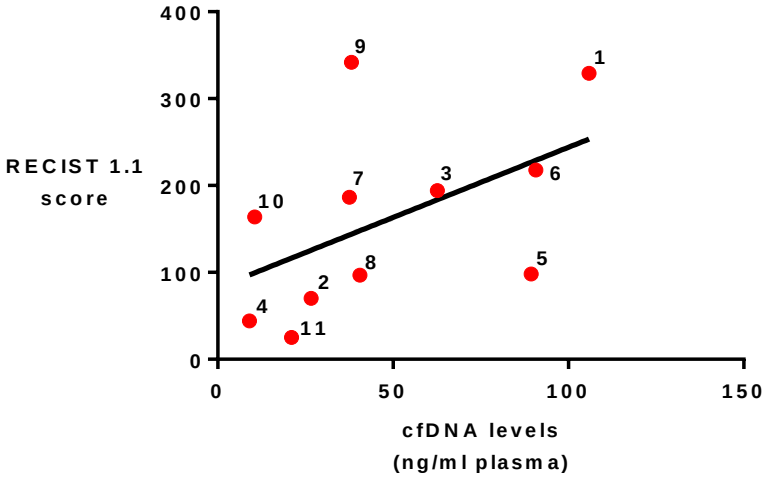


Figure 4.4: Correlation between disease burden and cfDNA levels in metastatic STS patients. Patients' participant numbers are shown on the graph. According to RECIST 1.1 criteria patients 2, 4, 6, 8, 9 and 10 had progressive disease and patient 11 had stable disease when cfDNA samples were collected. Disease state could not be calculated for patients 1, 3, 5 and 7. Overall a weak positive linear relationship can be seen between disease burden and cfDNA concentration in the group (R^2=0.26 / P=0.11).

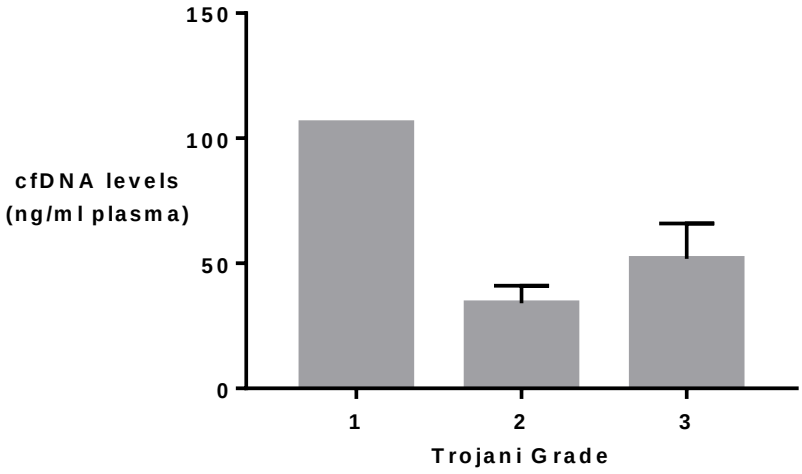


Figure 4.5: Metastatic STS patients' cfDNA levels categorised according to the Trojani tumour grade. The mean and standard error of the mean are represented by error bars. The mean cfDNA level was 34 (±7 SEM, n=2) for the grade 2 tumours and 51.8 (±14.2 SEM, n=6) for the grade 3 tumours. The single grade 1 STS analysed had a cfDNA level of 106ng/ml. No significant difference was seen in the cfDNA levels of the patients with different grades of STS (P=0.25, ANOVA).

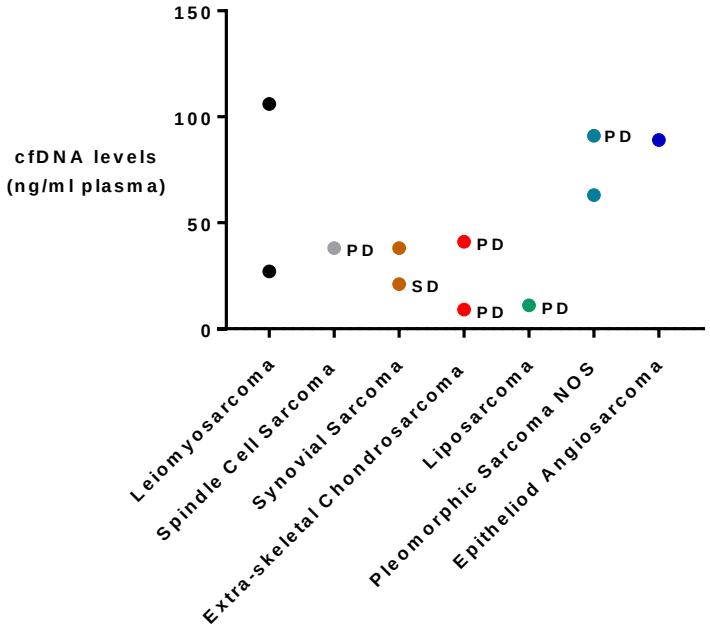


Figure 4.6: Metastatic STS patients' cfDNA levels categorised according to STS subtype and disease state. Disease state according to RECIST 1.1 criteria is recorded on the graph where available. **PD** = progressive disease, **SD** = stable disease.

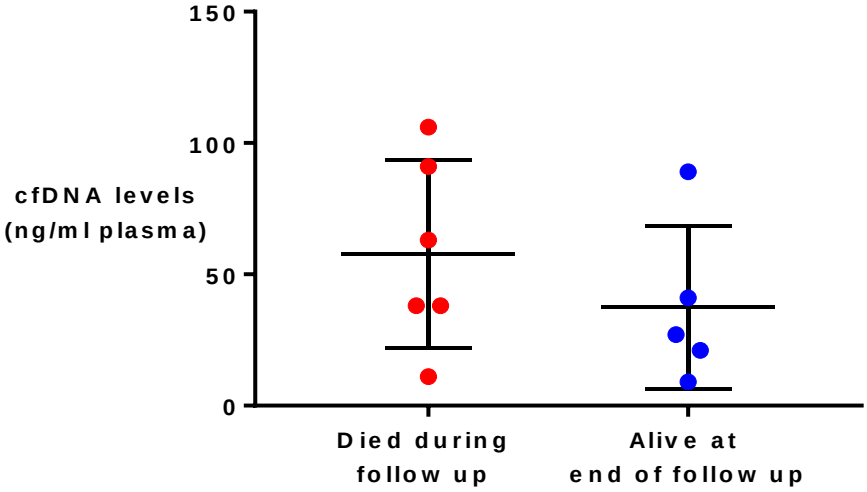


Figure 4.7: Comparison of cfDNA levels in metastatic STS patients that died and survived during follow up. The mean and standard error of the mean are represented by horizontal bars. Mean STS patient cfDNA concentration was 57.8ng/ml (± 14.6 SEM, n=6) compared with 37.4 ng/ml (± 13.9 SEM, n=5) in the healthy controls. This difference was not significant (P=0.34, unpaired t-test). A retrospective analysis indicates that the power for this comparison was just 17.2% ($\alpha = 0.05$).

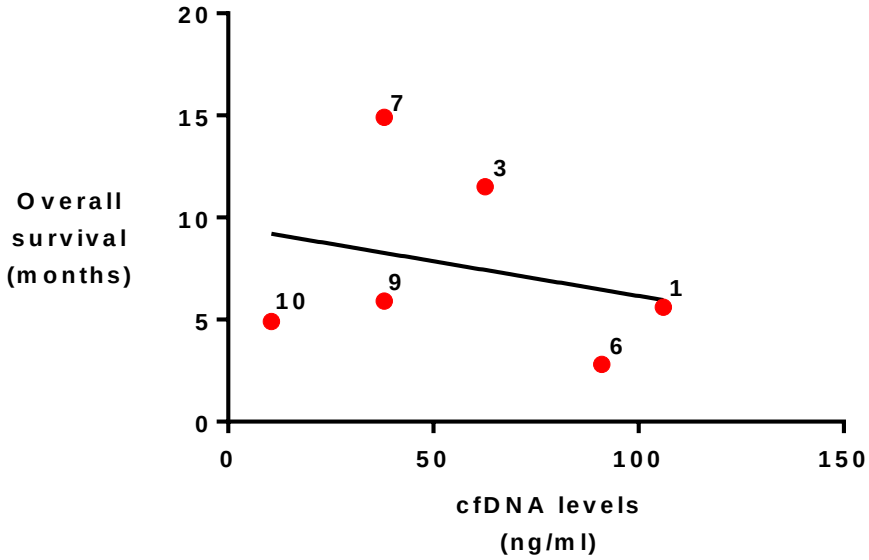


Figure 4.8: Correlation between cfDNA levels and length of survival. Data is shown for all the STS patients that died during follow up with participant numbers shown on the graph. No significant linear relationship can be seen between cfDNA levels and length of survival (R^2=0.07 / P=0.61).

4.5 Circulating tumour derived DNA characteristics

Evidence of ctDNA was identified in four of the 11 (36%) metastatic STS patients analysed (patients 2-5) (Table 4.2). The analyses of samples collected from one additional patient (participant 1) revealed two polymorphisms at different frequencies in the tumour DNA, BC DNA and cfDNA that can be interpreted in various ways, one of which supports the presence of ctDNA. Despite this a definitive statement about ctDNA in this patient is not possible, and for the purposes of the analyses below this patient was therefore assumed not to have ctDNA unless otherwise stated.

Between them the four STS patients with ctDNA had four different STS subtypes (leiomyosarcoma, undifferentiated pleomorphic sarcoma, epithelial angiosarcoma and extra skeletal myxoid chondrosarcoma) and a wide range of disease burdens (RECIST 1.1. score range 44-149) (Table 4.1). Disease state was determined for two of the four patients, both of whom had progressing disease at the time bloods were collected for analysis.

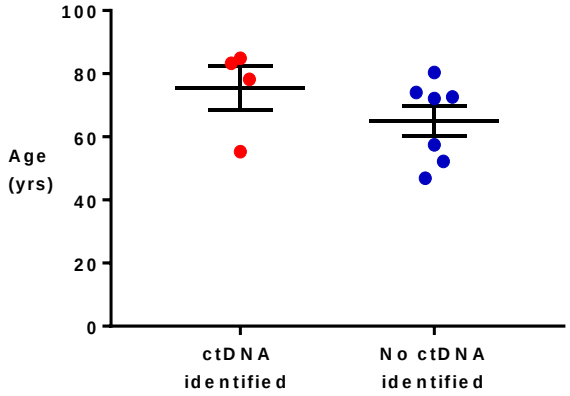
When those STS patients with and without ctDNA were compared no significant difference was seen between their ages, disease burdens or the grades of their primary STSs (P=0.24, 0.72 and 0.51 respectively) (Figure 4.9). Circulating cfDNA levels were however found to correlate more closely with disease burden in those patients with ctDNA compared to those without (Figure 4.10) (R^2 = 0.34 / P=0.42 excluding participant 1 / R^2 = 0.61 / P=0.12 including participant 1 vs. R^2 = 0.11 / P=0.52).

One of the four patients with ctDNA died during follow up with an overall survival of 11.5 months (Table 4.1). Five of the seven patients that showed no evidence of ctDNA died during follow up with a mean overall survival of 6.8 months.

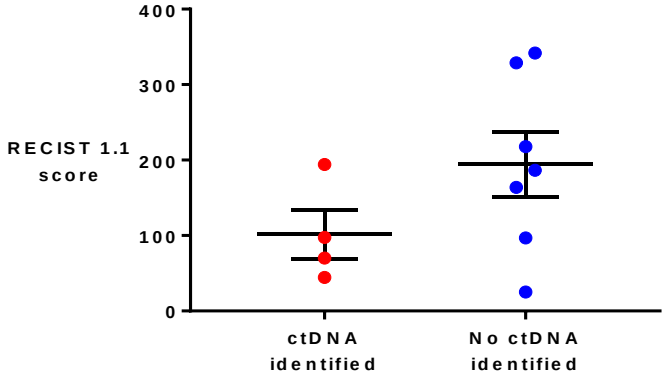
Detailed descriptions of the sequencing analyses of patients 1-5s' samples are described in detail in sections 4.5.1-4.5.5 and in table 4.2, with their clinical details shown in table 4.1.

								Patient cfDNA				Patient BC DNA			Patient Tumour DNA			Control cfDNA		
Pt	Chr	Position	Alleles (Ref /major: Var/minor)	Gene	Cosmic ID/ dbSNP	Predicted effect	Variant Caller identified	Replicate data	Depth	Variant reads	%	Depth	Variant reads	%	Depth	Variant reads	%	Depth	Variant reads	%
1	11	534242	(T/C)	HRAS	COSM249860 rs12628	p.H27H Silent	y	2527/3877 (65%) 1127/1654 (68%)	5531	3654	66	10660	6169	58	5395	5119	95	5202	13	0
	17	7576501	(C/T)	TP53	rs77697176	Intronic SNP	y	274/1482 (18%) 73/575 (13%)	2057	347	17	2957	908	31	716	16	2	941	0	0
2	17	7577094	(C/T)	TP53	COSM10704	p.R282W Arginine →tryptophan	y	257/5591 (5%) 191/4457 (4%)	10048	448	4	2479	3	0	1307	1034	79	6071	4	0
3	3	178921495	(C/T)	PIK3CA	COSM1666843 /4	p.S326F Serine →Phenylalanine	n	\	6926	36	1	1898	2	0	2532	17	1	2075	4	0
4	3	178927410	(A/G)	PIK3CA	COSM328028	p.I391M Isoleucine →Methionine	y	\	4008	110	3	2314	1	0	2139	1	0	3151	1	0
	17	7578210	(A/G)	TP53	COSM249885 rs1800372	p.R213R Silent	y	\	8114	293	4	2486	1	0	3438	3	0	4983	5	0
5	3	178916941	(G>A)	PIK3CA	COSM6145	p.E110K Glutamic acid→Lysine	n	\	23716	148	1	10463	7	0	5506	17	0	1271	2	0

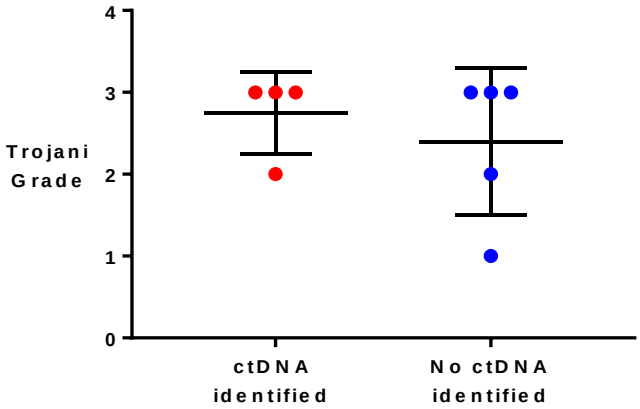
Table 4.2: Circulating variants identified in the metastatic STS patients with evidence of ctDNA. Data shown includes the sequencing analyses of patients' cfDNA, BC DNA and FFPE tumour tissue DNA. The analysis of the cfDNA of one healthy control is also shown as a measure of background noise**. SNP** - Single nucleotide polymorphism, **dbSNP** - Single nucleotide polymorphism database (https://www.ncbi.nlm.nih.gov/projects/SNP/), **COSMIC** - Catalogue of somatic mutations in cancer (http://cancer.sanger.ac.uk/cosmic), **Ref** – reference, **Var** – variant.



a)



b)



c)

Figure 4.9: Comparisons between the demographics of metastatic STS patients with and without evidence of ctDNA. Comparisons between the groups' ages, disease burdens (RECIST 1.1 score) and Trojani tumour grades are shown in figures a, b and c respectively. Horizontal bars represent mean values and standard error of the mean. No significant difference was seen between the ages, disease burden or tumour grades of the groups (P=0.24/0.18/0.51 respectively, unpaired t-test). A retrospective analysis indicates that the power for this comparison was just 11.2% ($\alpha = 0.05$).

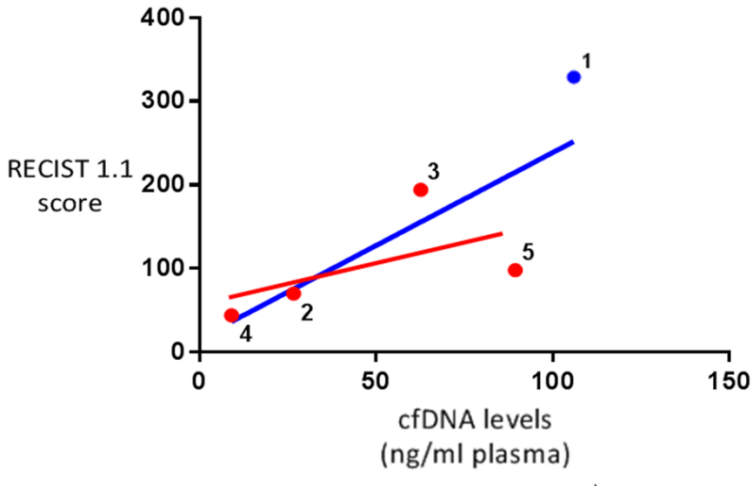


Figure 4.10: Correlation between cfDNA levels and disease burden in metastatic STS patients with evidence of ctDNA. The four patients with clear evidence of ctDNA detected are shown (red dots) and participant 1 (blue dot). If participant 1 is assumed to have ctDNA a strong positive linear relationship is seen between disease burden and cfDNA levels in these patients (blue line of best fit shown) (R^2=0.61/ P=0.12). If participant 1 is assumed not to have ctDNA and removed from this analysis a weaker positive correlation between cfDNA levels and disease burden is seen (red line of best fit shown) (R^2=0.34 / P=0.42).

4.5.1 Participant one:

Although the analysis of participant 1's STS DNA and cfDNA revealed no known cancer associated mutations, analysis of their BC DNA revealed two polymorphisms at high frequencies. One of these was an intronic SNP involving *TP53* (rs77697176, C/T) with a minor allele (T) read depth frequency of 31%. The other was a silent variant involving *HRAS* (rs12628_T/C_H27H) with a variant allele (C) frequency of 58%. The observed frequencies of the same *TP53* (rs77697176, T) and *HRAS* (rs12628, C) alleles in participant 1's STS DNA and cfDNA were 2% and 95%, and 17% and 66% respectively (Table 4.2). There are three potential explanations for the observed frequencies of these alleles at these positions: The first is the presence of somatic loss of heterozygosity (LOH) at both positions in the tumour, combined with significant sampling error in the BC samples and the presence of ctDNA. The second explanation for the observed frequencies is that the patient is mosaic for somatic LOH at both positions in certain cell lineages including the haematopoietic progenitor cell population and the lineage that gave rise to the STS. In this interpretation the polymorphisms' BC allele frequencies may be explained by the mosaic haematopoietic progenitor cell population's variable contribution to the circulating BC. Meanwhile the polymorphisms' tumour allele frequencies would follow its origin from a cell with LOH at both positions, whilst the polymorphisms' cfDNA allele frequencies may either reflect the presence of cfDNA shed from the mosaic BC population or the STS itself (ctDNA). The third interpretation is that participant 1 was born a homozygote C at both polymorphisms, but subsequently developed a *de novo* somatic C>T mutation in their haematopoietic progenitor lineage at both positions (a process termed clonal haematopoiesis). The resulting mosaic C>T mutations in the haematopoietic progenitor lineage would explain the polymorphisms' alleles' observed BC frequencies, whilst the variant alleles' cfDNA frequencies would follow the presence of cfDNA shed from participant 1 circulating BC population.

4.5.2 Participant two:

A well-recognised oncogenic *TP53* mutation (COSM10704_p.R282W) was identified in participant two's plasma at a variant allele frequency (VAF) of 4%. The same variant was absent from participant two's BC DNA but present at a frequency of 79% in their STS tissue confirming its tumoural origin.

4.5.3 Participant three:

A mutation in *PIK3CA* (COSM1666843/2_ p.S326F) was identified in participant three's plasma at a VAF of 1%. The variant was absent from the participant's BC DNA but present in their FFPE tumour tissue (albeit at a low frequency of 1%).

4.5.4 Participant four:

Two circulating mutations involving *PIK3CA* and *TP53* were detected in participant four's plasma at VAFs of 3% (COSM328028 _p.I391M) and 4% (COSM249885 _R213R) respectively. Neither variant was present in the participant's BC DNA, and interesting no somatic mutations (including COSM328028 _p.I391M or COSM249885 _R213R) were detected in their tumour tissue.

4.5.5 Participant five:

Analysis of participant five's plasma revealed a circulating *PIK3CA* mutation (COSM6145_ p.E110K) at a VAF of 1%. The variant was absent from the participant's BC DNA, but identified at low frequency (1%) in their tumour tissue.

4.6 Discussion

4.6.1 Participants

To begin our investigation into the circulating nucleic acids characteristics of STS patients the work outlined in this chapter aimed to characterise cfDNA and ctDNA in patients with confirmed metastatic disease. This decision was based on the hypothesis that circulating nucleic acid levels would be higher in STS patients with disseminated disease, based on cfDNA levels reported in other malignancies (Zaher et al. 2013). A total of 11 patients were enrolled. Although this constitutes a small cohort these individuals reflect the wider STS population well in terms of age (mean 68, range 46-86 years) gender (5F:6M) and tumour subtype, with a total of 7 different types of STS enrolled including cases of the relatively common undifferentiated pleomorphic sarcoma, liposarcoma, synovial sarcoma and leiomyosarcoma. To act as a comparison during our cfDNA level analysis a group of five healthy adults were also enrolled. The only significant difference between this control group and the STS patients was their age (P=0.002). Although such demographic differences should ideally be avoided when making comparisons, age has been shown to have no effect on cfDNA levels in healthy individuals (Tamkovich et al. 2016) making this difference likely inconsequential.

4.6.2 Experimental approach

A commercial kit was used to extract cfDNA. This minimised the risks of variation in yield between extractions and has been shown to provide high yields of good quality cfDNA (Sorber et al. 2017; Fong et al. 2009). A second kit was also used to extract FFPE tumour DNA, which again has been shown to isolate good quality DNA suitable for downstream sequencing (Darwanto et al. 2017) (in part due to its use of Uracil-N-Glycosilase to remove deaminated cytosine residues created during FFPE fixation (Do & Dobrovic 2012)).

To quantify cfDNA, BC DNA, FFPE tumour DNA and DNA libraries qPCR (cfDNA), a Qubit Fluorometer (BC/FFPE tumour DNA) and an Agilent Bioanalyzer (DNA libraries) were used. Testing showed that our in house qPCR assay showed good reproducibility (Figures 2.2-2.3) and both the Qubit Fluorometer and Agilent Bioanalyzer have also been shown to accurately quantify nucleic acids (including PCR products) in a reproducible way (Simbolo et al. 2013; Panaro et al. 2000).

To genotype our samples we performed targeted NGS (tNGS) using an Ion Torrent PGM™ sequencer (Life Technologies). This sequencing method has been validated against qPCR and other NGS techniques (Tuononen et al. 2013; Reiman et al. 2017) and is sensitive enough to confidently identify circulating mutations at frequencies down to 0.5-1%. Although readymade Ampliseq™ panels targeting common oncogenic mutations are available, to maximise our chances of detecting variants (and so ctDNA) we created our own custom designed panel covering the most commonly reported substitutions in STSs. Importantly using an Ampliseq tNGS assay allowed us to look for multiple SNVs in each patient's plasma simultaneously. This was particularly important as tumour tissue collected at the same time as each patient's plasma was unavailable for the metastatic patients analysed, meaning we had no way of knowing for sure the exact SNVs present in each patient's STSs at this point in time. Although the result of six of this panel's 57 amplicons were discarded due to poor sequencing depth, the remaining amplicons performed well with 98.7% of their reads aligning accurately to their target regions, resulting in a mean sequencing depth of greater than 2000 and an overall sequencing uniformity of over 90%.

4.6.3 Cell free DNA levels

The average cfDNA levels of the metastatic STS patients analysed was 48ng/ml. The absence of objective measures of disease burden in much of the published literature make comparisons between this value and cfDNA levels in other metastatic cancer patients difficult. However, it should be noted that levels much lower (non-small cell lung cancer 7.0 ng/ml (Li et al. 2016) / colorectal cancer 14.2ng/ml (Berger et al. 2017)) and higher (oesophageal squamous cell carcinoma 83 ng/ml (Totomochika et al. 2010) / prostate cancer 662.9ng/ml (Fawzy et al. 2016)) than this have been reported.

As predicted cfDNA levels were significantly higher in the STS patients than the healthy individuals, which is consistent with findings in other malignancies (Zaher et al. 2013). This difference may be the result of ctDNA undetected by our later analyses (see 4.6.4 for more detail) or a generally increased rate of cellular turnover in the STS patients. Regardless of the exact cause, the difference in cfDNA levels seen highlights cfDNA levels as a potential diagnostic marker for metastatic STS (a tool that may have particular use to detect metastatic recurrence following curative treatment).

Cell free DNA levels in the entire STS patient cohort positively correlated with patient disease burden, albeit only weakly (R^2=0.26 / P=0.11). If this is an accurate reflection of the greater STS population, then this correlation's weakness suggests that cfDNA may not response sensitively enough to changes in disease burden to allow metastatic STS patients to be monitored longitudinally for disease progression/response. In other malignancies where a lack of correlation is also seen between cfDNA levels and disease burden, cfDNA levels have subsequently been shown to provide a better reflection of more complex tumoural characteristics such as metabolic rate (Morbelli et al. 2017). We performed no such assessment of the STSs included in our analysis, but would advocate this work in the future. Recognising the small size of our cohort, it may be that the weak correlation seen in our data between cfDNA levels and disease burden strengthens as more STS patients (or subtype specific cohorts) are analysed. Of course this would re-highlight cfDNA levels as a potential marker of disease burden in late stage STS patients (which is currently assessed using expensive potentially iatrogenic imaging techniques (Pearce et al. 2012)).

Interestingly the strength of the correlation seen between disease burden and cfDNA levels rose when those STS patients with evidence of ctDNA were analysed in isolation. This suggests that a significant proportion of these patients' cfDNA is both tumour derived, and released into the circulation at a rate that correlates with disease burden. In contrast, the absence of a correlation between disease burden and cfDNA levels in those patients with no evidence of ctDNA suggests that the majority of their cfDNA is either not tumour derived, or released from tumoural tissue into the circulation (where our assays failed to detect it) at a rate that does not correlate with disease burden.

For a tumour to grow it must sustain an elevated metabolic rate (De Berardinis & Chandel 2016). Considering that disease state is defined by the growth of a patient's lesions and that cfDNA levels have been shown to correlate with metabolic activity in certain cancers (Morbelli et al. 2017), it could be hypothesised that STS cfDNA levels may also correlate with disease state. Although the only patient with stable disease in our cohort had the lowest cfDNA levels (25ng/ml plasma) our cohort's small size precludes any further statements on the relationship between cfDNA levels and disease state on the basis of our data alone, making this another relationship that needs further investigating.

High cfDNA levels have been linked with a poor overall survival in several cancer types (Yi et al. 2017; Ai et al. 2016). This was not a relationship identified during our analyses with no significant difference seen between the average cfDNA levels of those patients that died during follow up and those that survived (P=0.34), and no correlation seen between overall survival and cfDNA levels in those patients that died (R^2=0.07 / P=0.61). On the basis of our data this suggests that cfDNA levels are unlikely to play a role as a prognostic marker in late stage STS patients, although the analysis of larger cohorts required to conclude this.

In summary the cfDNA analysis outlined above has highlighted a potential diagnostic role for cfDNA levels in cases of advanced metastatic STS. This has been investigated further in chapter 5 by the analysing cfDNA levels longitudinally in STS patients undergoing attempted curative surgery, with the aim of determining how sensitively cfDNA levels alter (if at all) following the development of metastatic (or local) disease recurrence. Despite this exciting finding, it is overall disappointingly that the analysis failed to identify any significant correlation between cfDNA levels and disease burden, disease state, or patient outcome. Although this suggests cfDNA level may not be useful as a marker of disease progression or prognostic marker in cases of advanced STS, the small size and heterogeneous nature of our cohort necessitates larger, ideally subtype specific analyses. To try to address these absent relationships, we next sought to characterise ctDNA in the metastatic STS patients enrolled.

4.6.4 Circulating tumour derived DNA characteristics

Using our Ion AmpliSeq™ panel evidence of ctDNA was identified in four of the 11 (36%) metastatic STS patients analysed (participants 2-5). In the case of participant two one somatic, synonymous variant with characteristics consistent with an early clonal driver mutation (*TP53*_p.R282W_ COSM10704) was identified in the plasma and FFPE tumour tissue at frequencies of 4% and 79% respectively. The variant identified has previously been associated with multiple mesenchymal tumours (López-Guerrero et al. 2004; Tarpey et al. 2013) and provides clear evidence for the tumoural shedding of DNA. Analysis of participant three and five's plasma samples revealed two circulating mutations involving *PIK3CA* at a frequency of 1% (p.S326F_ COSM1666844 and p.E110K_ COSM6145 respectively). Both mutations have been previously associated with several malignancies (Beadling et al. 2011; Santarpia et al. 2012) and were also

identified in participant three and five's matched FFPE STS tissue, confirming the presence of ctDNA. Interestingly the tumoural frequencies of p.S326F and p.E110K were just 1%, which is lower than expected considering the effects of dilution. Assuming the cores which tumour DNA was extracted from contained a high proportion of tumoural tissue (suggested by H&E stained slides reviewed during the sampling process) these low frequencies are likely the result of significant intratumoural heterogeneity (ITH) resulting in those subclones containing p.S326F and p.E110K at high frequency to be missed during the FFPE sampling process. If true these cases therefore provide a good example of how ctDNA analysis can overcome some of the issues associated with ITH, which are often amplified by the analysis of solid tumour biopsies alone.

The analysis of patient four's plasma sample revealed two circulating variants at VAFs of 3% (COSM328028 *PIK3CA*_p.I391M) and 4% (COSM249885 *TP53*_R213R.) Both variants were absent from this patient's matched FFPE tumour tissue which may be explained in one of two ways. Firstly, significant ITH may again have resulted in the subclones containing the circulating variants from not being sampled during the process of FFPE tissue DNA extraction. Secondly, the circulating mutations may have developed at some stage after participant four's STS tissue was biopsied, but prior to the date of plasma collection (a period of time lasting 23 months.) In either scenario *PIK3CA*_p.I391M's recent association with several malignancies (Hucthagowder et al. 2012; Iacono et al. 2016) means that its presence in the circulation is highly likely to indicate the presence of ctDNA, whilst the presence of *TP53*_R213R highlights how somatic passenger mutations can also potentially act as useful tumour markers, even if they are synonymous and so unlikely to have any oncogenic effect themselves.

The four patients in which we identified ctDNA allow us to conclude that leiomyosarcoma, undifferentiated pleomorphic sarcoma, epithelioid sarcoma and soft tissue chondrosarcoma all shed DNA into the circulation. The wide range of these patients' disease burdens also allows us to determine that ctDNA can be detected in STS patients with both low and high volume disease. Despite this using the approach outlined we still failed to detect ctDNA in 7 of the patients analysed (participants 1, 6-11). There are 2 potential explanations for this. Most obviously our findings may reflect a true absence of ctDNA in these cases, that would mean that not all STS subtypes (or indeed every case of the same STS subtype) release DNA into the circulation. Alternatively ctDNA may have been present in these patients, but missed by our Ion AmpliSeq™

panel. This may be the result of a low burden of simple point mutations in these patients' tumours, and if accurate highlights how the absence of any point mutations consistently reported at high frequency in STSs is a potential barrier to the successful use of ctDNA to monitor STS patients. Regardless of the reason for our inability to detect ctDNA in 64% of the patients analysed, our failure to discriminate between them and healthy individuals obviously precludes us from volunteering the presence of ctDNA as a novel diagnostic STS biomarkers based on our data.

Our initial hypothesis was that patients with the highest tumour burden and most aggressive tumours were most likely to have ctDNA. Assuming that our Ion AmpliSeq™ panel had a 100% sensitivity for the detection of ctDNA, our data disagrees with this hypothesis by revealing no significant difference in the disease burden or tumour grade of those patients with and without ctDNA, and no positive correlation between ctDNA levels and RECIST 1.1 scores in those patients with ctDNA. .Again if it assumed that our panel was 100% sensitive, collectively this suggests that the tumoural shedding of DNA from STSs is neither proportional to disease burden, nor the other tumour characteristics that make up the Trojani grading system (tumour cell differentiation, necrosis and mitotic count). Although this in turn suggests that ctDNA levels are therefore unlikely to accurately reflect changes in STS disease burden, the small size and heterogeneous nature of our cohort necessitates larger (ideally subtype specific) analyses before conclusions can be drawn on the potential use of ctDNA to monitor advanced STS patients for disease progression.

In terms of ctDNA and patients' disease state and prognosis we hypothesised that ctDNA would most likely to be identified in patients with progressive disease, where it would be associated with a poor prognosis. This prediction was based on the high levels of cell turnover present in enlarging tumours, combined with the presence of ctDNA's poor prognostic significance in several other malignancies (Fan et al. 2017; Lee et al. 2018). Again this hypothesis was not borne out by our data which only revealed ctDNA in 2 of the 6 patients with progressive disease, and 2 or the 6 patients that died during follow up. Although the limitations of our cohort again require the analyses of larger cohorts to allow the ctDNA characteristics of individual STS subtypes to be comment on, our data provides no evidence that ctDNA levels may be useful as a marker of progressive disease or prognosis in cases of advanced STS.

Among the 7 patients without evidence of ctDNA, only participant 1 suggested some ambiguity. Two polymorphisms with frequencies of interest were identified in this case – one intronic single nuclear polymorphism involving *TP53* (major allele C /minor allele T) and one silent variant involving *HRAS* (reference base T/variant base C). In participant 1's BC DNA the frequency of the *TP53* polymorphism's minor allele (T) was 31% and the *HRAS* variant's variant allele's (C) frequency was 58%. The frequencies of the same *TP53* (T) and *HRAS* (C) alleles in participant 1's STS DNA and cfDNA were 2% and 95%, and 17% and 66% respectively (Table 4.2). Taken together the patterns of these polymorphisms' alleles' frequencies can be explained by various scenarios:

The first of these assumes that participant 1 is germline heterozygous at both polymorphisms, and that the observed allele frequencies in the tumour are the result of somatic loss of the T allele at both sites (loss of heterozygosity.) In this scenario the deviation seen in both polymorphisms' allele frequencies in the BC DNA away from their expected 50% may be explained by sampling error, whilst the deviation seen in their cfDNA frequencies (away from their expected 50% towards their tumoural frequencies) may be explained by the presence of ctDNA.

A second explanation for the polymorphisms' alleles' frequencies is also based on the assumption that participant 1 is germline heterozygous at both positions, but also that they are mosaic for LOH in certain cell lineages including the haematopoietic progenitor cell population and the cell lineage that gave rise to the STS. In this interpretation the deviation seen in the polymorphisms' BC allele frequencies away from their expected 50% may be explained by a variable contribution to the circulating BC from the mosaic haematopoietic progenitor cell population, whist the presence of single alleles at each locus within the tumour could be explained by its origin from a cell with LOH at both loci. Finally the deviation seen in the alleles' cfDNA frequencies away from their expected 50% may be explained in 2 ways – either the presence of cfDNA shed from the mosaic haematopoietic BC population, or DNA shed directly from the tumour (ctDNA).

A third explanation for the findings in participant 1 assumes they were born a germline homozygote C at both polymorphisms, and again shows mosaicism in the haematopoietic progenitor lineage but this time for two independently early somatic C>T mutations acquired by clonal haematopoiesis (a process previously reported in over 20% of STSs (Coombs et al. 2017)). In this scenario the mosaic C>T mutations would explain the polymorphisms' alleles' observed BC frequencies, whilst the deviation seen in the

mutant alleles' cfDNA frequencies away from those seen in the tumour would result from the contribution of participant 1's BC DNA to their cfDNA pool. This interpretation subsequently does not involve ctDNA, and so precludes any conclusion on the overall ctDNA characteristics of participant 1.

4.6.5 Summary

In summary, using tNGS we have successfully characterised ctDNA in a cohort of metastatic STS patients. Although the data emerging from this analysis builds significantly on previously published work, several unanswered questions remain. The absence of any evidence of ctDNA in 64% of the patients analysed either suggests that 1) the tumoural shedding of DNA is not a characteristic of all STSs, or 2) the experimental approach we adopted is not specific enough to reliably detect ctDNA in STS patients.

4.6.6 Future work

Further studies are needed to determine which of these explanations is accurate. Regardless of their outcome, our data suggests that using ctDNA to monitor even late stage STS patients will prove challenging. Our sample size was too small and heterogeneous to allow us to comment on the relationships between patients' ctDNA characteristics and their tumours' characteristics, prognoses, or indeed the uniformity with which different STS subtypes release ctDNA. To address this, future studies performed should be large enough to allow these relationships to be investigated, and despite our inability to identify a correlation between ctDNA levels and disease burden, these should also ideally be performed longitudinally to allow any dynamic changes in ctDNA to be identified. Another logical step following on from the work outlined in this chapter is the characterisation of circulating nucleic acids in a group of non-metastatic STS patients. This is work which we decided to conduct using a combination of experimental approaches to investigate if adopting a truly personalised approach using patient/mutation specific ddPCR assays would allow us to identify ctDNA in a higher proportion of the STS patients. The results of our follow up work adopting this experimental approach is outlined in chapter 5.

Chapter 5: The circulating nucleic acid characteristics of non-metastatic STS patients

5.1 Background and aim

In recent years a drive to increase STS awareness has meant that the majority of STS patients now present to sarcoma treatment centres with potentially curable disease. In the UK following treatment these patients currently enter a standardised follow up regime consisting of serial clinical examinations and radiological investigations. Unfortunately despite this surveillance programme cases of STS recurrence are often too extensive when diagnosed to facilitate limb sparing, or even curative treatment.

Following our work investigating the circulating nucleic acid characteristics of metastatic STS patients we moved our attention to a cohort of non-metastatic patients. We analysed this group using a combination of experimental techniques in a longitudinal way with the aim of determining if a) adopting a more 'patient specific' approach would allow us to identify ctDNA in a higher proportion of STS patients and b) circulating nucleic acids are a potentially valuable source of prognostic or diagnostic markers of STS disease recurrence.

5.2 Overview experiment approach

For the work outlined in this chapter a group of non-metastatic STS patients undergoing attempted curative surgery provided a sample of fresh frozen STS tissue and multiple plasma samples collected throughout their follow up (Figure 5.1). To measure cfDNA levels absolute qPCR of GAPDH was used to quantify cfDNA in these plasma samples. Next to look for ctDNA patients' plasma and tumour samples were analysed using Taqman ddPCR assays or a second AmpliSeq™ SNV panel custom designed to target the three most commonly mutated genes in STS (*TP53, RB1* and *ATRX*). Patients' cfDNA levels and ctDNA profiles were then correlated with their clinical characteristics and outcomes to look for any potential use as diagnostic or prognostic markers.

5.3 Clinical characteristic of non-metastatic STS patients

5.3.1 Patient demographics and tumour characteristics:

Plasma samples were collected from 26 patients with non-metastatic STSs (13M:13F) with a mean age of 63.3 (range 22-87). These patients had a variety of STS subtypes of greatly varied sizes (mean resected volume 498cm^3 (range 4.4-3289) and a mean Trojani grade of 2.4 (Table 5.1).

Intra-operative plasma samples were collected from 22 of the 26 patients (12M:10F, mean age 63.8, range 27.2-87.2). Post-operative plasma samples were collected from every patient, although 2 of these were lost to long term follow up precluding any comments on recurrence.

5.3.2 Patient management:

Sixteen patients had their STSs removed with wide surgical margins. Ten of these patients received perioperative radiotherapy (five neoadjuvant / four adjuvant/ one both) which was combined with neoadjuvant chemotherapy in two cases. The remaining six patients received no (neo)adjuvant treatment (Table 5.1).

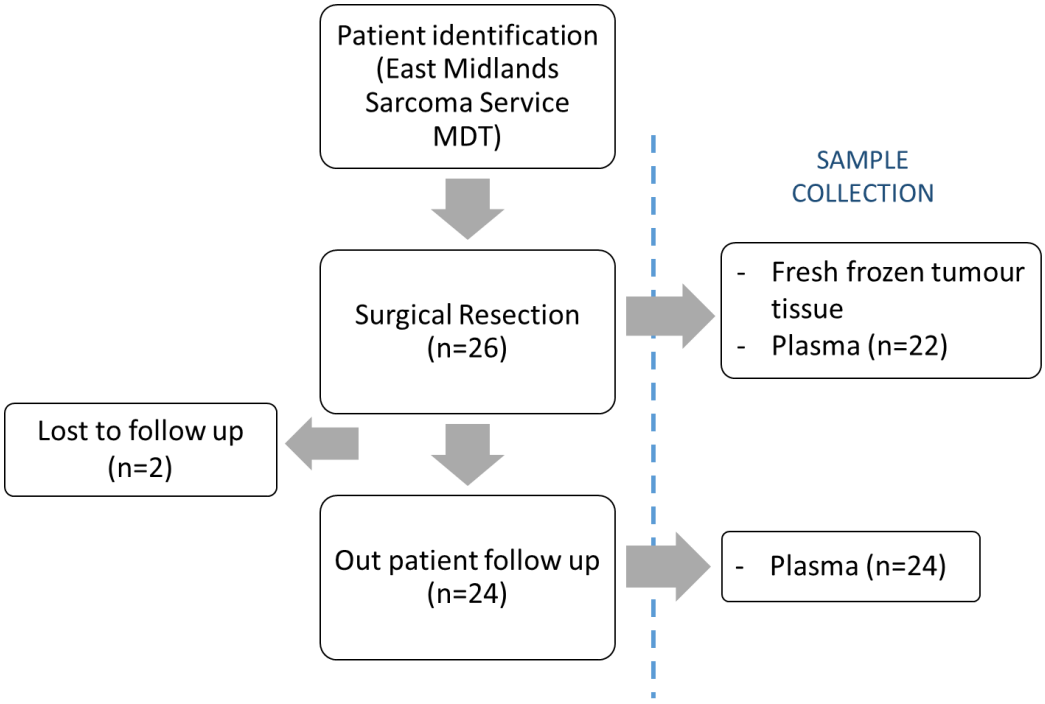
Five patients initially underwent an unplanned marginal (R1) resection. This was accepted in one case whom received adjuvant radiotherapy. The four remaining patients underwent further surgery to successfully achieve wide histological margins. Only one of these patients received (neoadjuvant) radiotherapy.

Five patients underwent a planned marginal resection due to the involvement of key neurovascular structures. All of these patients received radiotherapy (four neoadjuvant / one adjuvant). No intra-lesional resections were performed.

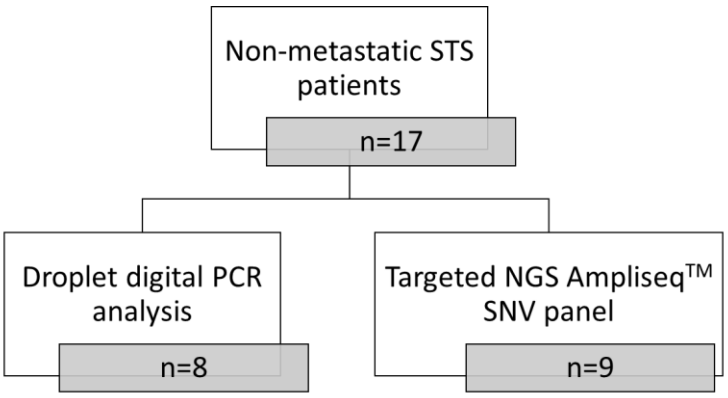
5.3.3 Patient outcome: Two patients were lost to follow up during our analysis. Eight of the remaining 24 patients (33%) developed metastatic STS recurrence during follow up at a mean of 34.7 weeks post-operatively (range 5.7-80.7) (Table 5.2/Figure 5.2). One case of local recurrence was detected 30 week post-operatively in a patient that underwent 2 surgeries to secure wide surgical margins and received no (neo)adjuvant therapy (patient 34). The mean length of follow up for those patients that remained disease free was 73 weeks (range 36-146). The overall 12 month disease-free survival in the cohort was 70.1% (Figure 5.2).

No significant difference was seen in the Trojani tumour grades of those patients that did and did not suffer disease recurrence, although those patients that recurred had significantly larger tumours ($1063cm^3$ vs $249cm^3$, P= 0.04) (Figure 5.3).

5.3.4 Survival/oncological outcome: At the cessation of follow up 7 patients were alive with disease (AWD), 15 had no evidence of disease (NED), none had died from other causes (DOC), 2 patients had died from their disease (DOD) and 2 were lost to follow up (Table 5.2). The two patients that died of metastatic disease died 23 and 44 weeks following completion of their surgical management.



a)



b)

Figure 5.1: Flow diagram of non-metastatic patient sample analyses. 5.1a shows the sample collection protocol used with the number of patients that provided intra- and post-operative plasma samples shown. Circulating tumour derived DNA was profiled in follow up plasma samples collected from 17 of the 26 patients that underwent surgery. 5.1b shows how these patients' samples were analysed. In those patients whose STS tissue was sequenced using illumina WES as outlined in detail in chapter 3 ctDNA was analysed using ddPCR. The remaining patients ctDNA was analysed using tNGS and a second custom designed Ion AmpliSeq™ SNV panel ('Sarcoma V2').

Patient	Age (years) / Gender	STS Subtype (+Trojani tumour Grade)	Tumour volume (cm^3)	Surgical resection margins (Marginal / Wide)	Radiotherapy / Chemotherapy (Nil /Neo/Adj)
1	76.1 / M	Myxofibrosarcoma (3)	663	Marginal (planned)	Radiotherapy (Neo)
3	63.1 / M	Extraskeletal Myxoid Chondrosarcoma (unknown)	588	Marginal (planned)	Radiotherapy (Adj)
6	55.3 / M	Undifferentiated Pleomorphic Sarcoma (2)	8.3	1^{st}:Marginal (unplanned) 2^{nd}:Wide	Radiotherapy (Neo)
9	62.3 / F	Leiomyosarcoma (3)	Unknown	Marginal (planned)	Radiotherapy (Neo)
10	59.7 / F	Synovial Sarcoma (2)	9	Wide	Nil
17	27.6 / M	Myxofibrosarcoma (1)	539	Wide	Nil
18	80.0 / F	Haemangiosarcoma (unknown)	Unknown	Wide	Nil
21	76.5 / F	Myxoid Liposarcoma (2)	198	Wide	Radiotherapy (Neo)
22	65.4 / F	Undifferentiated Pleomorphic Sarcoma (3)	364	Wide	Chemotherapy (Neo)+Radiotherapy (Adj)
23	53.2 / M	Undifferentiated Pleomorphic Sarcoma (2)	117	Wide	Radiotherapy (Neo)
24	68.9 / M	Myxofibrosarcoma (2)	144	Marginal (unplanned)	Radiotherapy (Adj)
25	36.7 / F	Undifferentiated Pleomorphic Sarcoma (3)	630	Wide	Nil
26	62.8 / M	Dedifferentiated Liposarcoma (2)	759	Wide	Radiotherapy (Neo)
27	67.0 / F	Undifferentiated Pleomorphic Sarcoma (2)	129	Wide	Radiotherapy (Neo)
28	70.6 / F	Myxofibrosarcoma (2)	113	1^{st}:Marginal (unplanned) 2^{nd}:Wide	Nil
29	74.0 / M	Leiomyosarcoma (3)	525	Marginal (planned)	Radiotherapy (Neo)
30	22.2 / M	Soft Tissue Ewing's Sarcoma (3)	151	Wide	Chemotherapy (Neo)+Radiotherapy (Adj)
31	45.8 / M	Undifferentiated Pleomorphic Sarcoma (3)	2947	Marginal (planned)	Radiotherapy (Neo)
32	64.0 / F	Myxofibrosarcoma (3)	4.4	1^{st} /2^{nd}:Marginal (unplanned) 3^{rd}:Wide	Nil
33	79.7 /M	Leiomyosarcoma (3)	3289	Wide	Radiotherapy (Neo/Adj)
34	69.0 /M	Undifferentiated pleomorphic sarcoma (2)	27	1^{st}:Marginal (unplanned) 2^{nd}:Wide	Nil
35	87.2 / F	Myxofibrosarcoma (3)	38	Wide	Nil
36	74.2 / M	Dedifferentiated liposarcoma (2)	576	Wide	Radiotherapy (Neo)
37	74.4 / F	Myxofibrosarcoma (2)	9	Wide	Nil
38	48.7 / F	Leiomyosarcoma (2)	61	Wide	Radiotherapy (Adj)
41	81.2 / F	Myxofibrosarcoma (3)	70	Wide	Radiotherapy (Adj)

Table 5.1: Clinical characteristics and management of non-metastatic STS patients. Adj-Adjuvant-; Neo-Neoadjuvant. M-Male; F-Female

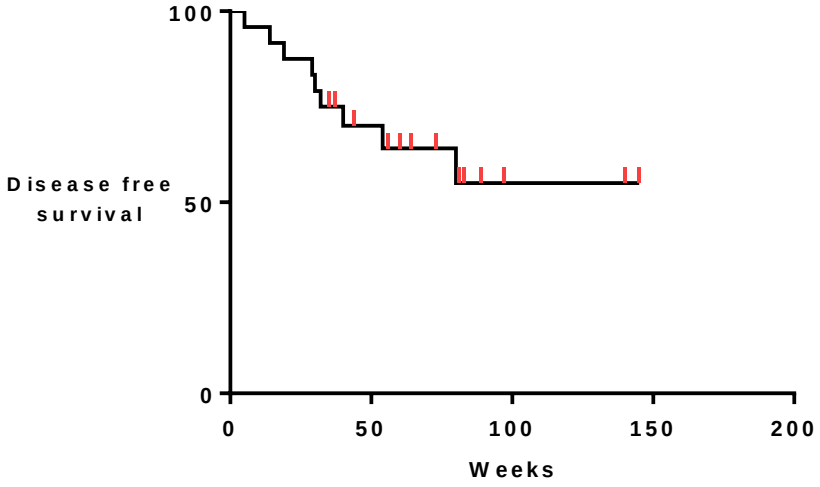
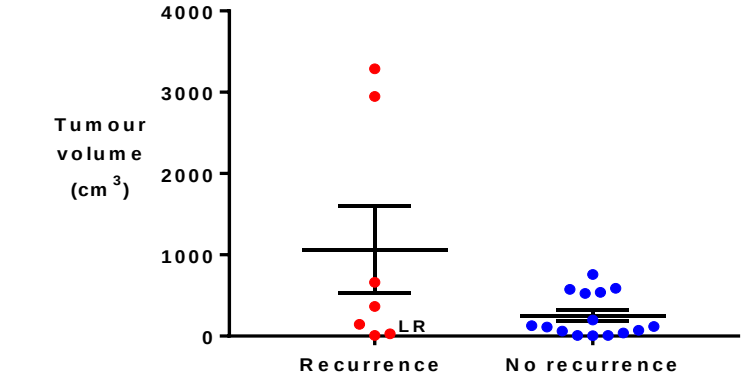
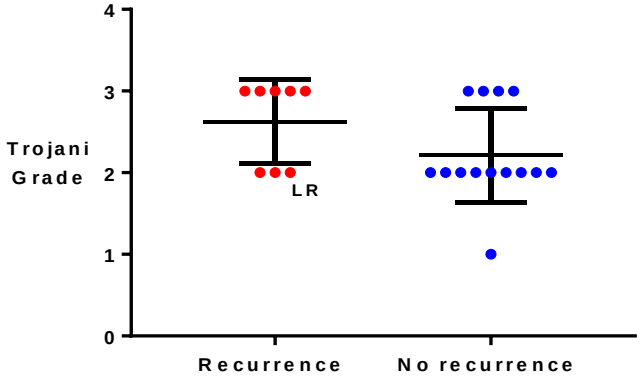


Figure 5.2: Kaplan-Meier curve of non-metastatic STS patients analysed. Data is shown for all patients not lost to follow up (n=24). Patients each had a minimum of 6 months clinical follow up. Censored points are represented by red ticks.



a)



b)

Figure 5.3: Tumour characteristics of recurrent and non-recurrent non-metastatic STS patients. Data for all patients not lost to follow up is shown where available. The mean and standard error of the mean are represented by horizontal bars. Data from the sole patient which suffered local recurrence is highlighted (**LR**).

5.3a compares resected tumour size (volume) between groups. The mean volume in patients that recurred was 1063cm^3 ± 539 (SEM) (n=7) compared with 249cm^3 ±68.4 (SEM) (n=15) in those patients that remained disease free during follow up. This difference was significant (P=0.04, unpaired t-test). Data on tumour volume was unavailable for 2 patients that suffered metastatic recurrence. A retrospective analysis indicates that the power for this comparison was 32.3% ($\alpha = 0.05$).

5.3b compares Trojani tumour grade between groups. The mean grade in those patients that recurred was 2.6 ±0.2 (SEM) (n=8) compared with 2.2 ±0.15 (SEM) (n=14) in those patients that did not recur. This difference was not significant (P=0.11, unpaired t-test). Tumour grade was unavailable for 1 patient that suffered metastatic recurrence and 1 patient that remained disease free. A retrospective analysis indicates that the power for this comparison was 43.2% ($\alpha = 0.05$).

5.4 Cell free DNA characteristics of non-metastatic STS patients

5.4.1 Intraoperative plasma cfDNA levels

Mean intraoperative total cfDNA concentration in the cohort was 10.5ng/ml plasma (±1.9 SEM, median 9.0, range 2.3-35.3) (Table 5.2). There was no significant difference seen in the intraoperative cfDNA levels of patents with different Trojani grade tumours, and no correlation between total cfDNA levels and tumour size (Figure 5.4). There was also no clear relationship identified between intraoperative cfDNA levels and STS subtype (Figure 5.5).

The mean intraoperative cfDNA levels in the non-metastatic STS patients was significantly lower that the metastatic patients analysed in chapter 4 ($P<0.0001$) although not significantly higher than the cfDNA levels seen in the healthy control group ($P=0.06$) (Supplementary 5.1).

The mean intraoperative cfDNA levels in those patients that developed recurrence during follow up was 10.5ng/ml (± 3.8 SEM, median 7.7, range 2.5-27.5) compared to 10.7ng/ml (± 2.1 SEM, median 8.6, range 2.3-35.3) in those that remained disease free. This difference was not significant (Figure 5.6).

5.4.2 Post-operative cfDNA levels

Post-operative cfDNA levels were calculated for 26 patients with a mean of 8.8ng/ml plasma (± 1.2 SEM, median 7.8, range 2-30) (Table 5.2). Paired intra- and post-operative cfDNA levels were available for 22 patients. Although the average cfDNA levels of this group fell following surgery, this was not significant (Figure 5.7).

The non-metastatic STS patients' post-operative cfDNA levels were significantly lower that the metastatic patients cfDNA analysed in chapter 4 ($P<0.0001$) although again not significantly higher than the healthy control group's cfDNA levels ($P=0.07$) (Supplementary 5.1).

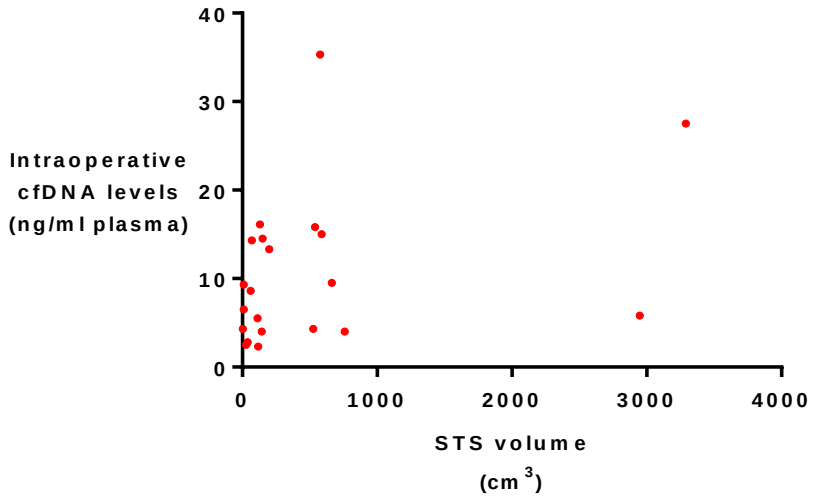
No significant difference was seen in the postoperative cfDNA levels or difference in intra- and post-operative cfDNA levels of those patents that suffered recurrence during follow up and those that remained disease free (Figure 5.6). Similarly there was no significant difference in post-operative cfDNA levels or difference in intra- and post-operative cfDNA levels of those patents that had their tumours resected widely at completion of their surgical management and those that underwent marginal R1 resections (planned or unplanned) (Figure 5.8).

The mean interval between patients' surgeries and the date their 1st post-operative plasma samples were collected was 27.7 days (range 5-76 days). No correlation was seen between the size of this interval and the magnitude of the change in cfDNA concentration seen intra- and post-operatively (see figure 5.9).

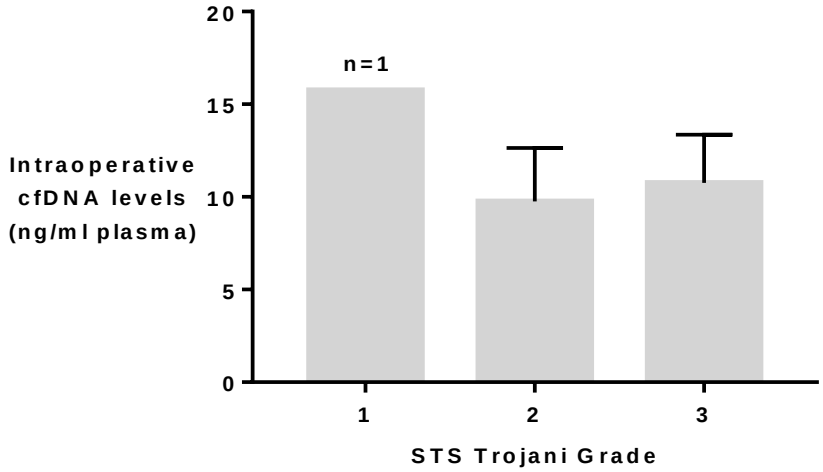
5.4.3: Cell free DNA levels and disease recurrence: Matched plasma samples post-operatively and at the point recurrence was diagnosed were collected from 7 of the 9 patients that recurred during follow up. Although cfDNA levels were an average of 1.6ng/ml plasma higher when recurrence was detected (10.3± 3.7 SEM vs 8.7± 2.4 SEM) this difference was not significant (Figure 5.10).

Patient	Follow up (months)	Intra-op. cfDNA levels (ng/ml)	Post-op. cfDNA levels (ng/ml)	STS recurrence	cfDNA level at recurrence (ng/ml)	Oncological outcome	Disease free survival (months)	Overall survival (months)	ctDNA detected (Y/N/n/a) (assay utilised)
1	18.8	9.5	5.5	Metastatic	6.5	AWD	18.8	-	N (tNGS)
3	30.1	15	7.8	No	n/a	NED	-	-	N (ddPCR)
6	9.5	Not available	2	Metastatic	3.5	AWD	9.5	-	Y (ddPCR)
9	1.3	13.8	3.2	Metastatic	3.2	AWD	1.3	-	N (ddPCR)
10	28.9	9.3	12	No	n/a	NED	-	-	N (tNGS)
17	17	15.8	7.3	No	n/a	NED	-	-	N (tNGS)
18	3.3	Not available	7.5	Metastatic	16.9	AWD	3.3	-	Y (ddPCR)
21	18.9	13.3	9.3	No	n/a	NED	-	-	Y (tNGS)
22	12.6	Not available	9.25	Metastatic	16	AWD	12.6	-	Y (ddPCR)
23	15.6	2.3	6.3	No	n/a	NED	-	-	Y (tNGS)
24	6.9	4.0	2	Metastatic	7.75	DOD	6.9	10.3	N (ddPCR)
25	11.8	Not available	4	unknown	3	Lost to FU	-	-	N (ddPCR)
26	15.1	4.0	3	No	n/a	NED	-	-	N (ddPCR)
27	11.2	10.9	16.1	No	n/a	NED	-	-	Not applicable
28	13.3	5.5	7	No	n/a	NED	-	-	N (tNGS)
29	9.3	4.3	4.8	No	n/a	NED	-	-	Not applicable
30	11.9	14.5	4.3	unknown	n/a	Lost to FU	-	-	Not applicable
31	4.6	5.8	30	Metastatic	16.75	DOD	4.6	5.3	N (tNGS)
32	10.2	4.3	9.5	No	n/a	NED	-	-	N (tNGS)
33	7.6	27.5	10.8	Metastatic	17	AWD	7.6	-	N (tNGS)
34	8.8	2.5	15.3	Local	n/a	NED	30	-	Not applicable
35	6.5	2.8	4.5	No	n/a	NED	-	-	Not applicable
36	4.9	35.25	17.3	No	n/a	NED	-	-	Not applicable
37	6.5	6.5	7.8	No	n/a	NED	-	-	Not applicable
38	4.4	8.6	7.6	No	n/a	NED	-	-	Not applicable
41	4.4	14.3	19.7	No	n/a	NED	-	-	Not applicable

Table 5.2: **Clinical outcome and circulating nucleic acid characteristics of non-metastatic STS patients. AWD**-alive with disease; **DOC**-died from other causes; **DOD**-died from disease; **NED**-no evidence of disease. **tNGS**-Targeted next generation sequencing; **ddPCR**-Droplet digital PCR.



a)



b)

Figure 5.4: Correlation between non-metastatic patients' STS characteristics and intraoperative cfDNA levels. 5.4a is a scatter graph correlating STS volume and intraoperative cfDNA levels. No significant correlation is seen between these factors (R^2=0.11 / P=0.13). 5.4b is a bar chart comparing the intraoperative cfDNA levels of patients with tumours of different Trojani grades. The mean and standard error of the mean are represented by error bars. The mean intraoperative cfDNA level was 9.8 (±2.9 SEM, n=11) for the grade 2 tumours and 10.8 (±2.6 SEM, n=9) for the grade 3 tumours. The single grade 1 STS analysed had an intraoperative cfDNA level of 15.8ng/ml. Tumour grade was unknown for one case. No significant difference was seen between groups (P=0.8, ANOVA).

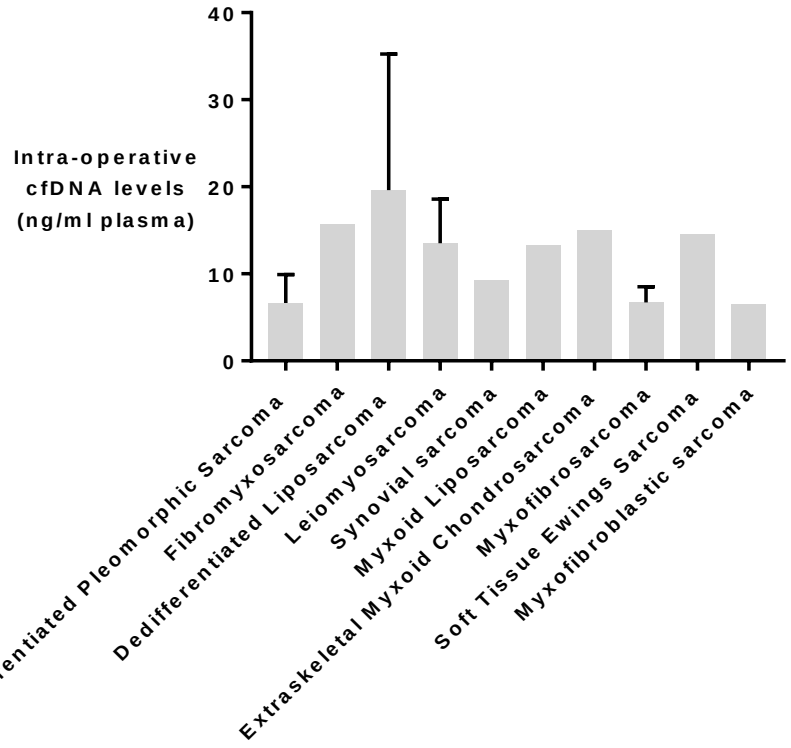
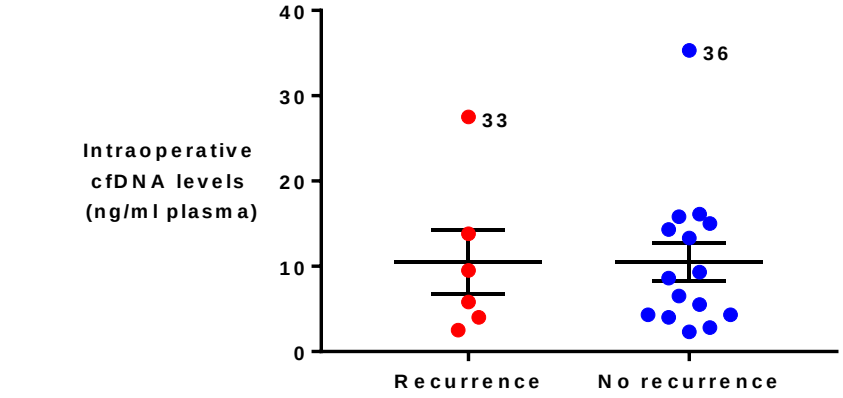
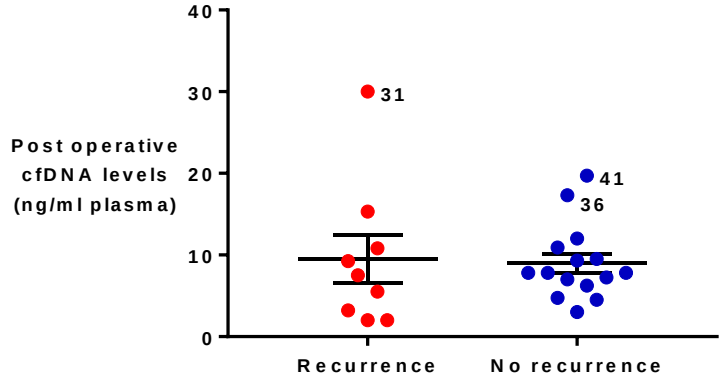


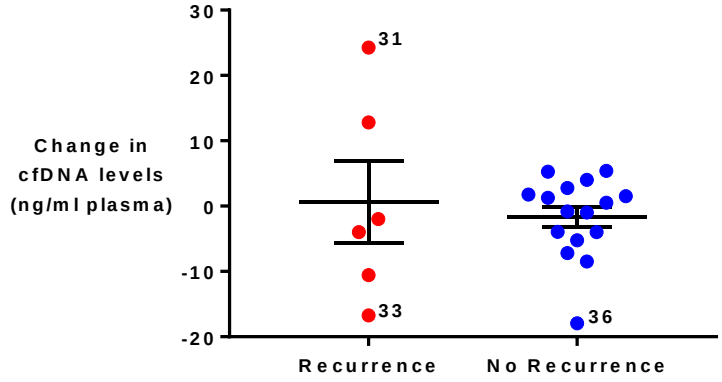
Figure 5.5: Intra-operative cfDNA levels of patients analysed categorised by STS subtype. Error bars representing the SEM are shown where multiple patients with the same STS subtype were analysed (undifferentiated pleomorphic sarcoma n=4, dedifferentiated liposarcoma n=2, leiomyosarcoma n=4, myxofibrosarcoma n=6). Cell free DNA levels were not significantly different between groups (P=0.8, ANOVA).



a)



b)



c)

Figure 5.6: Comparison of perioperative cfDNA levels and recurrence in non-metastatic STS patients. Horizontal bars represent the mean and SEM. The participant numbers of outliers are also shown. 5.6a compares the intra-operative cfDNA levels in both groups. Mean intraoperative cfDNA levels for those patients that suffered recurrence was 10.5ng/ml (± 3.8 SEM, median 7.7, range 2.5-27.5) compared to 10.5ng/ml (± 2.2 SEM, median 8.6, range 2.3-35.3) in those that remained disease free during follow up (P=0.99, unpaired t-test.) 5.6b compares the same groups' cfDNA levels

at their first post-operative follow up appointments. Mean cfDNA levels at this stage were 9.5ng/ml (± 1.2 SEM, median 7.5, range 2-30) in those patients that suffered recurrence and 9.0 ng/ml (± 1.2 SEM, median 7.8, range 3-19.7) in those that did not (P=0.85, unpaired t-test). 5.6c compares the difference in intra- and post-operative cfDNA levels between each group. On average cfDNA levels rose by a mean of 0.63ng/ml (± 6.2 SEM, median -3, range -16.8-24.3) following surgery in those patients that suffered recurrence. In contrast cfDNA levels fell by a mean of -1.6ng/ml (± 1.5 SEM, median -0.2, range -18-5.4) in those patients that remained disease free (P= 0.61, unpaired t-test). Retrospective analyses for the comparisons shown in Figures a and b revealed powers of 2.5% and 3.6% respectively (α = 0.05).

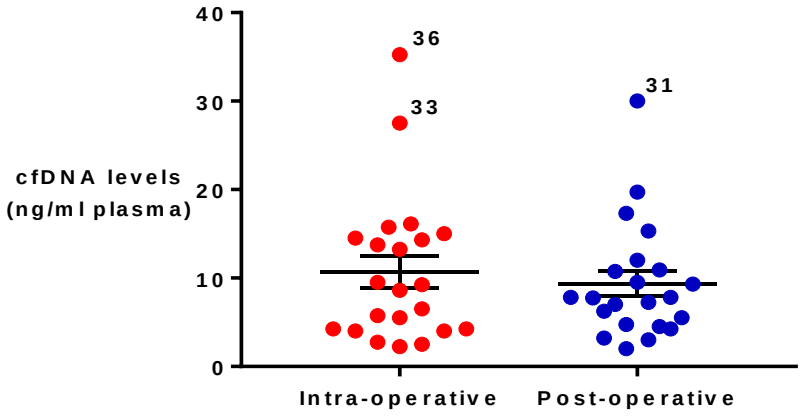
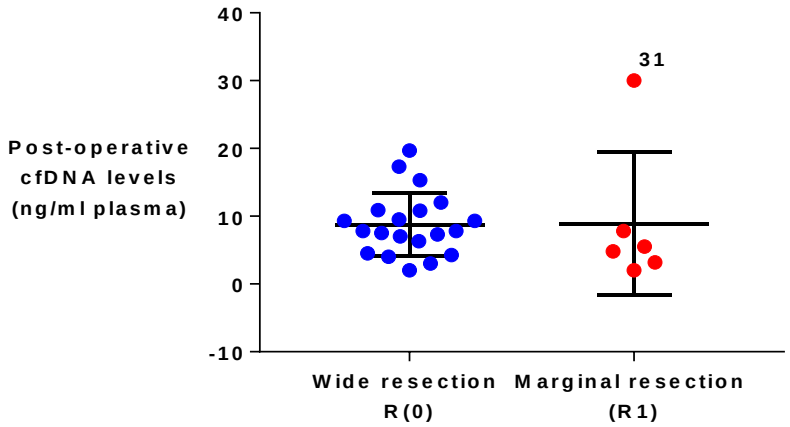
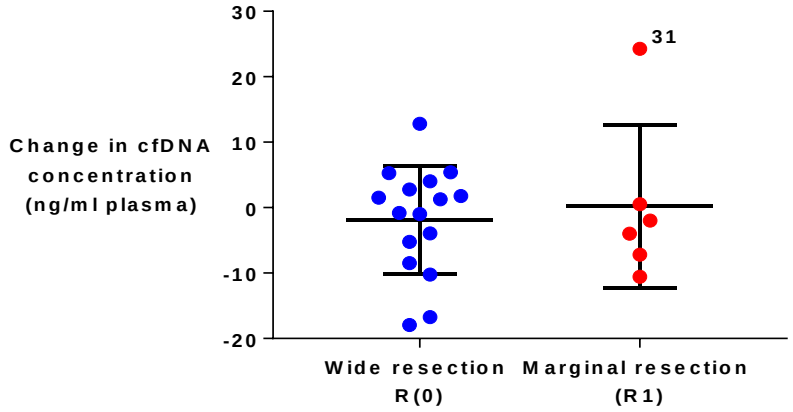


Figure 5.7: Perioperative cfDNA levels in non-metastatic STS patients. Mean intraoperative cfDNA levels were 10.7ng/ml (± 1.8 SEM, median 8.9, range 2.3-35.3) compared with 9.4ng/ml (± 1.4 SEM, median 7.8, range 2-30) post-operatively. This difference was not significant (P=0.56, paired t-test). Outliers' participant numbers are shown. A retrospective analysis indicates that the power for this comparison is 8.3% (α = 0.05).



a)



b)

Figure 5.8: Comparison of perioperative cfDNA levels and surgical resection margins in non-metastatic STS patients.

5.8a compares post-operative cfDNA levels in those patients that underwent wide (R0) resections and those that underwent marginal (R1) resections (unplanned and planned). No significant difference was seen in the post-operative cfDNA levels of these groups (P=0.97, unpaired t-test). A retrospective analysis indicates that the power for this comparison was just 2.6% ($\alpha = 0.05$).

5.8b plots the differences seen between each patient's intra-operative and post-operative cfDNA levels. The graph compares these differences in those patient that underwent a wide and marginal resection. No significant difference was seen between groups (P= 0.66 respectively, unpaired t-test). The participant number of 31 is shown as potential outlier.

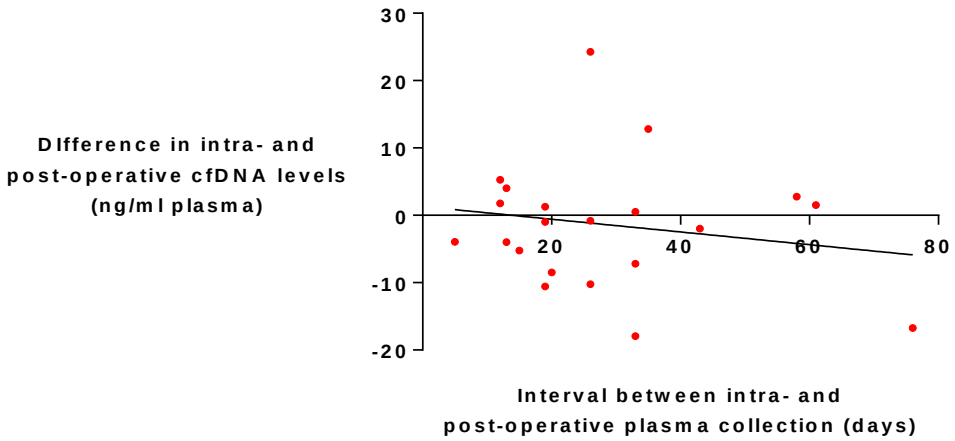


Figure 5.9: Correlation between perioperative cfDNA levels and timing of plasma sample collection. No significant correlation was seen between these variables (R^2= 0.03 / P=0.41).

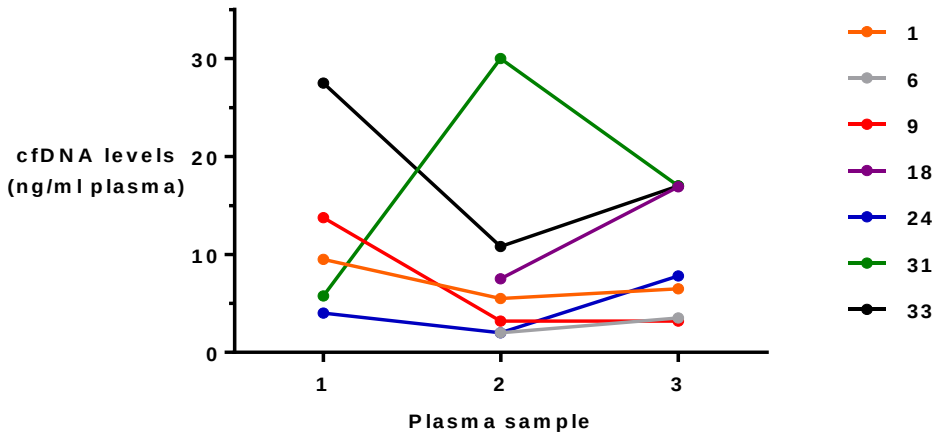


Figure 5.10: Longitudinal trends in cfDNA levels in recurrent STS patients. Cell free DNA levels intraoperatively (plasma sample 1), at patients' first post-operative follow up appointment (plasma sample 2) and at the point recurrence was shown (plasma sample 3). Patients' cfDNA levels rose by a mean of 1.6ng/ml from their 1st post-operative outpatient appointment to the point when recurrence was detected (10.3± 3.7 SEM vs 8.7± 2.4 SEM) although this difference was not significant (P=0.59, paired t-test). Intra-operative samples for patients 18/31 were not available for analysis.

5.5 Circulating tumour derived DNA characteristics of non-metastatic STS patients

In recent years both ddPCR and semiconductor tNGS have become used increasingly commonly to characterize ctDNA in cancer patients with early and late stage disease (Chen et al. 2016; Kaisaki et al. 2016; Riva et al. 2017; Huang et al. 2016). Both techniques have specific advantages including ddPCR's high sensitivity (down to 0.04% in some series (Yang et al. 2016)) and an Ampliseq^{TM} panel's ability to sequence multiple hotspot regions simultaneously. To maximise these advantages and our chances of successfully detecting ctDNA, we elected to use both techniques for our analysis.

5.5.1 Droplet digital PCR analysis of ctDNA characteristics of non-metastatic STS patients

We utilised ddPCR in all of those non-metastatic patients analysed from whom fresh frozen STS tumour tissue was collected on the day of their STS resections. We analysed this tissue using WES as outlined in chapter 3, and using the resulting data identified any SNVs present in their STSs. Individual ddPCR assays were then designed to detect these SNVs circulating in the corresponding patient's plasma. The analysis of these patients' tumour samples in this manner provided us with less need to look for multiple SNVs simultaneously (an advantage of tNGS) but instead allowed us to chase circulating variants in a truly mutation/patient specific manner whilst simultaneously taking advantage of ddPCR's excellent specificity. In total eight non-metastatic STS patients were analysed using ddPCR with 1-3 functioning ddPCR assays designed and used for each patient (Tables 2.3 and 2.4). This approach identified evidence of ctDNA based on a variant allele frequency (VAF) of >0.4% either intraoperatively or throughout follow in 3 of the 8 patients analysed (38%) (Table 5.2). The 3 patient vignettes are described discussed below.

5.5.1.1 Assay development and validation

Droplet digital PCR assays were developed and optimised as described in section 2.17.2. Each target variant present at a tumoural frequency of >20% was successfully identified in patients' tumour tissue using sanger sequencing (Figure 5.11). Primers' specificities were then confirmed using standard and Evagreen PCR (Figures 5.12 and 5.13) before each complete assay's optimal cycling conditions and sensitivities were confirmed down to a VAF of 0.3% (Figure 5.14 and 5.15).

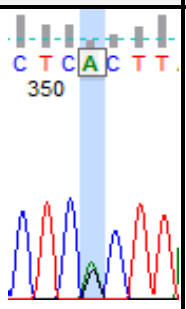
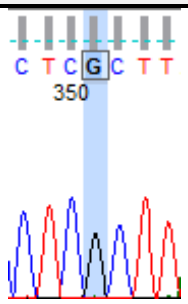
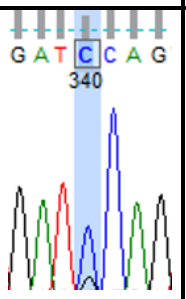
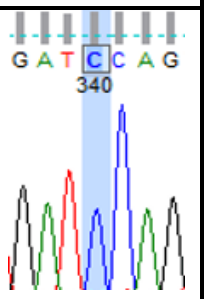
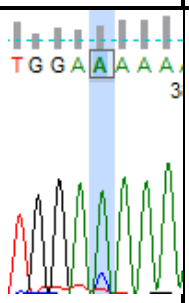
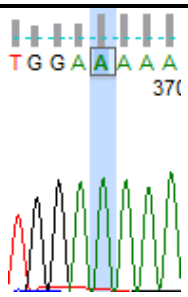
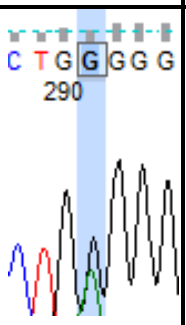
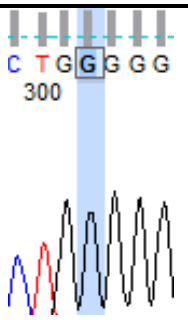
Patient	Variant	Tumoural variant frequency	Tumour DNA	BC DNA
6	*TP53* (17:7577022 G>A)	56%		
	BRIP1 (17:59761496 C>G)	21%		
22	*DACH1* (13:72053389 A>C)	21%		
	EPHB6 (7:142563798 G>A)	42%		

Figure 5.11: Sanger sequence chromatographs of single nucleotide variants targeted using ddPCR in patients 6 and 22. Participant 18's tumour DNA was also sanger sequenced but no evidence of *FLT4* (5:180046092, G>A) was identified due to the variant's low frequency.

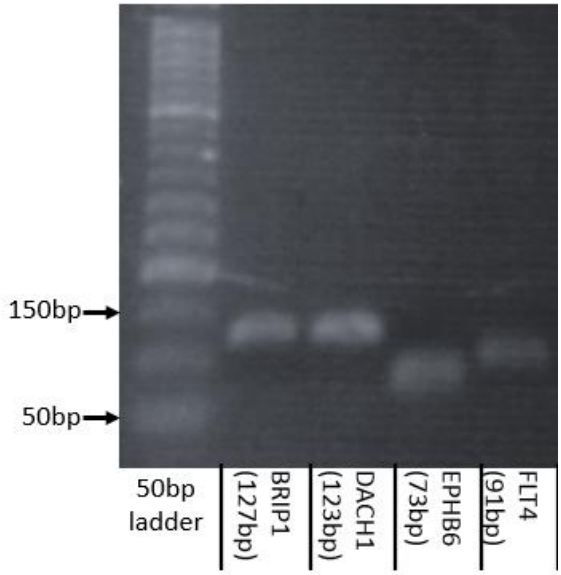


Figure 5.12: Gel electrophoresis of ddPCR primers used in participants 6, 18 and 22. Gene names and predicted amplicon sizes are shown.

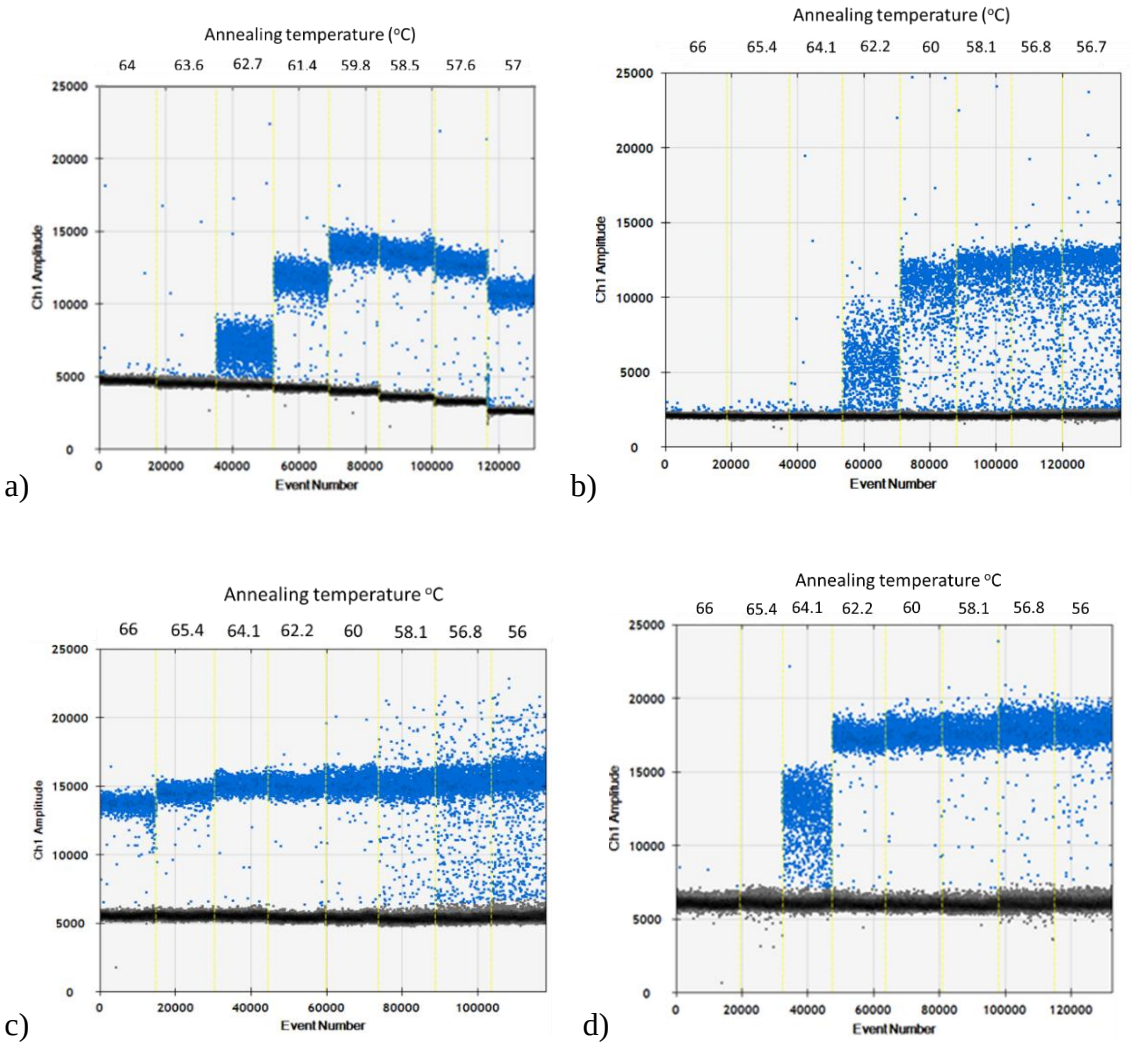
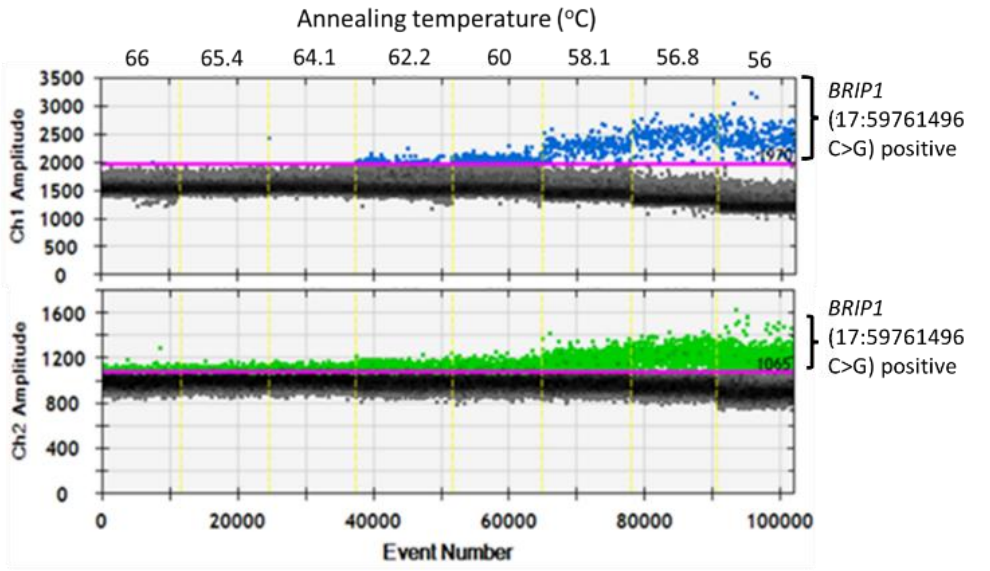
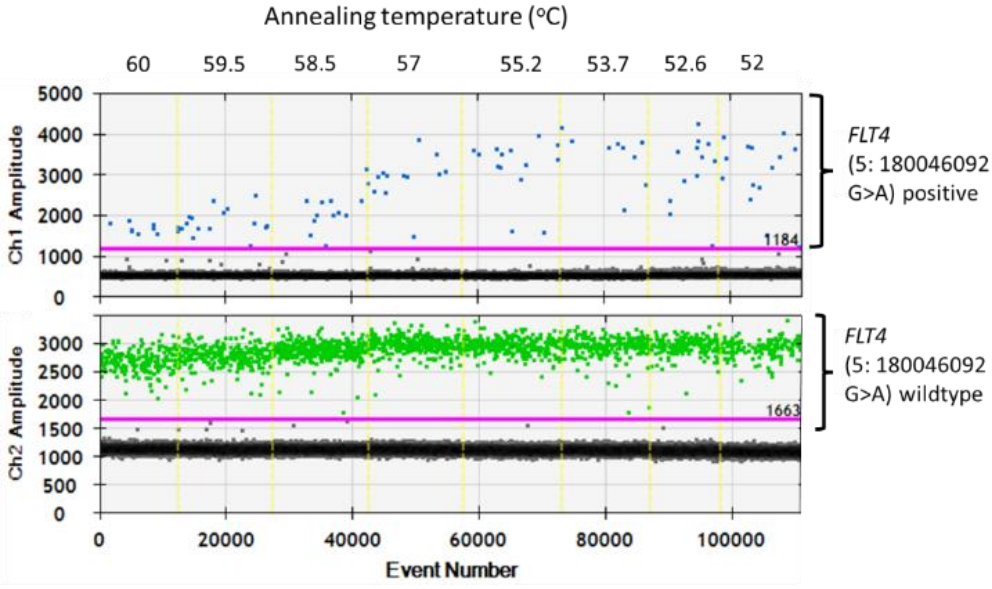


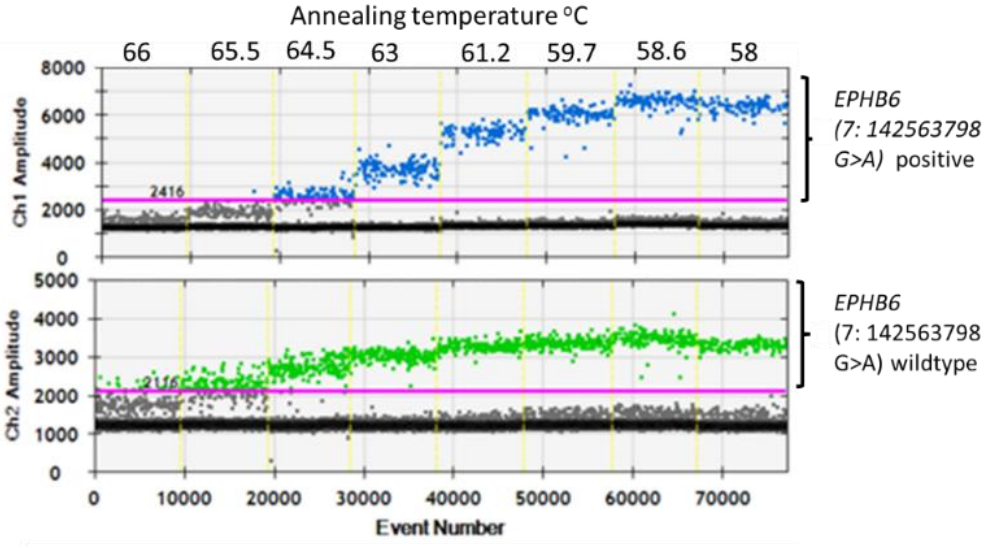
Figure 5.13: Droplet digital PCR assay primers' annealing temperature gradients. Figures show Evagreen annealing temperature gradients for ddPCR assay primers performed on the QX200™ Droplet Digital™ PCR platform. The x axis represent cumulative droplet count and the y axis fluorescence (FAM) amplitude. Human genomic DNA was used as a template for amplification (10ng/reaction). Results for primers target *BRIP1*. *FLT4*, *EPHB6* and *DACH1* variants are shown (Figure 5.13a, b, c, and d respectively).



5.14a)



5.14b)



5.14c)

5.14d)

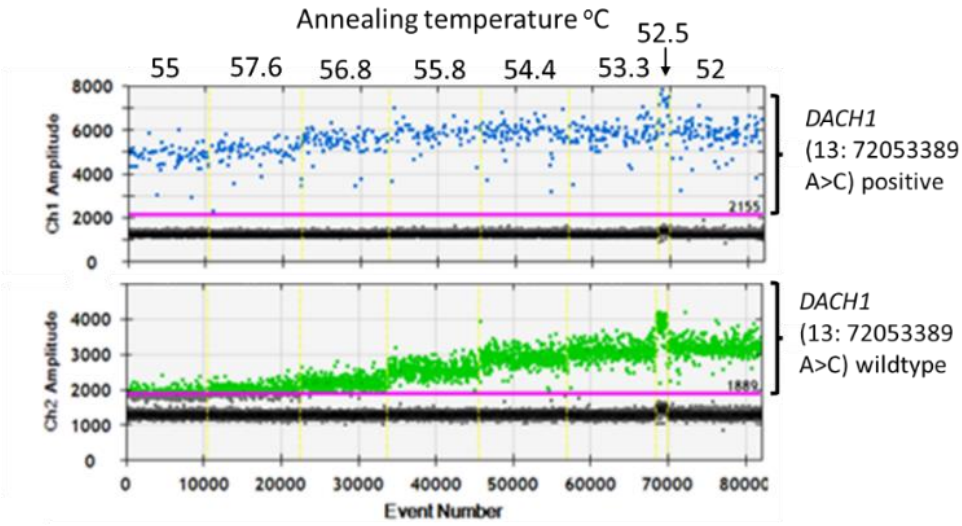
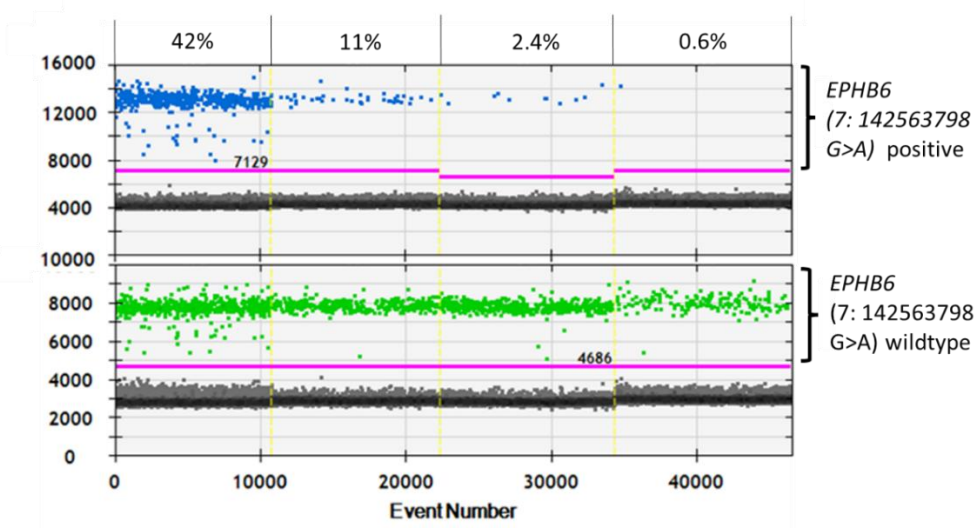
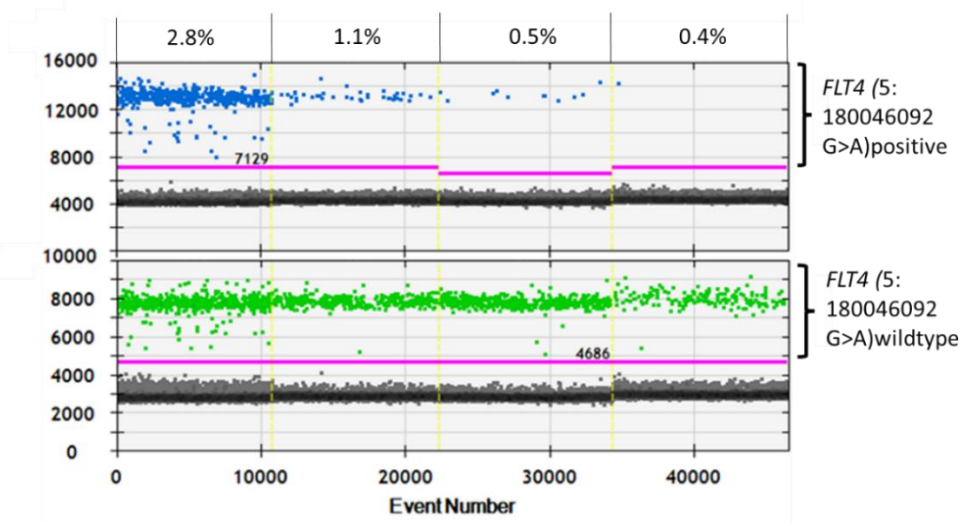


Figure 5.14: Optimisation of ddPCR assay cycling conditions.

Figure 5.14a-b represent annealing temperature gradients performed on the QX200™ Droplet Digital™ PCR platform for ddPCR assays developed in house for use on patients 6, 81 and 22. The x axes represent cumulative droplet count (event number) and annealing temperature. The fluorescent amplitudes of channel 1 (FAM/blue=mutant allele) and channel 2 (HEX/green=wildtype allele) are shown on the graphs' y axes. Participant STS DNA was used as a template (10ng/reaction). Data is shown for assays targeting *BRIP1*, *FLT4*, *EPHB6* and *FLT4* (Figure 1a, b, c and d respectively). Annealing temperatures of 57°C (*BRIP1*), 53°C (*FLT4*), 59°C (*EPHB6*) and 57°C (*DACH1*) were selected for use.



5.15a)



5.15b)

Figure 5.15: Validation of ddPCR assay sensitivities. Variant allele concentration gradients for ddPCR assays targeting *EPHB6* (7:142563798 G>A) and *FLT4* (5:180046092 G>A) are shown. The x axes represent cumulative droplet count (event number) and variant allele frequencies. The fluorescent amplitudes of channel 1 (FAM/blue=mutant allele) and channel 2 (HEX/green=wildtype allele) are shown on the graphs' y axes. Participant tumour DNA doped with human genomic DNA was used as template in each reaction. Figure 5.15a shows data for the *EPHB6* (7: 142563798 G>A) assay with predicted VAFs of 44%, 11%, 2.75% and 0.69%. Figure 5.15b shows data for the *FLT4* (5:180046092 G>A) assay with predicted VAFs of 2.3%, 1.1%, 0.6% and 0.3%.

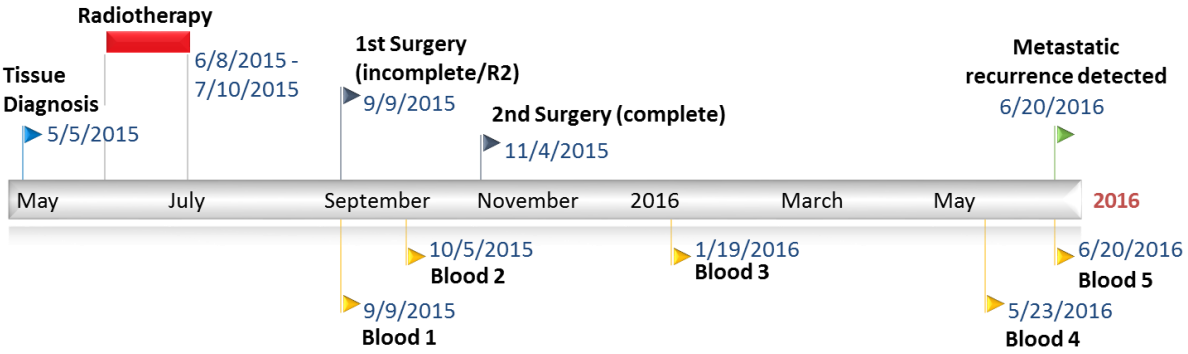
5.5.1.2 Patient 6:

Clinical characteristics: Patient 6 (M/54 years) had an undifferentiated pleomorphic sarcoma (grade 2) treated with neoadjuvant radiotherapy and wide surgical resection (achieved after 2 surgeries). Following surgery five plasma samples were collected from participant 6 before a diagnosis of metastatic (pulmonary) recurrence was made 33 weeks post operatively (Figure 5.16).

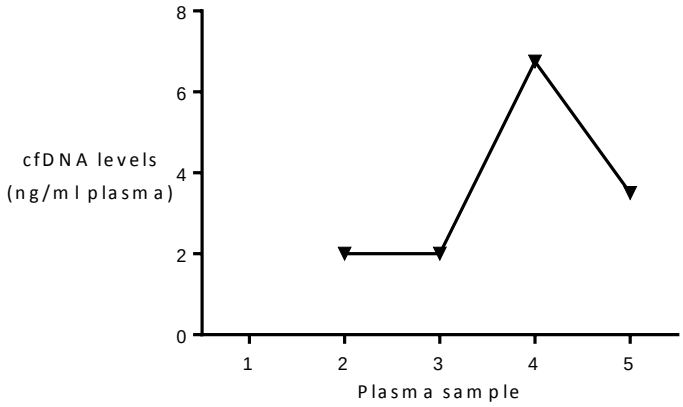
Circulating tumour derived DNA characteristics: Based on exome data two somatic mutations were analysed in patient 6's cfDNA - *TP53* (17:7577022 C>T) and *BRIP1* (17:59761496 C>G) (Table 5.2).

No circulating *TP53* (17:7577022 C>T) was detected in the plasma samples collected intra-operatively or 4 weeks post operatively. Following this the levels of circulating *TP53* (17:7577022 C>T) rose in plasma samples collected 11 and 29 weeks post operatively, before falling again to become undetectable in the sample collected 33 weeks post operatively at the point of recurrence (Figures 5.16 and 5.17).

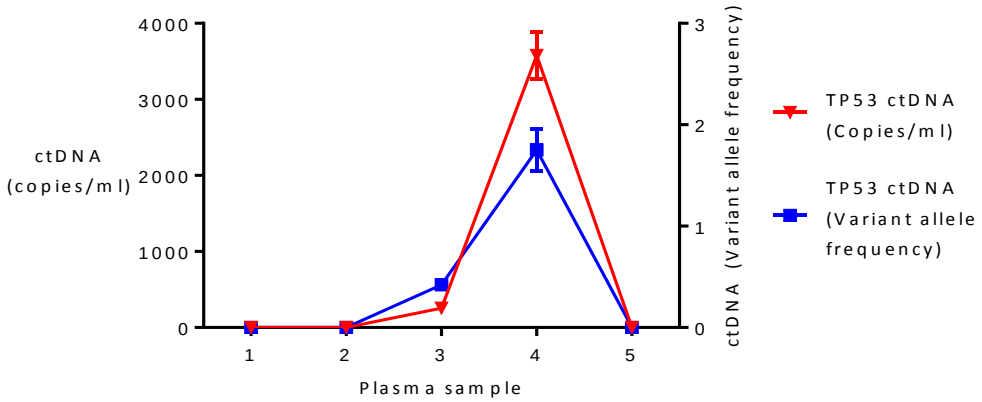
No evidence of circulating *BRIP*1 (17:59761496 C>G) was identified in plasma samples collected 4 or 11 weeks post operatively (see supplementary 5.2). Insufficient cfDNA was available to look for *BRIP1* (17:59761496 C>G) in the remaining plasma samples.



a)



b)



c)

Figure 5.16: Management, outcome and circulating nucleic acid characteristics of patient 6. 5.16a is a timeline that plots patient 6's progress from diagnosis throughout treatment and follow-up until the point of recurrence. The dates that every plasma sample was collected are also shown. 5.16b shows overall cfDNA concentration in each plasma sample collected. This analysis was not possible for plasma sample 1 collected intraoperatively. 5.16c shows the levels of

circulating *TP53* (17:7577022 C>T) in each plasma sample collected. The right axis represents the variant allele's circulating fractional abundance (i.e. the % of circulating nucleic acids that contain the variant allele). The left axis represents the plasma concentration of the variant allele in copies/ml. Circulating *TP53* (17:7577022 C>T) was identified in plasma samples 3 (VAF 0.42, 254 copies/ml), and 4 (VAF 1.75 (SEM 0.15), 3580 copies/ml (SEM 307)). Error bars are shown where DNA yields facilitated biological replicates and represent the standard error of the mean.

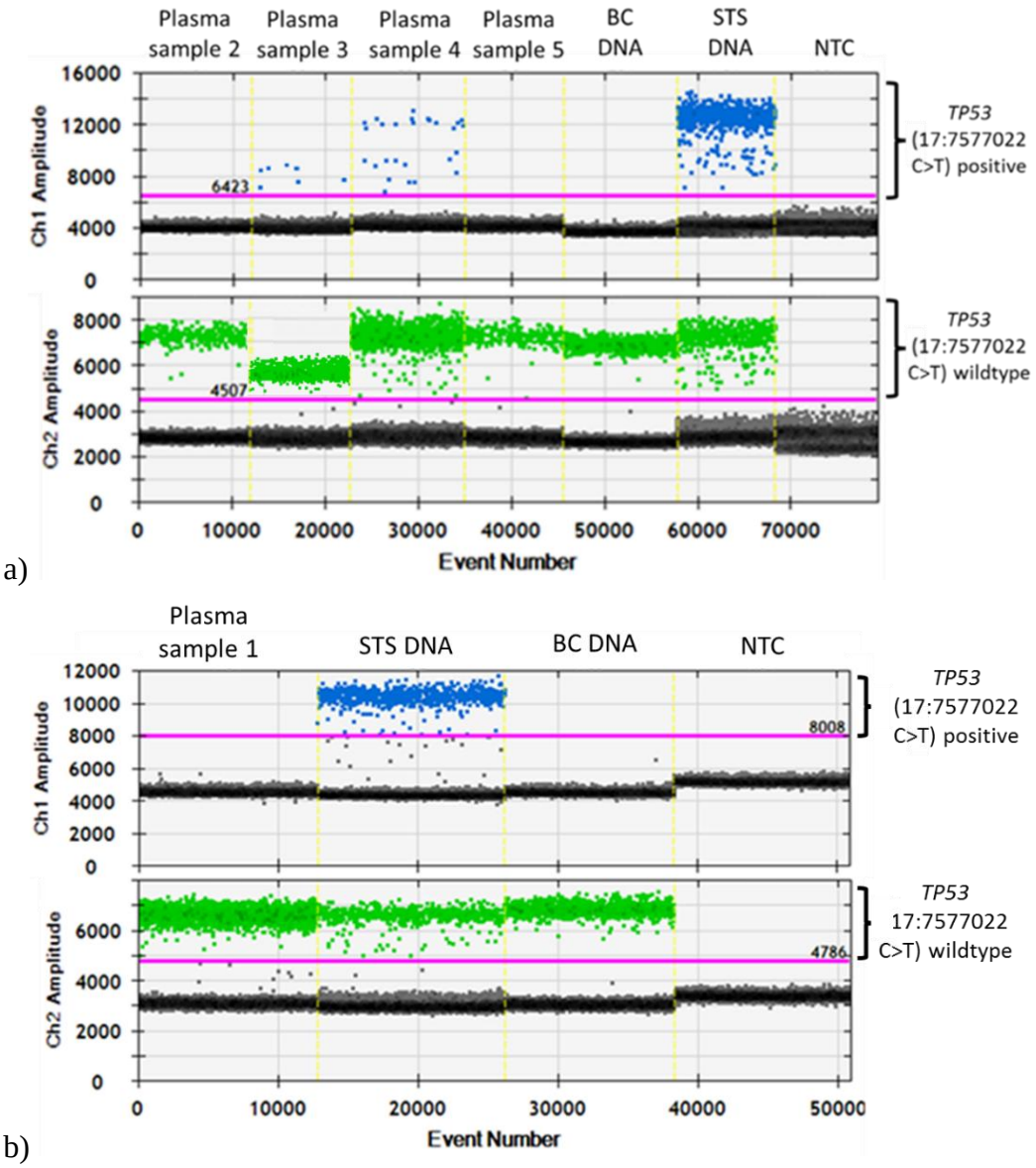
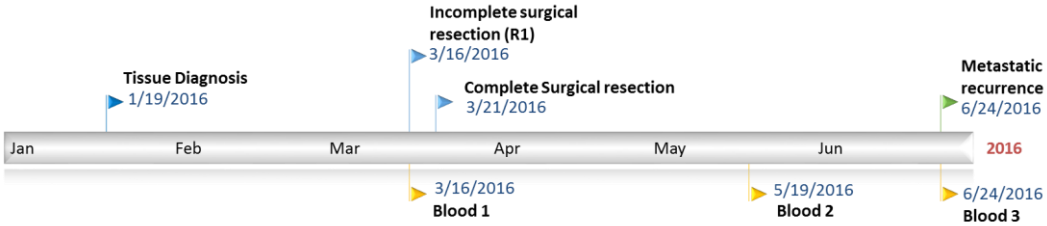


Figure 5.17: Droplet digital PCR mutation analysis of patient 6 targeting *TP53* (17:7577022 C>T). Cumulative droplet count (event number) are represented on the graphs' x axes and the fluorescent amplitudes of channel 1 (FAM/blue=mutant allele) and channel 2 (HEX/green=wildtype allele) are shown on the graphs' y axes. Samples of BC and STS DNA (10ng/reaction) and a non-template control (NTC) reaction were used as controls. 5.17a shows the analysis of plasma samples 2-5 collected 4, 11, 29 and 33 weeks post-operatively respectively. 5.17b shows the analysis of plasma sample 1 collected intra-operatively. Circulating *TP53* (17:7577022 C>T) was identified in plasma samples 3 (VAF 0.42, 254 copies/ml), and 4 (VAF 1.75 (SEM 0.15), 3580 copies/ml (SEM 307)).

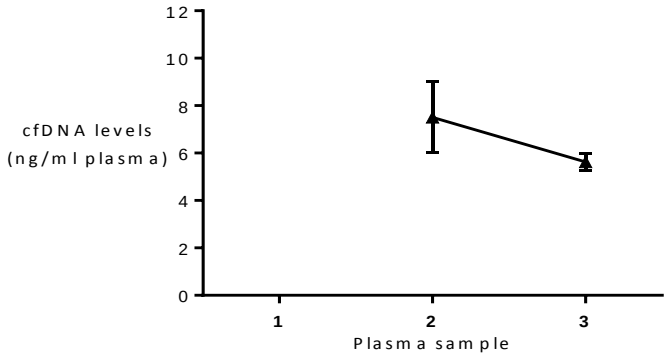
5.5.1.3 Patient 18:

Clinical characteristics: Patient 18 (F/79 years) had an extensive Haemangiosarcoma (grade unknown) affecting her leg for which she underwent an above knee amputation following an initial failed attempt at a limb sparing wide local resection. During follow up two post-operative plasma samples were collected from patient 18 before metastatic (pulmonary) recurrence was diagnosed 14 weeks post operatively (Figure 5.18).

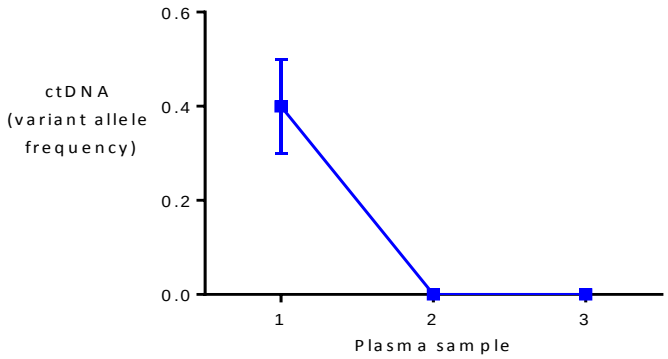
Circulating tumour derived DNA characteristics: One somatic mutation was analysed in patient 18's plasma cfDNA - *FLT4 (*5: 180046092 G>A). This variant was identified in participant 018's circulation intraoperatively although in neither sample collected 5 or 14 weeks post operatively (Figure 5.18 and 5.19).



a)



b)



c)

Figure 5.18: Management and circulating nucleic acid characteristics of participant 18. 5.18a is a timeline that plots patient 18's progress from diagnosis throughout treatment and follow-up until the point of disease recurrence. The dates that each plasma sample was collected are also shown. 5.18b is a graph that shows overall cfDNA concentration in plasma samples 2 and 3. Insufficient DNA precluded this analysis for the plasma sample collected intraoperatively. 5.18c shows the levels of circulating plasma *FLT4* (5: 180046092 G>A) in each plasma sample collected. Circulating *FLT4* (5: 180046092 G>A) was identified at a VAF of 0.4 (SEM 0.1) intraoperatively although fell away to become undetectable throughout follow up. Error bars are shown where DNA yields facilitated biological replicates and represent the standard error of the mean.

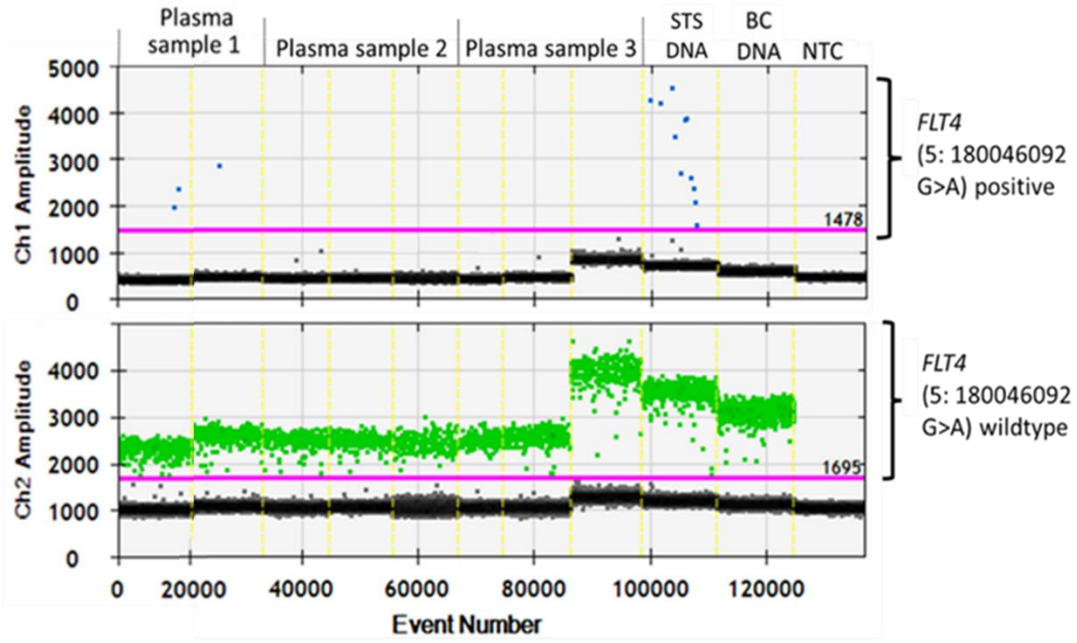


Figure 5.19: Droplet digital PCR analysis of patient 18's plasma cfDNA targeting *FLT4* (5: 180046092 G>A). Cumulative droplet count (event number) are represented on the graphs' x axes and the fluorescent amplitudes of channel 1 (FAM/blue=mutant allele) and channel 2 (HEX/green=wildtype allele) are shown on the graphs' y axes. Samples of participant 18's BC and STS DNA (10ng/reaction) and a non-template control (NTC) reaction were used as controls. Circulating *FLT4* (5: 180046092 G>A) was only identified in plasma sample 1 collected intra-operatively at a fractional abundance of 0.4 (SEM 0.1) (1091 copies/ml, SEM 273). The results of biological duplicates (sample 1) or triplicates (samples 2/3) are shown.

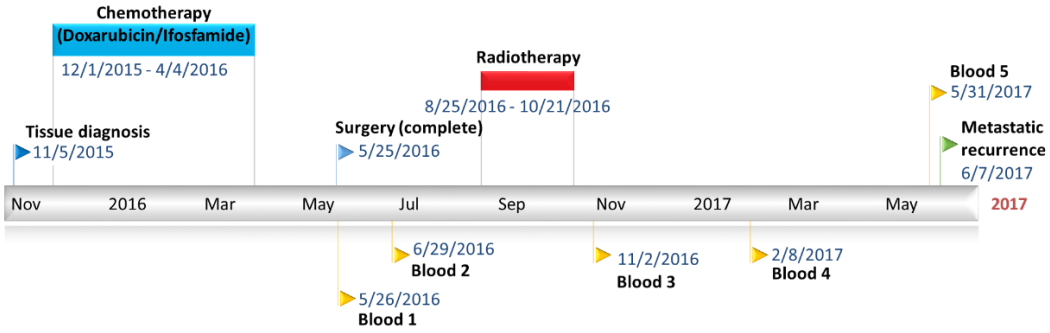
5.5.1.4 Participant 22:

Clinical characteristics: Patient 22 (F/64 years) had an undifferentiated pleomorphic sarcoma (grade 3) treated with neoadjuvant chemotherapy followed by a wide surgical resection and adjuvant radiotherapy. During follow up four plasma samples were collected from patient 22 at 5, 23, 37 and 53 weeks post operatively before a diagnosis of metastatic (pulmonary) recurrence was made 54 weeks post operatively (Figure 5.20).

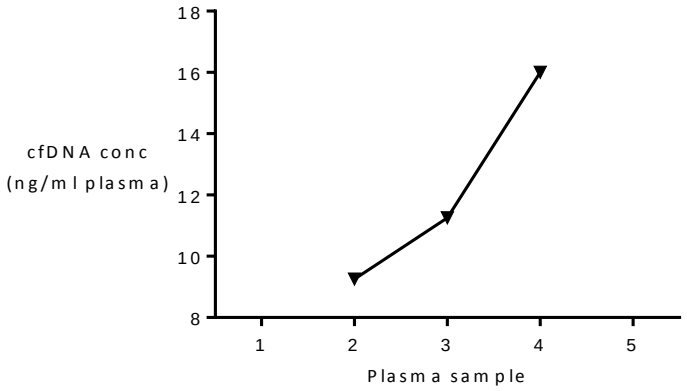
Circulating tumour derived DNA characteristics: Two variants were analysed in patient 22's plasma - *DACH1* (13: 72053389A>C) and *EPHB6* (7: 142563798 G>A).

Circulating *EPHB6* (7:142563798 G>A) was identified intraoperatively in participant 22. Following surgery the variant initially disappeared from the circulation but returned 23 weeks post-operatively before steadily rising up to the point of recurrence 53 weeks post-operatively (Figure 5.20 and 5.21).

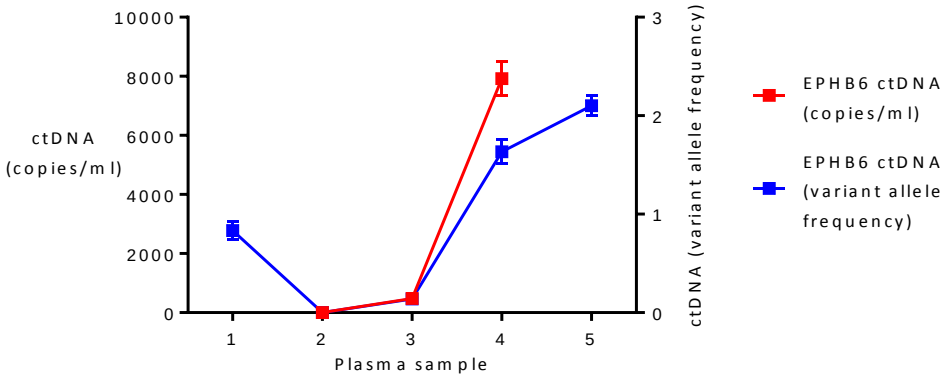
No evidence of circulating *DACH1* (13: 72053389A>C) was identified at any stage of patient 22's analysis (Supplementary 5.3).



a)



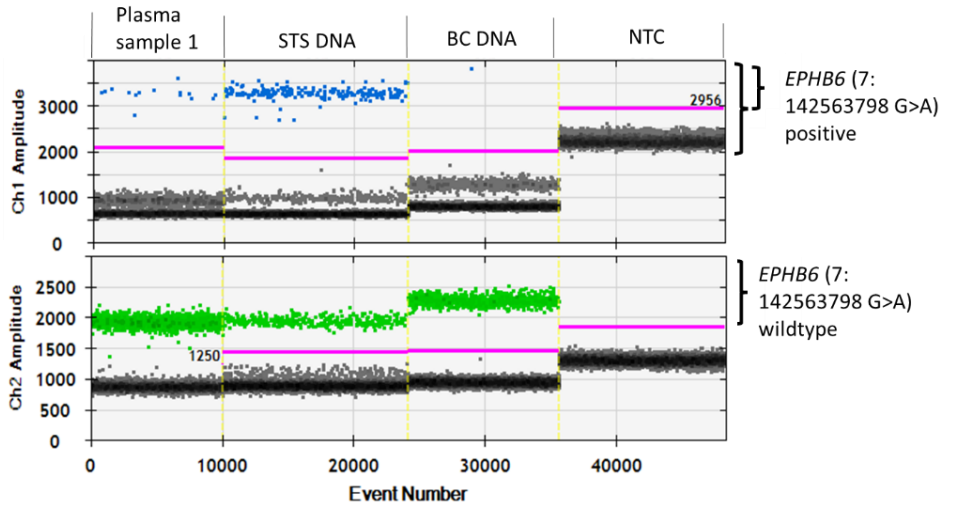
b)



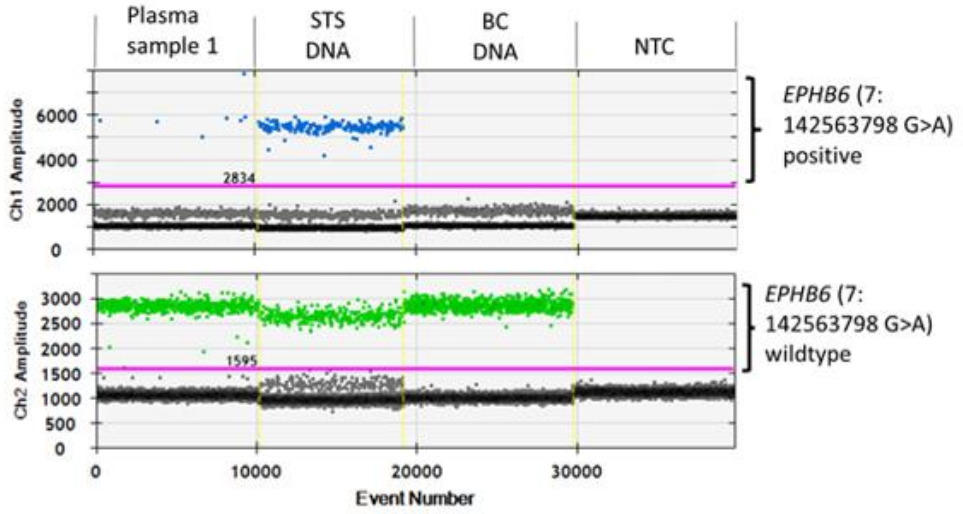
c)

Figure 5.20: Management, outcome and circulating nucleic acid characteristics of patient 022. 5.20a is a timeline that plots patient 22's progress from the point of diagnosis throughout treatment and follow-up until the point of recurrence. The dates that each plasma sample was collected are also shown. 5.20b shows overall cfDNA concentration in plasma samples 2-4. Insufficient DNA precluded this analysis for plasma collected intraoperatively and plasma sample 5 collected 1 week prior to the diagnosis of recurrence. 5.20c shows the levels of circulating plasma *EPHB6* (7:142563798 G>A) in each plasma sample collected. The right axis represents the variant allele's circulating fractional abundance (i.e. the % of circulating nucleic acids that contain the variant allele) and the left allele represents the plasma concentration of

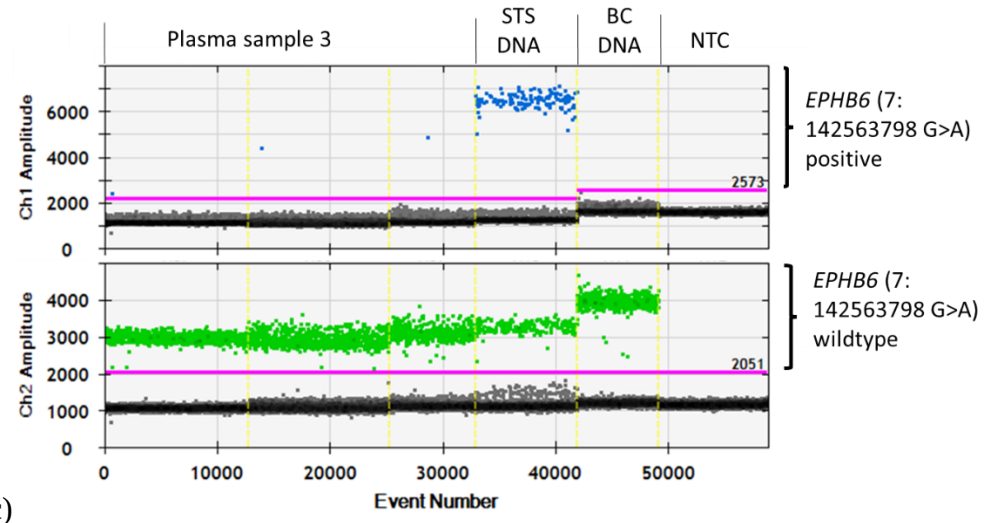
the variant allele in copies/ml. It was not possible to deduce the absolute number of copies of *EPHB6* (7:142563798 G>A) intraoperatively or in sample 5 as cfDNA levels are unavailable for these samples. Error bars representing standard error of the mean are shown at the time points where biological replicates were possible. Circulating *EPHB6* (7:142563798 G>A) was identified intraoperatively (VAF 0.85 (SEM 0.15) and in plasma samples 3 (VAF 0.14 (SEM 0.03), 477 copies/ml (SEM 102)), 4 (VAF 1.6 (SEM 0.12), 7919 copies/ml (SEM 582)) and 5 (VAF 2.1 (SEM 0.1)).



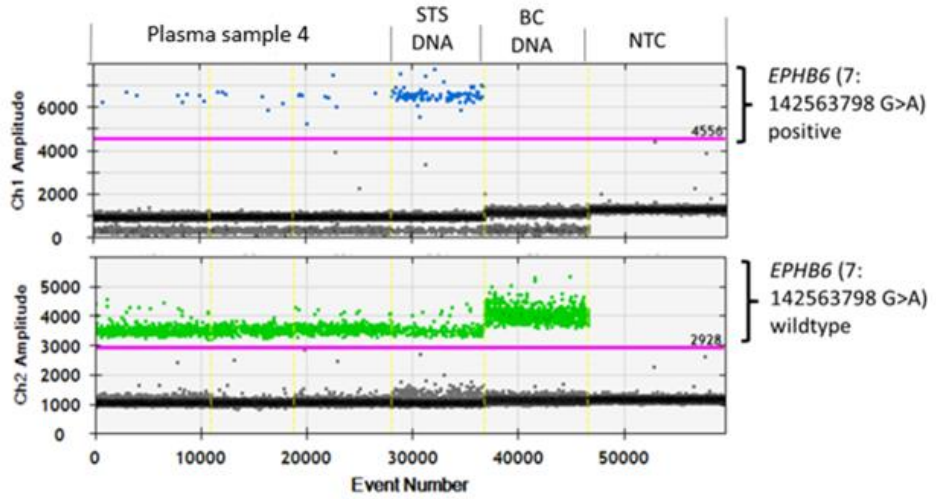
5.21a)



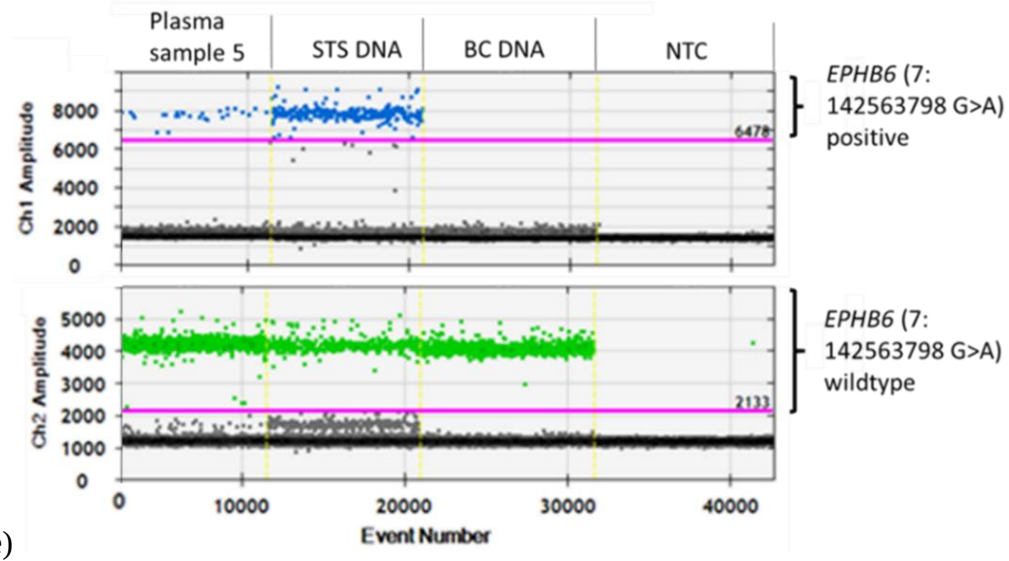
5.21b)



5.21c)



5.21d)



5.21e)

Figure 5.21: Droplet digital PCR analysis of patient 22's plasma cfDNA targeting *EPHB6* (7: 142563798 G>A). Cumulative droplet count (event number) and annealing temperature are represented on the graphs' x axes. The fluorescent amplitudes of the 2 channels used are represented on the graphs' y axis (channel 1/FAM=mutant allele, channel 2/HEX=wildtype allele). Participant 22's BC DNA, STS DNA, and a non-template control (NTC) reaction were used as controls. Figures 5.21a and b, c, d and e show the analyses of plasma samples 1, 3, 4 and 5 respectively. Circulating *EPHB6* (7:142563798 G>A) was identified intraoperatively at a fractional abundance of 0.85 (SEM 0.15). The variant was absent from plasma sample 2 (data not shown) although reappeared in sample 3 (fractional abundance 0.14 (SEM 0.03), 477 copies/ml (SEM 102)) before going onto steadily rise in concentration in samples 4 (fractional abundance 1.6 (SEM 0.12) 7919 copies/ml (SEM 582)) and 5 (fractional abundance 2.1 (SEM 0.1). The results of biological replicates are shown where DNA yields allowed.

5.5.2 Iontorrent semiconductor NGS analysis of ctDNA characteristics of non-metastatic STS patients

Samples from 9 non-metastatic STS patients were analysed using the Sarcoma V2 Iontorrent AmpliSeq™ panel. Although matched STS tissue was collected from 6 of these patients, each of these failed our internal QC to be sent for WES, predominantly because of macroscopic tissue necrosis or contamination with normal tissue. As a result we had no reliable way to determine which SNVs were present in any of these 9 patients' STSs, and so to take advantage of tNGS' ability to analyse multiple variants simultaneously we elected to analyse their plasma samples using an Ampliseq panel. The AmpliSeq™ panel utilised for this purpose performed well with a median sample read depth of 8052 (2163- 29592) and 99.7% of reads aligning accurately to their target regions. However a manual review of every amplicon's performance in IGV revealed that 6 performed poorly, defined as a mean read depth of <1000. All of these amplicons were subsequently excluded from our analysis, two of which amplified regions in *RB1* and 4 regions in *ATRX*.

Sequencing of STS tissue revealed one cosmic registered mutation in 2 patients at frequencies of 1% and 18% (*ATRX* X:76814213 T>C and *TP53* 17:7578526 G>T respectively) (Table 5.3). However, neither of these mutations were detected in the matched plasma cfDNA when libraries were created with 10ng of starting cfDNA.

Plasma analysis revealed circulating cosmic registered mutations unique to the plasma (i.e. not identified in matched STS tissue) in 2/9 patients analysed at frequencies of 0.5-1% (Table 5.3 and Figure 5.22). These 4 cases are described below.

5.5.2.1 Iontorrent Sarcoma V2 panel validation

The Iontorrent sarcoma V2 panel was validated by sequencing the tumour and BC of two patient that each had one somatic SNVs identified during the WES analysis outlined in chapter 3 in region covered by the panel's amplicons. This analysis successfully identified both *TP53* variants at frequencies comparable to those seen following WES (See Supplementary 5.4).

5.5.2.2 Patient 21: Patient 21 (F/77yrs) had a myxoid liposarcoma (grade 2) treated with neoadjuvant radiotherapy and a wide surgical resection. Following surgery patient 21 was followed up for 83 weeks during which time no evidence of disease recurrence was noted. Four plasma samples were collected during this time at one, 10, 28 and 40 weeks post operatively. None of patient 21's STS was available for analysis.

Low frequency circulating variants were identified in the plasma samples collected from patient 1 intraoperatively (*RB1* 13:49050968 A>G, 0.9%) and 10 weeks post-operatively (*RB1* 13:48947576 T>C 0.5%) (Table 5.5).

5.5.2.3 Patient 23: Patient 23 (M/53yrs) had an undifferentiated pleomorphic sarcoma (grade 2) treated with neoadjuvant radiotherapy and wide surgical resection. During follow up which lasted 70 weeks no evidence of recurrence was detected and 4 plasma samples were collected 2, 21, 35 and 49 weeks post-operatively.

No cancer associated single nucleotide variants were identified in patient 23's tumour DNA. Despite this two low frequency circulating COSMIC ID registered variants were detected in plasma collected 21 weeks post-operatively (*TP53* 17:7577116 T>C 0.6%, *TP53* 17:7578346 G>A 0.7%) (Table 5.3).

								STS tissue DNA			cfDNA DNA			
Pt	Chr	Loci	Gene	Coding strand	Base change	Cosmic ID	Predicted effect	Depth (reads)	Variant reads	%	Depth (reads)	Variant reads (%)	Total cfDNA (ng/ml)	Plasma sample (weeks post-op)
17	X	76814213	*ATRX*	-	T>C	4971451/2	p.(Asp2106Gly)	5477	23+/28-	1%	Not detected			
21	13	49050968	*RB1*	+	A>G	4807437/8	p.(Glu884Glu)	Not available			1628	9+/5- (0.9%)	13.3	IO
	13	48947576	*RB1*	+	T>C	136213/4	p.(Ile388Thr)				2407	5+/7- (0.5%)	5.8	PO (10)
23	17	7577116	*TP53*	-	T>C	1386598 / 45924	p.(Val274Val)	Not detected			30247	92+/82- (0.6%)	5.0	PO (21)
	17	7578346	*TP53*	-	G>A	45841	Intronic	Not detected			30344	104+/100- (0.7%)		
32	17	7578526	*TP53*	-	G>T	303849-52	p.(Cys135Phe)	25409	2003+ / 2534-	18%	Not detected			

Table 5.3: Tumour and plasma variants identified in non-metastatic STS patients analysed using targeted NGS. Data is shown from the analysis of patient samples using the sarcoma V2 AmpliSeq™ panel. **IO** – intra-operative; **PO** – post-operative.

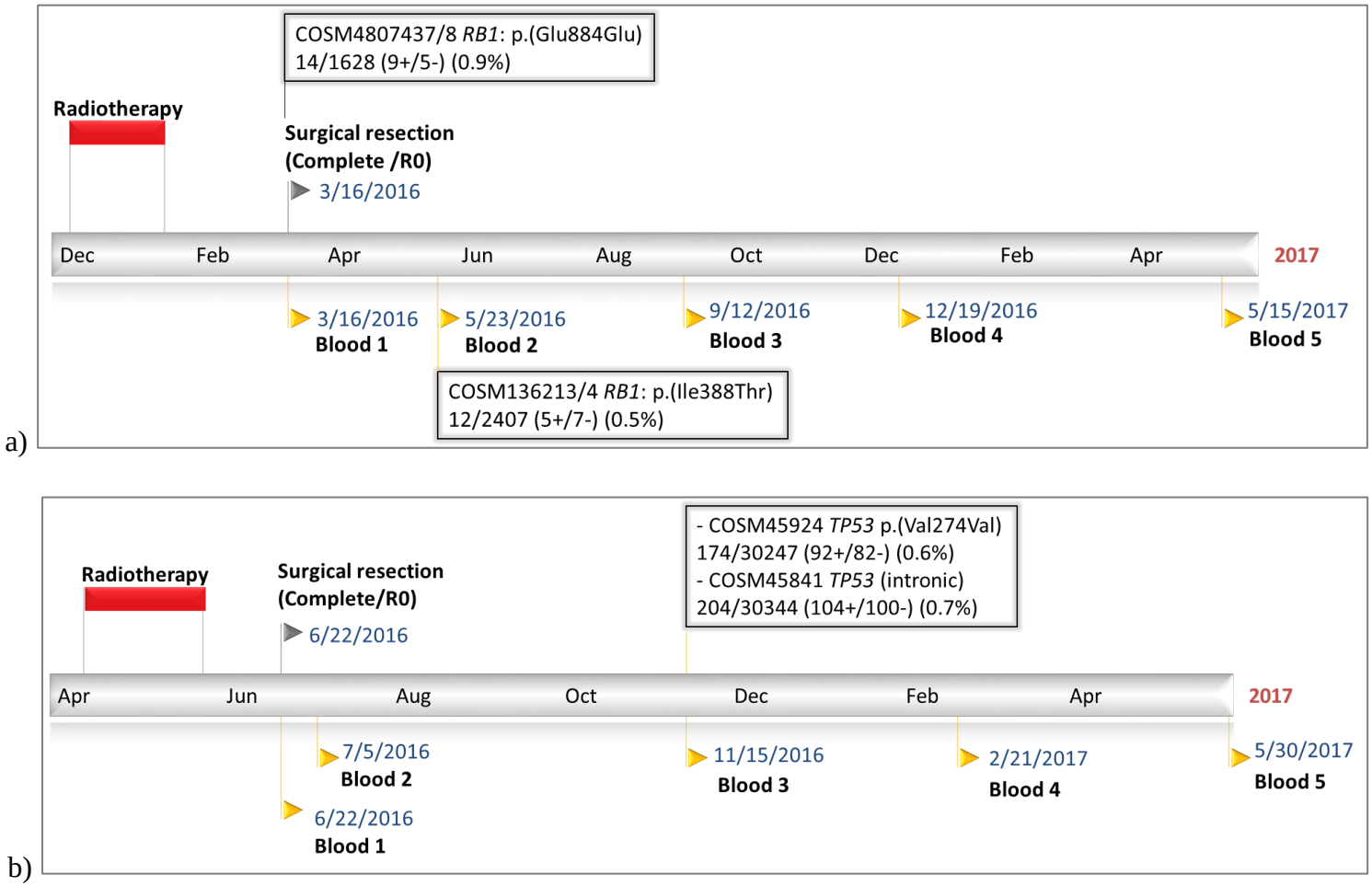


Figure 5.22: Circulating variants detected using targeted next generation sequencing in on metastatic STS patients. Figure 5.22a shows data for patient 021. Figure 5.22b shows patient 023.

5.6 Discussion

5.6.1 Patient demographics: The aim of the work described in this chapter was to investigate the circulating nucleic acid characteristics of a cohort of non-metastatic STS patients undergoing curative treatment. The patients enrolled for this analysis had similar demographics to those expected of STS patients in terms of gender and age, and included a wide variety of STSs in terms of tumour size, grade and subtype. This allows the data emerging from the analysis of these patients to be tentatively extrapolated to STS patients generally, despite the cohort's relatively small size.

5.6.2 Patient management: The management of every STS patient analysed was planned by a National Institute of Health and Clinical Excellence accredited STS treatment centre MDT. The unplanned marginal resection rate seen in our cohort was 19% (5/26 patients). This is higher than reported by some (as low as 5.3% in some series (O'Donnell et al. 2014)) although this may reflect differences in sample sizes (26 vs 1371). Five STS patients in the cohort (24%) analysed underwent planned marginal resections. Again this is a higher rate than reported by some other groups (3.4% (Gerrand et al. 2001) - 13% (Gunar K Zagars et al. 2003)) this is likely a reflection of the shift seen in recent years towards limb sparing surgery for extremity STSs, especially considering the presence of just 2 amputations in our series. Overall 17/26 patients (65%) in the analysed cohort received (neo)adjuvant radiotherapy, which is a similar rate to those reported by other groups (Bagaria et al. 2014). Only 2/26 patients (8%) in our cohort received systemic (neo)adjuvant treatment, which given the limited indications for the use of current chemotherapeutic regimes in non-metastatic STS patients is also unsurprising (Dangoor et al. 2016).

5.6.3 Patient outcome: The patients in our cohort have a 12 month recurrence rate of 30% including a 35% rate of metastatic recurrence and 4% rate of local recurrence. Although published outcomes on STS patients over periods shorter than 2-5 years are not common, metastatic and local recurrence rates of 12% (Hovgaard et al. 2017) - 25% (Rothermundt et al. 2014) and 18% (Hovgaard et al. 2017) respectively have been reported.

Unsurprisingly poor prognostic factors for the development of metastatic STS recurrence include large (>5cm), high grade tumours. Considering this, the relatively high rate of metastatic recurrence seen in our cohort may be explained by large average size (498cm^3) and grade (2.7) of the tumours included, as well as the inclusion of multiple cases of leiomyosarcoma and synovial sarcoma (Zagars et al. 2003) (STS subtypes known to carry a relatively high risk of metastatic recurrence.) This is supported by the fact that those patients that suffered metastatic recurrence in our cohort had significantly larger, (1063cm^3 vs 265cm^3, p=0.04) higher grade (2.6 vs 2.3) tumours than those patients that remained disease free when compared directly.

Only one patent in our cohort suffered local recurrence during our follow up. This patient had a small (27cm^3) superficial grade 2 undifferentiated pleomorphic sarcoma completely excised following 2 resections. In line with current guidelines (Dangoor et al. 2016) radiotherapy was not initially utilised in this case, although was administered following re-resection of the locally recurrent disease. A notable proportion (up to 15%) of cases of STS local recurrence are diagnosed >5 years following treatment (Toulmonde et al. 2014). As a result, the low rate of local recurrence seen in the cohort compared with other published series may be explained by the relatively short follow up period (mean 73 weeks).

5.6.4 Total Cell free DNA levels

The mean intraoperative cfDNA levels in the patient cohort was 10.5ng/ml although wide variation was seen between patients (range 3.8-63.8ng/ml). Data on total cfDNA levels of non-metastatic cancer patients with localised disease amenable to curative surgery is relatively rare compared to patients with advanced disease. The only other previous report of cfDNA levels in a non-metastatic STS patient was published in 2017 by Namlos et al who identified cfDNA levels of 110ng/ml in a patient with a large (878cm^3) STS one day prior to undergoing a hemipelvectomy (Namløs et al. 2017). Data is similarly rare for other malignancies, although reported levels of 8.0-344ng/ml (non-small cell lung cancer (Szpechcinski et al. 2015; Sozzi et al. 2001)), 0.5-235ng/ml (stage 1-2 breast cancer (Tangvarasittichai et al. 2015)), 30.1ng/ml (colorectal cancer (Mead et al. 2011)) and 59ng/ml (pancreatic cancer (Singh et al. 2015)) mean that overall our data suggests that non-metastatic STS patients may have lower cfDNA levels than other cancer patients.

No correlation was identified between intraoperative cfDNA levels and tumour size or grade. This may be explained by the absence of significant levels of ctDNA, the presence of ctDNA shed at a rate proportional to tumour characteristics other than size or grade (for example metabolic activity or growth rate) or variation in the ctDNA characteristics of individual STS subtypes in the cohort. Similarly no significant change was seen in total cfDNA levels between pre and post-operative samples. In the presence of significant levels of ctDNA, a logical hypothesis would be that overall cfDNA levels would fall considerably following surgery, especially considering the rapid clearance of cfDNA from the circulation (Fatouros et al. 2010; Lo et al. 1999). As a result the limited drop in mean post-operative cfDNA levels seen (only 1.9ng.ml) may reflect an absence of intraoperative ctDNA. Alternatively the relatively long intervals present between many of the patients' surgeries and the collection of their first post-operative plasma samples (mean 28 days) may have masked more significant drops in cfDNA present in the immediate post-operative period. Of note, patient 031's cfDNA levels rose markedly following surgery (5.8 to 30ng/ml). In the absence of an obvious clinical cause for this (for example infection) this most likely due to sample contamination due to lysed white blood cell DNA. Although excluding this individual's data increased the average drop seen in cfDNA following surgery to 2.5ng/ml, this change is was still lower than previously reported in other malignancies (35ng/ml in colorectal cancer for example (Cassinotti et al. 2013)) although variation in follow up regimes make direct comparisons between studies difficult.

5.6.4.1 Prognostic value of perioperative cfDNA levels

A retrospective comparison of those patients that suffered disease recurrence and those that remained disease free throughout follow up revealed no difference in intra- or postoperative cfDNA levels. Similarly no difference was seen in the magnitude of the drop in cfDNA seen following surgery, arguing against a potential role for using perioperative cfDNA levels as a predictive tool for STS relapse. The intra- and postoperative cfDNA levels of the 2 patients (24/31) that died during follow up were 4/5.8ng/ml and 2/30ng/ml. These relatively low intraoperative cfDNA levels also offer no suggestion that cfDNA quantification during surgery will provide any useful prognostic information for STS patients.

5.6.4.2 Diagnostic value of perioperative cfDNA levels

The absence of any difference in the post-operative cfDNA levels of those patents that underwent wide and marginal resections suggests that cfDNA quantification is not a suitable way to diagnose minimal residual disease following surgery. This is fairly unsurprising considering that various non-tumour related factors may alter cfDNA levels perioperatively, and the great accuracy with which a histopathologist can already characterise a resected specimen's margins.

Plasma samples were collected from 7 patients at the point they were diagnosed with STS recurrence. In these cases cfDNA levels were an average of just 1.5ng/ml higher than in the first plasma samples collected post-operatively. However if patient 031's data is excluded this difference increased to 4.0ng/ml, and cfDNA levels at the point of recurrence become significantly higher than the levels seen at the patients'1st post-operative outpatient appointments ($p=0.05$, paired t-test). However the relatively small magnitude of the rises seen in these cfDNA levels still suggests that extra tumoural sources of cfDNA may complicate the process of using cfDNA levels to diagnose STS recurrence alone.

5.6.4.3 Cell free DNA levels and (neo)adjuvant therapy: Rises in urinary cfDNA levels in rats exposed to ionising radiation (Abdullaev et al. 2015) and plasma cfDNA levels in humans undergoing radiotherapy have been reported (Cheng et al. 2009), presumably due to increased cell death. Unfortunately variation in the timings of plasma collection from our patients in relation to adjuvant radiotherapy preclude any statements on the effects of radiotherapy on cfDNA levels (or ctDNA characteristics) in our cohort. These effects must be investigated and considered in any future studies that propose any prognostic or diagnostic roles for cfDNA quantification. Obviously the effects of chemotherapy administration can also not be commented on based on only the 2 patients analysed in our cohort that received systemic chemotherapy.

5.6.5 Circulating tumour derived DNA characteristics:

Using a patient specific approach to profile SNVs using ddPCR and a tNGS SNV panel to compare patients' tumour and plasma samples evidence of ctDNA was identified in 5 /17 patients (29%) analysed - 3/8 cases (38%) analysed using ddPCR and 2/9 (22%) using tNGS. Although this figure is lower than reported in other early stage malignancies (55%,

(Bettegowda et al. 2014)) it is pleasing that these patients had a range of STS subtypes of a variety of grades and sizes. In addition evidence of ctDNA was identified at various points throughout these patients' management/follow up, perhaps most excitingly well before the detection of disease recurrence in patients 006 and 022 (up to >35 weeks). Taken together these cases highlight the kinds of scenarios in which ctDNA analysis has the potential to provide novel markers of micrometastatic STS disease, but also help to molecularly profile STSs longitudinally.

In 2 patients the SNV panel detected variants in the plasma that were not detected in the matched primary STS DNA. The apparent absence of these circulating variants from the resected STSs may be explained in several ways. Firstly, the variants may have been shed from tissues unrelated to the patients' STSs. Given our decision to define ctDNA as the presence of circulating COSMIC registered mutations in known cancer associated genes this seems unlikely, although the presence of an undiagnosed pre-malignant condition not detectable on patients' staging imaging cannot be ruled out. Secondly the circulating variants may have arisen from non-tumoural tissues adjacent to patients' tumours, close enough to be within the field of (neo)adjuvant radiotherapy but not so close to be resected during surgery. Thirdly this neoadjuvant radiotherapy may also have resulted in tumoural DNA damage in the region of the circulating mutations, resulting in inadequate amplification during the processes of ddPCR/tNGS library preparation. Finally the presence of significant intra-tumoural heterogeneity (ITH) in the resected STSs may have led to the subclonal populations containing the circulating mutations being missed during the process of STS tissue collection. If this is true, the detection of these circulating variants highlights eloquently how ctDNA can overcome the issues associated with ITH, although to confirm this multiregional sequencing would be required.

5.6.5.1 Prognostic value of ctDNA profiling:

Generally there was an inability to predict patients' clinical outcome based on their ctDNA profiles. This is highlighted by the low predictive values of the presence or absence of intra- or post-operative ctDNA. Overall intraoperative ctDNA was identified in 3/16 patients analysed not lost to follow up. Two of these cases recurred during follow up giving the presence of intraoperative ctDNA a positive predictive value (PPV) for recurrence of 66%. Intraoperative ctDNA was not detected in 13 of the patients analysed not lost to follow up. Of these, 7 remained disease free throughout follow up, giving the

absence of intraoperative ctDNA a negative predictive (NPV) for recurrence of 54%. Postoperative ctDNA was identified in 4 patients overall. Two of these individuals went onto suffer recurrence giving the presence of post-operative ctDNA at any stage during follow up a PPV for recurrence of 50%. Post-operative ctDNA was not detected in 12 patients. Six of these cases remained disease free giving the absence of post-operative ctDNA a NPP for recurrence of 50%.

In terms of overall survival intra- or post-operative ctDNA was only identified in one of the 2 patients that died during follow up. Clearly however the pilot nature of this analysis prevents any conclusions on this relationship being drawn, necessitating further work.

5.6.5.2 Diagnostic value of ctDNA profiling

The absence of any detectable ctDNA at recurrence in 7/8 patients that recurred during follow up suggests that ctDNA profiling performed as outlined here is not sensitive enough to diagnose STS recurrence. The absence of detectable ctDNA in these patients may reflect a true absence of ctDNA at recurrence, or alternatively a lack of significant cell death of those subclones containing the SNVs selected to profile ctDNA in these patients. In those patients analysed using ddPCR this may result from clonal evolution phasing the selected SNVs out of any metastatic lesions present (which may also explain the drop seen in circulating *TP53* (17:7577022 C>T) in patient 6 during follow up (Figure 5.16)) whilst for those patients analysed using tNGS may result from those SNVs key to the process of tumour recurrence not being covered by the Ampliseq^{TM} panel utilised.

5.7 Summary

The work outlined in this chapter has for the first time performed a longitudinal study of the total cfDNA and ctDNA characteristics of a cohort of non-metastatic STS patients. Overall this pilot work has failed to highlight a clear role for monitoring cfDNA levels in non-metastatic STS patients. However, it should be noted that following the exclusion of one outlier patient's data a significant difference was seen between the analysed recurrent patients' cfDNA levels at their 1st post-operative follow up, and the point at which their recurrence was diagnosed, warranting further work. By successfully identifying ctDNA in 2 'disease free' patients many weeks before the development of radiologically detectable disease the work presented has also highlighted a potential role

for ctDNA profiling. Despite this, it has also highlighted several significant challenges that must be addressed before ctDNA analysis can be proposed as realistic sources for novel STSs markers. The absence of evidence of ctDNA in 71% of the patients analysed must be recognised. If accurate this suggest that ctDNA is not shed by every STS subtype, or indeed by every patient with the same type of STS. An alternate explanation for this finding is that ctDNA was present in these cases, but was not identified by our experimental approach. If this interpretation is accepted, this highlights another key challenge to using ctDNA to monitor STS patients - determining which SNVs should be used to profile ctDNA in a groups of tumours as heterogeneous and genetically complex as STSs. Another potential issue highlighted by our analysis is the low levels that ctDNA was found to circulate at in those STS patients in which it was identified (VAF<3%). Moving forwards the practical ramifications of this are that future studies must enrol enough patients so that even subtle changes in ctDNA levels can be identified over background sequencing noise, and correlated with disease progression with confidence.

5.8 Future work

Despite the findings outlined in this chapter pilot nature of this study combined with the analysed patients' relatively short, often varied follow up regimes prevents any concrete conclusion being drawn. Although plans are in place to continue following up our cohort, moving forwards this necessitates larger longitudinal studies using standardised follow up regimes in relation to patients' surgeries and any (neo)adjuvant therapies used. These should also aim to follow up patients for >3 years to capture the majority of cases of recurrence, and continue to monitor these cases throughout any systemic treatment to allow correlation between cfDNA/ctDNA characteristics and treatment response. Given the genomic diversity seen between STS subtypes these studies should ideally ensure that enough patients are enrolled to allow different STS subtypes to be analysed independently, accepting that this would likely equate to over 1000 participants. In our pilot group ctDNA was detected in a similar proportion of patients using ddPCR and tNGS. Despite this, the heterogeneous nature of STSs combined with the absence of any SNVs found consistently across STSs subtypes suggests the ability of an Ampliseq^{TM} panel to simultaneously investigate many (up to 360 (Costa et al. 2017)) mutations may be most successful when attempting to track ctDNA in STS patients. Alternatively, multiple ddPCR assays could be used on the same patients, although the cost ramifications of this would be significant, particularly if ddPCR is supplemented by WES

to aid variant selection. Finally, a significant proportion of STSs are characterised by large gains and/or losses of chromosomal regions resulting in significant copy number alterations. In certain cancers, changes in the copy number of genes with direct therapeutic ramifications have successfully been tracked in ctDNA (Page et al. 2017), and consequently any future work investigating ctDNA in STS patients should ensure copy number alterations, as well as more simple point mutations, are targeted in the plasma.

Chapter 6. Concluding remarks:

The work outlined in this thesis was performed to further understand the genotype of STSs, and to start to explore the potential role of circulating nucleic acids as a novel source of STS biomarkers. A by-product of this work has been the creation of new pathways within the East Midlands that now facilitate the on-going collection and storage of tissue and blood samples from newly diagnosed STS patients. This biobanking process is key for research involving rare tumour such as STSs to continue, both locally and nationally /internationally.

Our Illumina WES analysis provided us with high quality accurate SNV/indel data, validated by correlation with simultaneous Sanger Sequencing and ddPCR analysis. As expected this data revealed a relatively low mutational load in the analysed STSs, with no significant relationship between load and the use of neoadjuvant treatment or specific STS subtypes. Many different genes were found to contain SNVs/indels in the analysed STSs. As well as several genes that have previously been identified as commonly mutated in STSs (*TP53, ATRX* and *RB1*) several other genes not formerly recognised as regularly containing to SNVs/indels in STSs were also highlighted (*MUC2, MUC4* and *FRG1B.*) Interestingly only *PARP4* was identified as a potential driver in our cohort based on its SNV/indel characteristics, which to our knowledge this is a novel finding.

We next used the same raw WES data to characterise the somatic CNAs present in the analysed STSs, validating our findings by correlating them with a simultaneous analysis performed using Nanostring technology. This work identified several genomic regions with an abnormal copy in a high proportion of the analysed tumours including 7p22.1 (amplified in 21/31cases) 17p11.2 (amplified in 20/31 cases), 20q13.33 (amplified in 20/31 cases) and 1q42.2 (deleted in 20/31 cases). This is in contrast to our SNV/indel analysis, which revealed no recurrent point mutations in cohort, and suggests that grouping STSs by their somatic CNAs may provide a more clinically useful way to stratify the tumours compared to the analysis of their SNV/indel mutational signatures alone.

The production, publication and collation of high quality sequencing data is key to developing a better understanding of the process of sarcomagenesis, as well as developing our understanding of the genomic characteristics of individual STS subtypes. The work outlined above has also highlighted several specific areas for future work.

Although *MUC2*, *MUC4* and *FRG1B* only contained SNVs/indels in 13% of the cases analysed, work to 1) investigate the function of these genes and 2) to investigate the effects of their knockdown is required. Work should also be encouraged to investigate how detrimental *PARP4* mutations may provide an oncogenic effect, given the gene's propensity to mutate in our cohort. Following up on our copy number analysis work is also required to determine the role of those genes found in each of the commonly amplified/lost genomic regions identified in the analysed STSs. This will allow any genomic drivers of sarcomagenesis that may be present to be identified, and thereafter the potential for these drivers as a source of novel diagnostic markers or targets for novel treatments to be deduced.

To investigate the potential of circulating nucleic acids as a source of STS biomarkers we characterised total cfDNA levels and circulating ctDNA levels in 2 groups of STS patients.

First, samples from a group of metastatic STS patients were analysed, based on the hypothesis that the plasma of those individuals with disseminated STS tissue would have a relatively high probability of containing ctDNA (assuming that shedding cfDNA is a characteristic of STSs.) As predicted total cfDNA levels in these patients was significantly higher than in a healthy control group simultaneously analysed. Regardless of the cause for this difference, the high levels of cfDNA identified highlights total cfDNA as a potential diagnostic marker for metastatic STSs - a statement supported by the correlation also identified between cfDNA levels and disease burden in the same patients. Using an Ion AmpliSeq™ panel designed based on a bioinformatics search of the literature we next sought evidence of ctDNA in the same metastatic STS patients. This experimental approach revealed evidence of ctDNA in 36% of the patients analysed, confirming that the tumoural shedding of cfDNA is a characteristic of metastatic lesions of at least certain STS subtypes.

To further investigate the positive findings outlined above we next performed a longitudinal analysis of the circulating nucleic acid characteristics of a cohort of non-metastatic STS patients undergoing attempted curative surgery. Interestingly, following the exclusion of one outlier patent's data, the analysis of total cfDNA levels in those patients that recurred during follow up revealed that cfDNA levels were significantly higher at the point recurrence was diagnosed than at the same patients 1st post-operative

follow up appointments. This suggests that total cfDNA levels may also hold a role as a future diagnostic marker in non-metastatic patients, although clearly further work is required to investigate this. Recognising our inability to detect ctDNA in the majority of the metastatic patients analysed, we elected to alter our experimental approach when attempting to characterise ctDNA in non-metastatic patients. We did this in 2 ways – 1) by designing and using a second Ion AmpliSeq™ panel targeting the hotspot regions of only the 3 most commonly mutated genes in STSs and 2) by designing and using multiple patient/mutation specific ddPCR assays created on the basis of the SNV illumina analysis outlined in chapter 3. Using these new assays we successfully identified ctDNA in 29% of the non-metastatic patients analysed at various points throughout follow up, confirming that ctDNA is also released from a wide range of primary STS subtypes. Particularly excitingly was our ability to identify ctDNA well before the radiological detection of disease recurrence in several of the patients analysed, highlighting how in practice the analysis of ctDNA has the potential to provide novel markers of micrometastatic STS disease in apparently 'disease free' STS patients.

Despite the positive findings outlined above, our analysis of total cfDNA and ctDNA in both the metastatic and non-metastatic patients failed to identify any relationships between patients' circulating nucleic acid characteristics and their recurrence or survival. Based on our small sample this argues against a role for either total cfDNA levels or ctDNA characteristics as a source of future prognostic markers for STS patients. A further disappointing feature of our data is our inability to identify ctDNA in the majority of the metastatic and non-metastatic patient analysed. This either suggests that 1) not all STSs shed cfDNA or that 2) the assays we employed were too specific / not sensitive enough to detect any ctDNA present. Regardless of which explanation is accurate, before ctDNA can be realistically proposed as a source of novel diagnostic STS biomarkers further work is needed to refine our ability to use ctDNA profiling to both discriminate between metastatic STS patients and healthy individuals, and to consistently identify STS recurrence.

Despite the challenges outlined above future work into the topic of STS patient cfDNA/ctDNA should not be disparaged. Based on our analyses we would encourage work with the aim of 1) further investigating the significant relationships we have identified between total cfDNA levels and certain clinical STS characteristics, and 2) determining if our inability to detect ctDNA in the majority of the STS patients analysed

can be solved with an alternate experimental approach. One issue that will need to be addressed in this latter work is the genomic diversity of STSs, and the difficulties this creates when selecting which mutations to track in patient ctDNA. Moving forwards multifocal tumour sequencing combined with the characterisation of multiple alterations in patient plasma (SNVs or CNAs) is likely to be needed to overcome this issue.

SUPPLEMENTARY DATA

See attached CD.

BIBLIOGRAPHY

Abbosh, C. et al., 2017. Phylogenetic ctDNA analysis depicts early stage lung cancer evolution. *Nature*, 545(7655), pp.446–451.

Abdullaev, S.A. et al., 2015. Cell-free DNA in the urine of rats exposed to ionizing radiation. *Radiation and Environmental Biophysics*, 54(3), pp.297–304.

Abeshouse, A. et al., 2017. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell*, 171(4), pp.950–965.

Adorno, M. et al., 2009. A Mutant-p53/Smad Complex Opposes p63 to Empower TGFβ-Induced Metastasis. *Cell*, 137(1), pp.87–98.

Agostini, M. et al., 2011. Circulating cell-free DNA: A promising marker of pathologic tumor response in rectal cancer patients receiving preoperative chemoradiotherapy. *Annals of Surgical Oncology*, 18(9), pp.2461–2468.

Ahmad, S.S. et al., 2012. Advances in radiotherapy. *BMJ (Online)*, 345(7886).

Ahmed, I. & Hamacher, K.L., 2002. Angiosarcoma in a chronically immunosuppressed renal transplant recipient: report of a case and review of the literature. *The American Journal of dermatopathology*, 24(4), pp.330–335.

Ahn, S. et al., 2016. A nCounter CNV Assay to Detect HER2 Amplification: A Correlation Study with Immunohistochemistry and In Situ Hybridization in Advanced Gastric Cancer. *Molecular Diagnosis and Therapy*, 20(4), pp.375–383.

Ai, B. et al., 2016. Circulating cell-free DNA as a prognostic and predictive biomarker in non-small cell lung cancer. *Oncotarget*, 7(28), pp.44583–44595.

Alexandrov, L.B. et al., 2013. Signatures of mutational processes in human cancer. *Nature*, 500(7463), pp.415–421.

Aogi, K. et al., 2000. Telomerase activity in soft-tissue and bone sarcomas. *Clinical Cancer Research*, 6(12), pp.4776–4781.

Artandi, S.E. & Attardi, L.D., 2005. Pathways connecting telomeres and p53 in senescence, apoptosis, and cancer. *Biochemical and Biophysical Research Communications*, 331(3), pp.881–890.

Ayoubi, H.A., Mahjoubi, F. & Mirzaei, R., 2017. Investigation of the human H3.3B (H3F3B) gene expression as a novel marker in patients with colorectal cancer. *Journal of Gastrointestinal Oncology*, 8(1), pp.64–69.

Bagaria, S.P. et al., 2014. Compliance with national comprehensive cancer network guidelines in the use of radiation therapy for extremity and superficial trunk soft tissue sarcoma in the United States. *Journal of Surgical Oncology*, 109(7), pp.633–638.

Bailey, S.M., Goodwin, E.H. & Cornforth, M.N., 2004. Strand-specific fluorescence in situ hybridization: The CO-FISH family. *Cytogenetic and Genome Research*, 107(1–2), pp.14–17.

Ballinger, M.L. et al., 2016. Monogenic and polygenic determinants of sarcoma risk: an international genetic study. *The Lancet Oncology*, 17(9), pp.1261–1271.

Barkess, G. et al., 2012. The chromatin-binding protein HMGN3 stimulates histone acetylation and transcription across the *Glyt1* gene. *Biochemical Journal*, 442(3), pp.495–505.

Barretina, J. et al., 2010. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. *Nature Genetics*, 42(8), pp.715–721.

Barrett, A.N. et al., 2014. Stability of cell-free DNA from maternal plasma isolated following a single centrifugation step. *Prenatal Diagnosis*, 34(13), pp.1283–1288.

Batzer, M.A. & Deininger, P.L., 2002. Alu repeats and human genomic diversity. *Nature Reviews Genetics*, 3(5), pp.370–379.

Baumann, C., Viveiros, M.M. & De La Fuente, R., 2010. Loss of maternal ATRX results in centromere instability and aneuploidy in the mammalian oocyte and Pre-implantation embryo. *PLoS Genetics*, 6(9), e1001137.

Beadling, C. et al., 2011. Multiplex Mutation screening by mass spectrometry: Evaluation of 820 cases from a personalized cancer medicine registry. *Journal of Molecular Diagnostics*, 13(5), pp.504–513.

Beck, A.H. et al., 2010. Gene expression profiling for the investigation of soft tissue sarcoma pathogenesis and the identification of diagnostic, prognostic, and predictive biomarkers. *Virchows Arch*, 456(2): pp.141–151.

Beckingsale, T.B. & Shaw, C., 2017. Epidemiology of bone and soft-tissue sarcomas. *Orthopaedics and Trauma*, 31(3), pp.144–150.

Behjati, S. et al., 2013. Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nature Genetics*, 45(12), pp.1479–1482.

De Berardinis, R.J. & Chandel, N.S., 2016. Fundamentals of cancer metabolism. *Science Advances*, 2(5), e1600200.

Berger, A.W. et al., 2017. Treatment monitoring in metastatic colorectal cancer patients by quantification and KRAS genotyping of circulating cell-free DNA. *PLoS ONE*, 12(3), e0174308.

Bergmann, O. et al., 2015. Dynamics of Cell Generation and Turnover in the Human Heart. *Cell*, 161(7), pp.1566–1575.

Best, B.P., 2009. Nuclear DNA Damage as a Direct Cause of Aging. *Rejuvenation Research*, 12(3), pp.199–208.

Bettegowda, C. et al., 2014. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci Transl Med*, 6(224), pp.224ra24.

Bid, H.K. et al., 2013. RAC1: an emerging therapeutic option for targeting cancer angiogenesis and metastasis. *Molecular cancer therapeutics*, 12(10), pp.1925–34.

Boltze, C. et al., 2001. Distinction between lipoma and liposarcoma by MDM2 alterations: a case report of simultaneously occurring tumors and review of the literature. *Pathology, research and practice*, 197(8), pp.563–8.

Bork, P. et al., 1997. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *The FASEB Journal*, 11(1), pp.68–76.

Boskovic, T. et al., 2014. Pneumothorax after transthoracic needle biopsy of lung lesions under CT guidance. *Journal of Thoracic Disease*, 6(Suppl 1), S99–S107.

Both, J. et al., 2012. Identification of novel candidate oncogenes in chromosome region 17p11.2-p12 in human osteosarcoma. *PLoS One.*, 7(1), e30907.

Both, J. et al., 2017. Oncogenic Properties of Candidate Oncogenes in Chromosome Region 17p11.2p12 in Human Osteosarcoma. *Cytogenetic and Genome Research*, 150(1), pp.52–59.

Broberg, K. et al., 2009. Association between polymorphisms in RMI1, TOP3A, and BLM and risk of cancer, a case-control study. *BMC Cancer*, 11(9), pp.140.

Brouns, F., Stas, M. & De Wever, I., 2003. Delay in diagnosis of soft tissue sarcomas. *Eur J Surg Oncol*, 29(5), pp.440–445.

Bryan, T.M. et al., 1997. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nature Medicine*, 3(11), pp.1271–1274.

Bryan, T.M. et al., 1995. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J*, 14(17), pp.4240–4248.

Burningham, Z. et al., 2012. The Epidemiology of Sarcoma. *Clinical Sarcoma Research*, 2(1), p.14.

Bush, K.M. et al., 2013. Endogenous mammalian histone H3.3 exhibits chromatin-related functions during development. *Epigenetics and Chromatin*, 6(1), p.7.

Butler, T.M. et al., 2015. Exome sequencing of cell-free DNA from metastatic cancer patients identifies clinically actionable mutations distinct from primary disease. *PLoS ONE*, 10(8), e0136407.

Cassinotti, E. et al., 2013. Free circulating DNA as a biomarker of colorectal cancer. *International Journal of Surgery*, 11(Suppl 1), S54-7.

Champoux, J.J., 2001. DNA topoisomerases: structure, function, and mechanism. *Annual review of biochemistry*, 70, pp.369–413.

Chan, K.M. et al., 2013. The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Genes and Development*, 27(9), pp.985–990.

Chang, L. et al., 2015. Circulating tumor cells in sarcomas: a brief review. *Medical oncology (Northwood, London, England)*, 32(1), p.430.

Chen, K.-Z. et al., 2016. Circulating Tumor DNA Detection in Early-Stage Non-Small Cell Lung Cancer Patients by Targeted Sequencing. *Scientific Reports*, 24(6), p.31985.

Cheng, C. et al., 2009. Quantification of circulating cell-free DNA in the plasma of cancer patients during radiation therapy. *Cancer Science*, 100(2), pp.303–309.

Chinen, L.T. et al., 2014. Isolation, Detection, And immunomorphological characterization of circulating tumor cells (CTCs) from patients with different types of sarcoma using isolation by size of tumor cells: A window on sarcoma-cell invasion. *OncoTargets and Therapy*, 16(7), pp.1609–1617.

Chinnam, M. & Goodrich, D.W., 2011. RB1, Development, and Cancer. *Current Topics in Developmental Biology*, 94, pp.129–169.

Coombs, C.C. et al., 2017. Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell*, 21(3), pp.374–382.

Corey, R.M., Swett, K. & Ward, W.G., 2014. Epidemiology and survivorship of soft tissue sarcomas in adults: A national cancer database report. *Cancer Medicine*, 3(5), pp.1404–1415.

Cormier, J.N. & Pollock, R.E., 2004. Soft tissue sarcomas. *CA Cancer J Clinicians*, 54(2), pp.94–109.

Correa, H., 2016. Li-Fraumeni Syndrome. *Journal of pediatric genetics*, 5(2), pp.84–88.

Costa, A. et al., 2006. Telomere maintenance mechanisms in liposarcomas: Association with histologic subtypes and disease progression. *Cancer Research*, 66(17), pp.8918–8924.

Costa, J.L. et al., 2017. Multi institutional evaluation of a new NGS assay for mutation detection from cfDNA in lung cancer. *Cancer Research*, 77(Suppl 13), p.5694.

Crew, A.J. et al., 1995. Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with homology to the Kruppel-associated box in human synovial sarcoma. *The EMBO journal*, 14(10), pp.2333–40.

Crozat, A. et al., 1993. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature*, 363(6430), pp.640–644.

Dangoor, A. et al., 2016. UK guidelines for the management of soft tissue sarcomas. *Clinical Sarcoma Research*, 6(1), p.20.

Danila, D.C. et al., 2011. Circulating tumors cells as biomarkers: Progress toward biomarker qualification. *Cancer Journal*, 17(6), pp.438–450.

Daniotti, M. et al., 2007. Detection of mutated BRAFV600E variant in circulating DNA of stage III-IV melanoma patients. *International Journal of Cancer*, 120(11), pp.2439–2444.

Darwanto, A. et al., 2017. Use of the QIAGEN GeneReader NGS system for detection of KRAS mutations, validated by the QIAGEN Therascreen PCR kit and alternative NGS platform. *BMC Cancer*, 17(1), p.358.

Davicioni, E. et al., 2006. Identification of a PAX-FKHR gene expression signature that defines molecular classes and determines the prognosis of alveolar rhabdomyosarcomas. *Cancer Res*, 66(14):pp.6936–6946.

Dawson, S.-J. et al., 2013. Analysis of Circulating Tumor DNA to Monitor Metastatic Breast Cancer. *New England Journal of Medicine*, 368(13), pp.1199–1209.

Demetri, G.D. et al., 2016. Efficacy and safety of trabectedin or dacarbazine for metastatic liposarcoma or leiomyosarcoma after failure of conventional chemotherapy: Results of a phase III randomized multicenter clinical trial. *Journal of Clinical Oncology*, 34(8), pp.786–793.

Demoulin, JB. & Montano-Almendras CP., 2012. Platelet-derived growth factors and their receptors in normal and malignant hematopoiesis. *Am J Blood Res*, 2(1), pp.44-56.

Depla, A.L. et al., 2014. Treatment and prognostic factors of radiation-associated angiosarcoma (RAAS) after primary breast cancer: A systematic review. *European Journal of Cancer*, 50(10), pp.1779–1788.

Diaz, L.A. et al., 2012. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature*, 486(7404), pp.537–540.

van Dijk, E.L. et al., 2014. Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), pp.418–426.

Djaafri, I. et al., 2014. A novel tumor suppressor function of Kindlin-3 in solid cancer. *Oncotarget*, 5(19), pp.8970–8985.

Do, H. & Dobrovic, A., 2012. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil-DNA glycosylase. *Oncotarget*, 3(5), pp.546–558.

Dowler Nygaard, A. et al., 2014. Levels of cell-free DNA and plasma KRAS during treatment of advanced NSCLC. *Oncology reports*, 31(2), pp.969–74.

Dumontet, C. & Jordan, M.A., 2010. Microtubule-binding agents: A dynamic field of cancer therapeutics. *Nature Reviews Drug Discovery*, 9(10), pp.790–803.

Edge, S. et al., 2010. AJCC Cancer Staging Manual (7th ed). In *Springer*. p. 347–76.

Eisenhauer, E.A. et al., 2009. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2), pp.228–247.

Enneking, W.F., Spanier, S.S. & Goodman, M.A., 2003. A system for the surgical staging of musculoskeletal sarcoma. 1980. *Clinical orthopaedics and related research*, pp.4–18.

Ernst, A. et al., 2016. Telomere dysfunction and chromothripsis. *International Journal of Cancer*, 138(12), pp.2905–2914.

Estourgie, S.H., Nielsen, G.P. & Ott, M.J., 2002. Metastatic patterns of extremity myxoid liposarcoma and their outcome. *Journal of Surgical Oncology*, 80(2), pp.89–93.

Etemadmoghadam, D. et al., 2013. Resistance to CDK2 inhibitors is associated with selection of polyploid cells in CCNE1-amplified ovarian cancer. *Clinical Cancer Research*, 19(21), pp.5960–5971.

Fan, G. et al., 2017. Prognostic value of circulating tumor DNA in patients with colon cancer: Systematic review. *PLoS ONE*, 12(2), e0171991.

Farooqi, A.A. & Siddik, Z.H., 2015. Platelet-derived growth factor (PDGF) signalling in cancer: Rapidly emerging signalling landscape. *Cell Biochemistry and Function*, 33(5), pp.257–265.

Fatouros, I.G. et al., 2010. Time of sampling is crucial for measurement of cell-free plasma DNA following acute aseptic inflammation induced by exercise. *Clinical Biochemistry*, 43(16–17), pp.1368–1370.

Favero, F. et al., 2015. Sequenza: Allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of Oncology*, 26(1), pp.64–70.

Fawzy, A. et al., 2016. Quantitative analysis of plasma cell-free DNA and its DNA integrity in patients with metastatic prostate cancer using ALU sequence. *Journal of the Egyptian National Cancer Institute*, 28(4), pp.235–242.

Feng, G. et al., 2013. Quantification of plasma cell-free DNA in predicting therapeutic efficacy of sorafenib on metastatic clear cell renal cell carcinoma. *Disease Markers*, 34(2), pp.105–111.

Ferrari, A. et al., 2007. Soft-tissue sarcomas in children and adolescents with neurofibromatosis type 1. *Cancer*, 109(7), pp.1406–1412.

Fisher, R., Pusztai, L. & Swanton, C., 2013. Cancer heterogeneity: Implications for targeted therapeutics. *Br J Cancer*, 108(3), pp.479–485.

Fletcher, C., World Health Organization & International Agency for Research on Cancer, 2013. *WHO classification of tumours of soft tissue and bone*, p.468

Fong, S.L. et al., 2009. Comparison of 7 Methods for Extracting Cell-Free DNA from Serum Samples of Colorectal Cancer Patients. *Clinical Chemistry*, 55(3), pp.587–9.

Fournié, G.J. et al., 1995. Plasma DNA as a marker of cancerous cell death. Investigations in patients suffering from lung cancer and in nude mice bearing human tumours. *Cancer Letters*, 91(2), pp.221–227.

Fujiwara, Y. et al., 1999. Plasma DNA microsatellites as tumor-specific markers and indicators of tumor progression in melanoma patients. *Cancer Research*, 59(7), pp.1567–1571.

Furlanut, M. & Franceschi, L., 2003. Pharmacology of ifosfamide. *Oncology*, 65(Suppl 2), pp.2–6.

Furusawa, T. & Cherukuri, S., 2010. Developmental function of HMGN proteins. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1799(1–2), pp.69–73.

Furuta, E. et al., 2010. Metabolic genes in cancer: their roles in tumor progression and clinical implications. *Biochim Biophys Acta*, 1805(2), pp.141–152.

Gadalla, S.E. et al., 2013. EpCAM associates with endoplasmic reticulum aminopeptidase 2 (ERAP2) in breast cancer cells. *Biochemical and Biophysical Research Communications*, 439(2), pp.203–208.

Gaillard, H., García-Muse, T. & Aguilera, A., 2015. Replication stress and cancer. *Nature Reviews Cancer*, 15(5), pp.276–280.

Gerashchenko, T.S. et al., 2013. Intratumor heterogeneity: Nature and biological significance. *Biochemistry (Moscow)*, 78(11), pp.1201–1215.

Gerrand, C.H. et al., 2001. Classification of positive margins after resection of soft-tissue sarcoma of the limb predicts the risk of local recurrence. *The Journal of bone and joint surgery. British volume*, 83(8), pp.1149–1155.

Giannakis, M. et al., 2016. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Reports*, 15(4), pp.857–865.

Gielis, E.M. et al., 2015. Cell-Free DNA: An Upcoming Biomarker in Transplantation. *American Journal of Transplantation*, 15(10), pp.2541–2551.

Giglia-Mari, G., Zotter, A. & Vermeulen, W., 2011. DNA damage response. *Cold Spring Harbor Perspectives in Biology*, 3(1), pp.1–19.

Giubellino, A., Burke, T.R. & Bottaro, D.P., 2008. Grb2 signaling in cell motility and cancer. *Expert opinion on therapeutic targets*, 12(8), pp.1021–33.

Gögenur, M., Burcharth, J. & Gögenur, I., 2017. The role of total cell-free DNA in predicting outcomes among trauma patients in the intensive care unit: A systematic review. *Critical Care*, 21(14).

Goh, J.Y. et al., 2017. Chromosome 1q21.3 amplification is a trackable biomarker and actionable target for breast cancer recurrence. *Nature Medicine*, 23(11), pp.1319–1330.

Goodrich, D.W., 2006. The retinoblastoma tumor-suppressor gene, the exception that proves the rule. *Oncogene*, 25(38), pp.5233–5243.

van der Graaf, W.T.A. et al., 2012. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): a randomised, double-blind, placebo-controlled phase 3 trial. *Lancet*, 379(9829), pp.1879–1886.

Grady, W.M. et al., 2001. Detection of aberrantly methylated hMLH1 promoter DNA in the serum of patients with microsatellite unstable colon cancer. *Cancer Research*, 61(3), pp.900–902.

Greenman, C. et al., 2007. Patterns of somatic mutation in human cancer genomes. *Nature*, 446(7132), pp.153–158.

Grimer, R.J., 2006. Size matters for sarcomas! *Annals of the Royal College of Surgeons of England*, 88(6), pp.519–524.

Grimer, R.J. & Briggs, T.W.R., 2010. Earlier diagnosis of bone and soft-tissue tumours. *Journal of Bone and Joint Surgery - British Volume*, 92(11), pp.1489–1492.

Grossi, V. et al., 2015. Loss of STK11 expression is an early event in prostate carcinogenesis and predicts therapeutic response to targeted therapy against MAPK/p38. *Autophagy*, 11(11), pp.2102–2113.

Guillou, L. & Aurias, A., 2010. Soft tissue sarcomas with complex genomic profiles. *Virchows Archiv*, 456(2), pp.201–217.

Hakin-Smith, V. et al., 2003. Alternative lengthening of telomeres and survival in patients with glioblastoma multiforme. *Lancet*, 361(9360), pp.836–838.

Hanahan, D. & Weinberg, R.A., 2011. Hallmarks of cancer: The next generation. *Cell*, 144(5), pp.646–674.

Hayflick, L. & Moorhead, P.S., 1961. The serial cultivation of human diploid cell strains. *Experimental Cell Research*, 25(3), pp.585–621.

Hayward, N.K. et al., 2017. Whole-genome landscapes of major melanoma subtypes. *Nature*, 545(7653), pp.175–180.

Heaphy et al., 2011. Altered telomeres in tumors with ATRX and DAXX mutations. *Science*, 333(6041), p.425.

Heaphy, C.M. et al., 2011. Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. *American Journal of Pathology*, 179(4), pp.1608–1615.

Heidenreich, B. et al., 2014. TERT promoter mutations in cancer development. *Current Opinion in Genetics and Development*, 24(1), pp.30–37.

Heinrich, M.C. et al., 2003. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *Journal of Clinical Oncology*, 21(23), pp.4342–4349.

Heldin, CH. & Lennartsson, J. 2013. Structural and functional properties of platelet-derived growth factor and stem cell factor receptors. *Cold Spring Harb Perspect Biol*, 5(8), a009100

Henderson, E.R. & Blackburn, E.H., 1989. An overhanging 3' terminus is a conserved feature of telomeres. *Molecular and cellular biology*, 9(1), pp.345–8.

Hendy, O.M. et al., 2016. Circulating cell free DNA as a predictor of systemic lupus erythematosus severity and monitoring of therapy. *Egyptian Journal of Medical Human Genetics*, 17(1), pp.79–85.

Henson, J.D. et al., 2005. A robust assay for alternative lengthening of telomeres in tumors shows the significance of alternative lengthening of telomeres in sarcomas and astrocytomas. *Clinical Cancer Research*, 11(1), pp.217–225.

Henson, J.D. et al., 2009. DNA C-circles are specific and quantifiable markers of alternative- lengthening-of-telomeres activity. *Nature Biotechnology*, 27(12), pp.1181–1185.

Hientz, K. et al., 2017. The role of p53 in cancer drug resistance and targeted chemotherapy. *Oncotarget*, 8(5), pp.8921–8946.

Hoeijmakers, J.H.J., 2001. Genome maintenance mechanisms for preventing cancer. *Nature*, 411(6835), pp.366–374.

Hong, B. & Zu, Y., 2013. Detecting circulating tumor cells: Current challenges and new trends. *Theranostics*, 3(6), pp.377–394.

Hovgaard, T.B. et al., 2017. Follow-up after initial surgical treatment of soft tissue sarcomas in the extremities and trunk wall. *Acta Oncologica*, 56(7), pp.1004–1012.

Hsieh, M.C. et al., 2013. Racial and Ethnic Disparities in the Incidence and Trends of Soft Tissue Sarcoma Among Adolescents and Young Adults in the United States, 1995-2008. *J Adolesc Young Adult Oncol*, 2(3), pp.89–94.

Huang, A. et al., 2016. Detecting Circulating Tumor DNA in Hepatocellular Carcinoma Patients Using Droplet Digital PCR Is Feasible and Reflects Intratumoral Heterogeneity. *Journal of Cancer*, 7(13), pp.1907–1914.

Huang, H.-T. et al., 2014. Loss of function of SWI/SNF chromatin remodeling genes leads to genome instability of human lung cancer. *Oncology Reports*, 33(1), pp.283–291.

Hucthagowder, V. et al., 2012. Resequencing analysis of the candidate tyrosine kinase and RAS pathway gene families in multiple myeloma. *Cancer Genet*, Sep;205(9), pp.474–8.

Hughes, S. et al., 2006. The use of whole genome amplification to study chromosomal changes in prostate cancer: Insights into genome-wide signature of preneoplasia associated with cancer progression. *BMC Genomics*, 7(65).

Iacono, M. Lo et al., 2016. Retrospective study testing next generation sequencing of selected cancer-associated genes in resected prostate cancer. *Oncotarget*, 7(12), pp.1–11.

Iengar, P., 2012. An analysis of substitution, deletion and insertion mutations in cancer genes. *Nucleic Acids Research*, 40(14), pp.6401–6413.

Ikeda, Y. et al., 2016. Germline PARP4 mutations in patients with primary thyroid and breast cancers. *Endocrine-Related Cancer*, 23(3), pp.171–179.

Illingworth, R.S. et al., 2010. Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome. *PLoS Genetics*, 6(9), e1001134.

Iyevleva, A.G. & Imyanitov, E.N., 2016. Cytotoxic and targeted therapy for hereditary cancers. *Hereditary Cancer in Clinical Practice*, 14(1), p.17.

Jahr, S. et al., 2001. DNA fragments in the blood plasma of cancer patients: Quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Research*, 61(4), pp.1659–1665.

Jaques, D.P. et al., 1995. Hepatic metastases from soft-tissue sarcoma. *Annals of surgery*, 221(4), pp.392–397.

Jeanes, A., Gottardi, C.J. & Yap, A.S., 2008. Cadherins and cancer: How does cadherin dysfunction promote tumor progression? *Oncogene*, 27(55), pp.6920–6929.

Jeyapalan, J.N. et al., 2008. Evidence for alternative lengthening of telomeres in liposarcomas in the absence of ALT-associated PML bodies. *International Journal of Cancer*, 122(11), pp.2414–2421.

Johnson, C.J., Pynsent, P.B. & Grimer, R.J., 2001. Clinical features of soft tissue sarcomas. *Annals of the Royal College of Surgeons of England*, 83(3), pp.203–5.

Johnson, J.E. et al., 2005. Multiple mechanisms of telomere maintenance exist in liposarcomas. *Clinical Cancer Research*, 11(15), pp.5347–5355.

Judson, I. et al., 2018. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: a randomised controlled phase 3 trial. *The Lancet Oncology*, 15(4), pp.415–423.

Judson, I., 2010. Targeted therapies in soft tissue sarcomas. *Annals of oncology*, 21(Suppl 7), pp.277–80.

Kaisaki, P.J. et al., 2016. Targeted next-generation sequencing of plasma DNA from cancer patients: Factors influencing consistency with tumour DNA and prospective investigation of its utility for diagnosis. *PLoS ONE*, 11(9), e0162809.

Katsonis, P. et al., 2014. Single nucleotide variations: Biological impact and theoretical interpretation. *Protein Science*, 23(12), pp.1650–1666.

Kawakami, K. et al., 2000. Hypermethylated APC DNA in plasma and prognosis of patients with esophageal adenocarcinoma. *Journal of the National Cancer Institute*, 92(22), pp.1805–11.

Kidess, E. & Jeffrey, S., 2013. Circulating tumor cells versus tumor-derived cell-free DNA: rivals or partners in cancer care in the era of single-cell analysis? *Genome Med*, 5(8), p.70.

Kim, N. et al., 2013. Somatic Mutaome Profile in Human Cancer Tissues. *Genomics & Informatics*, 11(4), p.239.

Kleinerman, R.A. et al., 2007. Risk of soft tissue sarcomas by individual subtype in survivors of hereditary retinoblastoma. *Journal of the National Cancer Institute*, 99(1), pp.24–31.

Koelsche, C. et al., 2016. Differential nuclear ATRX expression in sarcomas. *Histopathology*, 68(5), pp.738–745.

Koelsche, C. et al., 2017. Histone 3.3 hotspot mutations in conventional osteosarcomas: a comprehensive clinical and molecular characterization of six H3F3A mutated cases. *Clinical Sarcoma Research*, 7(9).

Koffler, D. et al., 1973. The occurrence of single-stranded DNA in the serum of patients with systemic lupus erythematosus and other diseases. *The Journal of clinical investigation*, 52(1), pp.198–204.

Kopreski, M.S. et al., 1997. Detection of mutant K-ras DNA in plasma or serum of patients with colorectal cancer. *British Journal of Cancer*, 76(10), pp.1293–1299.

Kopreski, M.S. et al., 2000. Somatic mutation screening: identification of individuals harboring K-ras mutations with the use of plasma DNA. *Journal of the National Cancer Institute*, 92(11), pp.918–23.

Koschmann, C. et al., 2016. ATRX loss promotes tumor growth and impairs nonhomologous end joining DNA repair in glioma. *Science Translational Medicine*, 8(328), p.328ra28.

Kufe, D.W., 2013. MUC1-C oncoprotein as a target in breast cancer: Activation of signaling pathways and therapeutic approaches. *Oncogene*, 32(9), pp.1073–1081.

Kufe, D.W., 2009. Mucins in cancer: Function, prognosis and therapy. *Nature Reviews Cancer*, 9(12), pp.874–885.

Langerød, A. et al., 2007. TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Res*, 9(3), p.R30.

Larson, D.E. et al., 2012. Somaticsniper: Identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*, 28(3), pp.311–317.

Lauer, N.K. et al., 2002. Absence of telomerase activity in malignant bone tumors and soft-tissue sarcomas. *Sarcoma*, 6(1), pp.43–6.

Lee, J.-C. et al., 2015. Alternative lengthening of telomeres and loss of ATRX are frequent events in pleomorphic and dedifferentiated liposarcomas. *Modern Pathology*, 28, pp.1064–107367.

Lee, M. et al., 2014. Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. *Nucleic Acids Research*, 42(3), pp.1733–1746.

Lee, R.J. et al., 2018. Circulating tumor DNA predicts survival in patients with resected high-risk stage II/III melanoma. *Annals of Oncology*, 29(2), pp.490–496.

Lehnhardt, M. et al., 2009. Importance of specialized centers in diagnosis and treatment of extremity-soft tissue sarcomas. Review of 603 cases. *Chirurg*, 80(4), pp.341–7.

Leon, S.A. et al., 1977. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Research*, 37(3), pp.646–650.

Lewis, P.W. et al., 2013. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science*, 340(6134), pp.857–861.

Li, B.T. et al., 2016. A prospective study of total plasma cell-free DNA as a predictive biomarker for response to systemic therapy in patients with advanced non-small-cell lung cancers. *Annals of oncology*, 27(1), pp.154–9.

Li, X.L. et al., 2017. Integrin β4 promotes cell invasion and epithelial-mesenchymal transition through the modulation of Slug expression in hepatocellular carcinoma. *Scientific Reports*, 7(40464).

Liau et al., 2015. Comprehensive screening of alternative lengthening of telomeres phenotype and loss of ATRX expression in sarcomas. *Modern pathology*, 28(12), pp.1545–54.

Liau, J.Y. et al., 2015. Alternative lengthening of telomeres phenotype in malignant vascular tumors is highly associated with loss of ATRX expression and is frequently observed in hepatic angiosarcomas. *Human Pathology*, 46(9), pp.1360–1366.

Lin, C.-J., Conti, M. & Ramalho-Santos, M., 2013. Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development. *Development*, 140(17), pp.3624–3634.

Lin, M. et al., 2017. Effects of short indels on protein structure and function in human genomes. *Scientific Reports*, 7(9313).

Liu, Q. et al., 2016. C-terminus of MUC16 activates Wnt signaling pathway through its interaction with β-catenin to promote tumorigenesis and metastasis. *Oncotarget*, 7(24), pp.36800–36813.

Lo, Y.M. et al., 1999. Rapid clearance of fetal DNA from maternal plasma. *American journal of human genetics*, 64(1), pp.218–24.

López-Guerrero, J.A. et al., 2004. Deregulation of the G1 to S-phase cell cycle checkpoint is involved in the pathogenesis of human osteosarcoma. *Diagnostic molecular pathology*, 13(2), pp.81–91.

Lovejoy, C.A. et al., 2012. Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of Telomeres pathway. *PLoS Genetics*, 8(7), e1002772.

Lundblad, V. & Szostak, J.W., 1989. A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell*, 57(4), pp.633–643.

Maciejowski, J. et al., 2015. Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell*, 163(7), pp.1641–1654.

Maki, R.G. et al., 2007. Randomized Phase II Study of Gemcitabine and Docetaxel Compared With Gemcitabine Alone in Patients With Metastatic Soft Tissue Sarcomas: Results of Sarcoma Alliance for Research Through Collaboration Study 002. *Journal of Clinical Oncology*, 25(19), pp.2755–2763.

Manzat Saplacan, R.M. et al., 2017. The Role of PDGFs and PDGFRs in Colorectal Cancer. *Mediators Inflamm*.2017:4708076.

Mao, Z. et al., 2014. DNA repair by nonhomologous end joining and homologous recombination during cell cycle in human cells. *Cell Cycle*, 7(18), pp.2902–2906.

May, W.A. et al., 1993. The Ewing's sarcoma EWS/FLI-1 fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than FLI-1. *Molecular and cellular biology*, 13(12), pp.7393–8.

Mayakonda, A. & Koeffler, H.P., 2016. Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. *bioRxiv*, p.52662.

McIntyre, J.B. et al., 2017. Molecular alterations in indolent, aggressive and recurrent ovarian low-grade serous carcinoma. *Histopathology*, 70(3), pp.347–358.

Mead, R. et al., 2011. Circulating tumour markers can define patients with normal colons, benign polyps, and cancers. *British Journal of Cancer*, 105(2), pp.239–245.

Mermel, C.H. et al., 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biology*, 12(4), p.R41.

Mertens, F. et al., 2006. Prognostic significance of chromosome aberrations in high-grade soft tissue sarcomas. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 24(2), pp.315–20.

Mertens, F. et al., 2009. Translocation-Related Sarcomas. *Seminars in Oncology*, 36(4), pp.312–323.

Mertens, F., Panagopoulos, I. & Mandahl, N., 2010. Genomic characteristics of soft tissue sarcomas. *Virchows Archiv*, 456(2), pp.129–139.

Mészáros, B. et al., 2016. Systematic analysis of somatic mutations driving cancer: Uncovering functional protein regions in disease development. *Biology Direct*, 11(1).

Meyerson, M., 2000. Role of telomerase in normal and cancer cells. *Journal of Clinical Oncology*, 18(13), pp.2626–2634.

Min, L. et al., 2016. The roles and implications of exosomes in sarcoma. *Cancer and Metastasis Reviews*, 35(3), pp.377–390.

Mini, E. et al., 2006. Cellular pharmacology of gemcitabine. *Annals of Oncology*, 17(Suppl 5).

Mir, O. et al., 2015. Activity of regorafenib (RE) in leiomyosarcomas (LMS) and other types of soft-tissue sarcomas (OTS): Results of a double-blind, randomized placebo (PL) controlled phase II trial. *Journal of Clinical Oncology*, 33(Suppl 15), p.10504.

Mistry, M. et al., 2011. Cancer incidence in the United Kingdom: Projections to the year 2030. *British Journal of Cancer*, 105(11), pp.1795–1803.

Moore, L.D., Le, T. & Fan, G., 2013. DNA methylation and its basic function. *Neuropsychopharmacology*, 38(1), pp.23–38.

Morales, J. et al., 2014. Review of Poly (ADP-ribose) Polymerase (PARP) Mechanisms of Action and Rationale for Targeting in Cancer and Other Diseases. *Critical Reviews in Eukaryotic Gene Expression*, 24(1), pp.15–28.

Morbelli, S. et al., 2017. Circulating Tumor DNA Reflects Tumor Metabolism Rather Than Tumor Burden in Chemotherapy-Naive Patients with Advanced Non-Small Cell Lung Cancer: 18F-FDG PET/CT Study. *Journal of nuclear medicine*, 58(11), pp.1764–1769.

Mosieniak, G. & Sikora, E., 2010. Polyploidy: the link between senescence and cancer. *Current pharmaceutical design*, 16(6), pp.734–740.

Mouliere, F. et al., 2011. High fragmentation characterizes tumour-derived circulating DNA. *PLoS ONE*, 6(9), e23418.

Movva, S. et al., 2015. Multi-platform profiling of over 2000 sarcomas: identification of biomarkers and novel therapeutic targets. *Oncotarget*, 6(14), pp.12234–47.

Muntoni, A. et al., 2009. Telomere elongation involves intra-molecular DNA replication in cells utilizing alternative lengthening of telomeres. *Human Molecular Genetics*, 18(6), pp.1017–1027.

Murnane, J.P. et al., 1994. Telomere dynamics in an immortal human cell line. *The EMBO journal*, 13(20), pp.4953–62.

Nagata, S. et al., 2003. Degradation of chromosomal DNA during apoptosis. *Cell Death and Differentiation*, 10(1), pp.108–116.

Nakamura, T. et al., 2009. Lung radiofrequency ablation in patients with pulmonary metastases from musculoskeletal sarcomas. *Cancer*, 115(16), pp.3774–3781.

Nakauchi, C. et al., 2016. Detection of TP53/PIK3CA Mutations in Cell-Free Plasma DNA From Metastatic Breast Cancer Patients Using Next Generation Sequencing. *Clinical Breast Cancer*, 16(5), pp.418–423.

Namløs, H.M. et al., 2017. Use of liquid biopsies to monitor disease progression in a sarcoma patient: A case report. *BMC Cancer*, 17(1), p.29.

Napier, C.E. et al., 2015. ATRX represses alternative lengthening of telomeres. *Oncotarget*, 6(18), pp.16543–16558.

Nath, S. & Mukherjee, P., 2014. MUC1: A multifaceted oncoprotein with a key role in cancer progression. *Trends in Molecular Medicine*, 20(6), pp.332–342.

Nathenson, M.J. & Sausville, E., 2016. Looking for answers: the current status of neoadjuvant treatment in localized soft tissue sarcomas. *Cancer Chemotherapy and Pharmacology*, 78(5), pp.895–919.

National Cancer Institute, 2018. The Cancer Genome Atlas. *2018*. Available at: www.cbioportal.org.

O'Donnell, P.W. et al., 2014. The effect of the setting of a positive surgical margin in soft tissue sarcoma. *Cancer*, 120(18), pp.2866–2875.

O'Sullivan, B. et al., 2002. Preoperative versus postoperative radiotherapy in soft-tissue sarcoma of the limbs: A randomised trial. *Lancet*, 359(9325), pp.2235–2241.

Ogose, A. et al., 1999. Tumors of peripheral nerves: Correlation of symptoms, clinical signs, imaging features, and histologic diagnosis. *Skeletal Radiology*, 28(4), pp.183–188.

Olivier, M., Hollstein, M. & Hainaut, P., 2010. TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harbor perspectives in biology*, 2(1), a001008.

Olovnikov, A.M., 1973. A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *Journal of Theoretical Biology*, 41(1), pp.181–190.

Van Oosterom, A.T. et al., 2002. Results of randomised studies of the EORTC Soft Tissue and Bone Sarcoma Group (STBSG) with two different ifosfamide regimens in first- and second-line chemotherapy in advanced soft tissue sarcoma patients. *European Journal of Cancer*, 38(18), pp.2397–2406.

Page, K. et al., 2013. Influence of Plasma Processing on Recovery and Analysis of Circulating Nucleic Acids. *PLoS ONE*, 8(10), e77963.

Page, K. et al., 2017. Next generation sequencing of circulating cell-free DNA for evaluating mutations and gene amplification in metastatic breast cancer. *Clinical Chemistry*, 63(2), pp.532–541.

Palm, W. & de Lange, T., 2008. How Shelterin Protects Mammalian Telomeres. *Annual Review of Genetics*, 42(1), pp.301–334.

Panaro, N.J. et al., 2000. Evaluation of DNA fragment sizing and quantification by the Agilent 2100 bioanalyzer. *Clinical Chemistry*, 46(11), pp.1851–1853.

Pastore, G. et al., 2006. Childhood soft tissue sarcomas incidence and survival in European children (1978-1997): Report from the Automated Childhood Cancer Information System project. *European Journal of Cancer*, 42(13), pp.2136–2149.

Pearce, M.S. et al., 2012. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: A retrospective cohort study. *The Lancet*, 380(9840), pp.499–505.

Pennington, K.P. et al., 2014. Germline and somatic mutations in homologous recombination genes predict platinum response and survival in ovarian, fallopian tube, and peritoneal carcinomas. *Clinical Cancer Research*, 20(3), pp.764–775.

Perdigones, N. & Murtaza, M., 2017. Capturing tumor heterogeneity and clonal evolution in solid cancers using circulating tumor DNA analysis. *Pharmacology and Therapeutics*, 174, pp.22–26.

Pickett, H.A. et al., 2009. Control of telomere length by a trimming mechanism that involves generation of t-circles. *EMBO Journal*, 28(7), pp.799–809.

Presneau, N. et al., 2005. Identification of novel variant, 1484delG in the 3'UTR of H3F3B, a member of the histone 3B replacement family, in ovarian tumors. *International journal of oncology*, 26(6), pp.1621–1627.

Public health England, 2011. Soft Tissue Sarcomas: incidence and survival rates in England. Available at: http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/sarcomas/.

Quail, M.A. et al., 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 24(13), p.341.

Rabinowits, G. et al., 2009. Exosomal microRNA: A diagnostic marker for lung cancer. *Clinical Lung Cancer*, 10(1), pp.42–46.

Ravi, V., Patel, S. & Benjamin, R., 2015. Chemotherapy for soft-tissue sarcomas. *Oncology (Williston Park).*, 29(1), pp.43–50.

Reiman, A. et al., 2017. Validation of an NGS mutation detection panel for melanoma. *BMC Cancer*, 17(1), p.150.

Rhodes, A. & Cecconi, M., 2012. Cell-free DNA and outcome in sepsis. *Critical Care*, 16(6), p.170.

Riva, F. et al., 2017. Patient-Specific Circulating Tumor DNA Detection during Neoadjuvant Chemotherapy in Triple-Negative Breast Cancer. *Clinical Chemistry*, 63(3), pp.691–699.

Ross, A.A. et al., 1993. Detection and viability of tumor cells in peripheral blood stem cell collections from breast cancer patients using immunocytochemical and clonogenic assay techniques. *Blood*, 82(9), pp.2605–2610.

Rothermundt, C. et al., 2014. What is the role of routine follow-up for localised limb soft tissue sarcomas? A retrospective analysis of 174 patients. *British Journal of Cancer*, 110(10), pp.2420–2426.

Rubin, A.F. & Green, P., 2009. Mutation patterns in cancer genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), pp.21766–70.

Sabolch, A. et al., 2012. Risk factors for local recurrence and metastasis in soft tissue sarcomas of the extremity. *American journal of clinical oncology*, 35(2), pp.151–7.

Saiki, A.Y. et al., 2015. Identifying the determinants of response to MDM2 inhibition. *Oncotarget*, 6(10), pp.7701–7712.

Sangar, F. et al., 2014. Involvement of small ArfGAP1 (SMAP1), a novel Arf6-specific GTPase-activating protein, in microsatellite instability oncogenesis. *Oncogene*, 33(21), pp.2758–2767.

Santarpia, L. et al., 2012. Mutation profiling identifies numerous rare drug targets and distinct mutation patterns in different clinical subtypes of breast cancers. *Breast Cancer Research and Treatment*, 134(1), pp.333–343.

Satelli, A. et al., 2014. Universal marker and detection tool for human sarcoma circulating tumor cells. *Cancer Research*, 74(6), pp.1645–1650.

Saunders, C.T. et al., 2012. Strelka: Accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*, 28(14), pp.1811–1817.

Schneider-Stock, R. et al., 1999. Prognostic significance of telomerase activity in soft tissue sarcomas. *Int J Oncol*, 15(4), pp.775–780.

Schöffski, P. et al., 2014. Soft tissue sarcoma: An update on systemic treatment options for patients with advanced disease. *Oncology Research and Treatment*, 37(6), pp.355–362.

Schwartzentruber, J. et al., 2012. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*, 482(7384), pp.226–231.

Seddon, B. et al., 2017. Gemcitabine and docetaxel versus doxorubicin as first-line treatment in previously untreated advanced unresectable or metastatic soft-tissue sarcomas (GeDDiS): a randomised controlled phase 3 trial. *The Lancet Oncology*, 18(10), pp.1397–1410.

Shao, Z.M. et al., 2001. p53 mutation in plasma DNA and its prognostic value in breast cancer patients 3892. *Clinical Cancer Research*, 7(8), pp.2222–2227.

Sharma, A. & Schwartz, R.A., 2012. Stewart-Treves syndrome: Pathogenesis and management. *Journal of the American Academy of Dermatology*, 67(6), pp.1342–1348.

Sharma, S., Kelly, T.K. & Jones, P.A., 2009. Epigenetics in cancer. *Carcinogenesis*, 31(1), pp.27–36.

Sharpe, H.J. et al., 2015. Genomic Analysis of Smoothened Inhibitor Resistance in Basal Cell Carcinoma. *Cancer Cell*, 27(3), pp.327–341.

Simbolo, M. et al., 2013. DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples. *PLoS ONE*, 8(6), e62692.

Singh, N. et al., 2015. High levels of cell-free circulating nucleic acids in pancreatic cancer are associated with vascular encasement, metastasis and poor survival. *Cancer Investigation*, 33(3), pp.78–85.

Smith, H.G. et al., 2018. Major Amputations for Extremity Soft-Tissue Sarcoma. *Annals of Surgical Oncology*, 25(2), pp.387–393.

Soave, A. et al., 2017. Copy number variations of circulating, cell-free DNA in urothelial carcinoma of the bladder patients treated with radical cystectomy: a prospective study. *Oncotarget*, 8(34), pp.56398–56407.

Sobinoff, A.P. et al., 2017. BLM and SLX4 play opposing roles in recombination-dependent replication at human telomeres. *The EMBO Journal*, 26(19), pp.2907–2919.

Sorber, L. et al., 2017. A Comparison of Cell-Free DNA Isolation Kits: Isolation and Quantification of Cell-Free DNA in Plasma. *Journal of Molecular Diagnostics*, 19(1), pp.162–168.

Sorenson, G.D., 2000. Detection of mutated KRAS2 sequences as tumor markers in plasma/serum of patients with gastrointestinal cancer. *Clinical Cancer Research*, 6(6), pp.2129–2137.

Sorenson, G.D. et al., 1994. Soluble Normal and Mutated DNA Sequences from Single-Copy Genes in Human Blood. *Cancer Epidemiology Biomarkers and Prevention*, 3(1), pp.67–71.

Sozzi, G. et al., 2001. Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients. *Cancer Research*, 61(12), pp.4675–4678.

Sozzi, G. et al., 1999. Detection of microsatellite alterations in plasma DNA of non-small cell lung cancer patients: a prospect for early diagnosis. *Clin Cancer Res*, 5(10), pp.2689–2692.

Spindler, K.L.G. et al., 2015. Circulating free DNA as biomarker and source for mutation detection in metastatic colorectal cancer. *PLoS ONE*, 10(4), e0108247.

Stephens, P.J. et al., 2011. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 144(1), pp.27–40.

Steppan, D., Pratilas, C. & Loeb, D., 2017. Targeted therapy for soft tissue sarcomas in adolescents and young adults. *Adolescent Health, Medicine and Therapeutics*, 8, pp.41–55.

Stölzel, F. et al., 2016. Karyotype complexity and prognosis in acute myeloid leukemia. *Blood Cancer Journal*, 6(1), e386.

Stroun, M. et al., 2001. About the possible origin and mechanism of circulating DNA: Apoptosis and active DNA release. In *Clinica Chimica Acta*. pp.139–142.

Subhawong, A.P. et al., 2009. The alternative lengthening of telomeres phenotype in breast carcinoma is associated with HER-2 overexpression. *Modern pathology*, 22(11), pp.1423–1431.

Suzuki, Y. et al., 2017. Multiregion ultra-deep sequencing reveals early intermixing and variable levels of intratumoral heterogeneity in colorectal cancer. *Molecular oncology*, 11(2), pp.124–139.

Szpechcinski, A. et al., 2015. Cell-free DNA levels in plasma of patients with non-small-cell lung cancer and inflammatory lung disease. *British journal of cancer*, 113(3), pp.476–483.

Tacar, O., Sriamornsak, P. & Dass, C.R., 2013. Doxorubicin: An update on anticancer molecular action, toxicity and novel drug delivery systems. *Journal of Pharmacy and Pharmacology*, 65(2), pp.157–170.

Tamkovich, S.N. et al., 2016. Circulating nucleic acids in blood of healthy male and female donors. *Clin.Chem.*, 51(7), pp.1317–1319.

Tangvarasittichai, O., Jaiwang, W. & Tangvarasittichai, S., 2015. The Plasma DNA Concentration as a Potential Breast Cancer Screening Marker. *Indian Journal of Clinical Biochemistry*, 30(1), pp.55–58.

Tarpey, P.S. et al., 2013. Frequent mutation of the major cartilage collagen gene COL2A1 in chondrosarcoma. *Nature Genetics*, 45(8), pp.923–926.

Taylor, D.D. & Gercel-Taylor, C., 2008. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecologic Oncology*, 110(1), pp.13–21.

Tellez-Gabriel, M. et al., 2016. The Challenges of Detecting Circulating Tumor Cells in Sarcoma. *Frontiers in Oncology*, 6, p.202.

Thierry, A.R. et al., 2016. Origins, structures, and functions of circulating DNA in oncology. *Cancer and Metastasis Reviews*, 35(3), pp.347–376.

Thomas, D.M. & Ballinger, M.L., 2015. Etiologic, environmental and inherited risk factors in sarcomas. *Journal of Surgical Oncology*, 111(5), pp.490–495.

Tian, D. et al., 2008. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature*, 455(7209), pp.105–108.

Tierney, J.F., 1997. Adjuvant chemotherapy for localised resectable soft-tissue sarcoma of adults: Meta-analysis of individual data. *Lancet*, 350(9092), pp.1647–1654.

Tomasetti, C., Vogelstein, B. & Parmigiani, G., 2013. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proceedings of the National Academy of Sciences*, 110(6), pp.1999–2004.

Totomochika, S. et al., 2010. Increased serum cell-free DNA levels in relation to inflammation are predictive of distant metastasis of esophageal squamous cell carcinoma. *Experimental and Therapeutic Medicine*, 1(1), pp.89–92.

Toulmonde, M. et al., 2014. Long-term recurrence of soft tissue sarcomas: Prognostic factors and implications for prolonged follow-up. *Cancer*, 120(19), pp.3003–3006.

Trojani, M. et al., 1984. Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. *International journal of cancer.*, 33(1), pp.37–42.

Trovik, C.S., 2001. Local recurrence of soft tissue sarcoma. A Scandinavian Sarcoma Group Project. *Acta orthop Scand Suppl*, 72(300), pp.1–31.

Tsumita, T. & Iwanaga, M., 1963. Fate of injected deoxyribonucleic acid in mice. *Nature*, 198(4885), pp.1088–1089.

Tuononen, K. et al., 2013. Comparison of Targeted Next-Generation Sequencing (NGS) and Real-Time PCR in the Detection of EGFR, KRAS, and BRAF Mutations on Formalin-Fixed, Paraffin-Embedded Tumor Material of Non-Small Cell Lung Carcinoma-Superiority of NGS. *Genes Chromosomes and Cancer*, 52(5), pp.503–511.

Underhill, H.R. et al., 2016. Fragment Length of Circulating Tumor DNA. *PLoS Genetics*, 12(7), e1006162.

Usadel, H. et al., 2002. Quantitative adenomatous polyposis coli promoter methylation analysis in tumor tissue, serum, and plasma DNA of patients with lung cancer. *Cancer Research*, 62(2), pp.371–375.

Uusitalo, E. et al., 2016. Distinctive cancer associations in patients with neurofibromatosis type 1. *Journal of Clinical Oncology*, 34(17), pp.1978–1986.

Varley, H. et al., 2002. Molecular characterization of inter-telomere and intra-telomere mutations in human ALT cells. *Nature genetics*, 30(3), pp.301–5.

Vasioukhin, V. et al., 1994. Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. *British Journal of Haematology*, 86(4), pp.774–779.

Velcich, A. et al., 2002. Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science*, 295(5560), pp.1726–1729.

Verhagen, P.C.M.S. et al., 2002. Deletion of chromosomal region 6q14-16 in prostate cancer. *International Journal of Cancer*, 102(2), pp.142–147.

Vocke, C.D. et al., 2017. Comprehensive genomic and phenotypic characterization of germline FH deletion in hereditary leiomyomatosis and renal cell carcinoma. *Genes Chromosomes and Cancer*, 56(6), pp.484–492.

Vogelstein, B. et al., 2013. Cancer genome landscapes. *Science*, 340(6127), pp.1546–1558.

Wang, B.G. et al., 2003. Increased plasma DNA integrity in cancer patients. *Cancer research*, 63(14), pp.3966–8.

Wang, Y. et al., 2004. Effect of the codon 72 polymorphism (c.215G>C, p.Arg72Pro) in combination with somatic sequence variants in the TP53 gene on survival in patients with advanced ovarian carcinoma. *Human Mutation*, 24(1), pp.21–34.

Ward, E. et al., 2001. Update of the follow-up of mortality and cancer incidence among European workers employed in the vinyl chloride industry. *Epidemiology*, 12(6), pp.710–718.

Ward, K. et al., 2015. Epithelial cell adhesion molecule is expressed in a subset of sarcomas and correlates to the degree of cytological atypia in leiomyosarcomas. *Molecular and clinical oncology*, 3(1), pp.31–36.

Wehler, T. C. et al., 2008. PDGFRα/β expression correlates with the metastatic behavior of human colorectal cancer: a possible rationale for a molecular targeting strategy. *Oncology Reports*, 19(3), pp.697–704.

Weitz, J., Antonescu, C.R. & Brennan, M.F., 2003. Localized extremity soft tissue sarcoma: Improved knowledge with unchanged survival over time. *Journal of Clinical Oncology*, 21(14), pp.2719–2725.

Woll, P.J. et al., 2012. Adjuvant chemotherapy with doxorubicin, ifosfamide, and lenograstim for resected soft-tissue sarcoma (EORTC 62931): A multicentre randomised controlled trial. *The Lancet Oncology*, 13(10), pp.1045–1054.

Wong, I.H.N. et al., 1999. Detection of aberrant p16 methylation in the plasma and serum of liver cancer patients. *Cancer Research*, 59(1), pp.71–73.

Xu, L., Li, S. & Stohr, B. a., 2013. The Role of Telomere Biology in Cancer. *Annual Review of Pathology*, 24(8), pp.49–78.

Yamada, T. et al., 1998. Detection of K-ras gene mutations in plasma DNA of patients with pancreatic adenocarcinoma: correlation with clinicopathological features. *Clinical cancer research*, 4(6), pp.1527–32.

Yang, X. et al., 2016. Quantification of mutant alleles in circulating tumor DNA can predict survival in lung cancer. *Oncotarget*, 7(15), pp.20810–24.

Yeager, T.R. et al., 1999. Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukemia (PML) body. *Cancer Research*, 59(17), pp.4175–4179.

Yi, Z. et al., 2017. Plasma cell-free DNA and survival in non-small-cell lung cancer: A meta-analysis. *Molecular and clinical oncology*, 7(2), pp.167–172.

Yoo, J. & Robinson, R.A., 2000. Expression of telomerase activity and telomerase RNA in human soft tissue sarcomas. *Archives of Pathology and Laboratory Medicine*, 124(3), pp.393–397.

Yoshikawa, H. et al., 1997. Skeletal metastases from soft-tissue sarcomas. Incidence, patterns, and radiological features. *The Journal of bone and joint surgery. British volume*, 79(4), pp.548–52.

Yuen, B.T.K. & Knoepfler, P.S., 2013. Histone H3.3 Mutations: A Variant Path to Cancer. *Cancer Cell*, 24(5), pp.567–574.

Zack, T.I. et al., 2013. Pan-cancer patterns of somatic copy number alteration. *Nature Genetics*, 45(10), pp.1134–1140.

Zagars et al., 2003. Prognostic factors for patients with localized soft-tissue sarcoma treated with conservation surgery and radiation therapy: An analysis of 1225 patients. *Cancer*, 97(10), pp.2530–2543.

Zagars, G.K. et al., 2003. Prognostic factors for disease-specific survival after first relapse of soft-tissue sarcoma: Analysis of 402 patients with disease relapse after initial conservative surgery and radiotherapy. *International Journal of Radiation Oncology Biology Physics*, 57(3), pp.739–747.

Zagars, G.K. et al., 2003. Surgical margins and reresection in the management of patients with soft tissue sarcoma using conservative surgery and radiation therapy. *Cancer*, 97(10), pp.2544–53.

Zaher, E. et al., 2013. Cell-free DNA concentration and integrity as a screening tool for cancer. *Indian Journal of Cancer*, 50(3), p.175.

Zare, F. et al., 2017. An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC Bioinformatics*, 18(1), p.286.

Zenonos, K., 2013. RAS signaling pathways, mutations and their role in colorectal cancer. *World Journal of Gastrointestinal Oncology*, 5(5), p.97.

Zhang, J. et al., 2015. Whole genome and transcriptome sequencing of matched primary and peritoneal metastatic gastric carcinoma. *Sci Rep*, 5, p.13750.

Zilfou, J.T. & Lowe, S.W., 2009. Tumor suppressive functions of p53. *Cold Spring Harbor perspectives in biology*, 1(5), a001883.