UNIVERSITY OF BIRMINGHAM

Identifying new driver genes in Well Differentiated Retroperitoneal Liposarcoma

by

Robert William Tyler

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Institute of Cancer and Genomic Sciences
College of Medical and Dental Sciences
March 2021

Abstract

Introduction

Well differentiated retroperitoneal liposarcoma (WDLS) has a high local recurrence rate, a poor response to neoadjuvant therapy and has the potential to become a high-grade dedifferentiated tumour (DDLS). Efforts to target the hallmark MDM2 and CDK4 amplification have been met with limited successes and led to efforts to target outside of this region. Preliminary work identified a possible somatic mutation (chr9: 68,303, 273A>T) in *FOXD4L3*. Secondly, a recurrent fusion within the TCA cycle gene *SDHA* was found. Both of these genes required further exploration, as well as a broader approach to identify new genes of interest.

Aims

Validate the presence of the somatic mutation (chr9: 68,303, 273A>T) in *FOXD4L3* in a larger patient cohort. Demonstrate *SDHA* fusion in vitro and then explore *SDHA* functionality in vivo, using metabolic tracing. Identify novel driver genes in WDLS through the combination of CRISPR and single-cell RNA sequencing.

Methods

Nanopore sequencing of WDLS tissue samples was performed. Sanger sequencing was performed on both WDLS samples, as well as healthy controls. *SDHA* expression was examined using an *SDHA* fusion assay and Western blotting. To explore the in vivo function of *SDHA*, $^{13}C_6$ metabolic tracing experiments were undertaken in four patients undergoing surgery for WDLS. To identify new targets in WDLS, a Genome wide CRISPR knockout (GeCKO) screen was performed on a

WDLS cell line (93T449). To complement this, total RNA sequencing on paired tumour/normal samples was performed as well as single-cell RNA sequencing (scRNA-seq).

Results

The previously assumed somatic mutation in FOXD4L3 (chr9: 68,303, 273A>T)) was identified as a single nucleotide variant – rs7034645. In WDLS, rs7034645 was more highly prevalent with an odds ratio of 5.3, classifying this as a high-risk germline variant. Elevated expression levels of *SDHA* were observed in normal tissue compared to tumour. Metabolic tracing experiments and RNA sequencing suggested an impairment of SDHA function in WDLS, although with no convincing evidence of a complete loss of function attributable to gene fusions. Significant levels of aspartate and succinate were found in tumours shedding new light on their metabolic phenotype. Both the Wnt and mTOR signalling pathways were consistently implicated in the GeCKO screen, total RNA sequencing and scRNA-seq, which merit further analysis.

Conclusion

Rs7034645 is the first risk variant described in WDLS. Fusions within *SDHA* do not appear responsible for tumourigenesis, but there is evidence for a downregulation of SDHA both in the transcriptome and metabolome. Furthermore, tracing experiments offer the first insights into the metabolome of WDLS with reliance on aspartate worthy of future investigation. The Wnt and mTOR signalling pathways are possible drivers in WDLS, which demand further inquiry.

Dedication

To my Dad, without whom I would never have embarked on this venture. You are missed every day.

To my wife Susan for supporting my career path every step of the way, with so much understanding, laughter and love.

Acknowledgements

I would like to acknowledge the help of my supervisor Professor Andrew Beggs, for his support and encouragement over the last three years. In particular for making time for my lists of questions, and the good-humoured conversations that resulted. I thank Mr. Anant Desai for creating this research post, supporting my academic and clinical endeavours and acting as a mentor.

Thanks to Dr. Dan Tennant for explaining the world of metabolomics to me, and ensuring the scientific support was there to run the SHARPS study. Alongside Dan, particular thanks to Dr. Paul Walker and Dr. Katarina Kluckova without whom the metabolomic analysis would not have been possible.

I will be forever grateful to the unsung stars of the Beggs Lab: Dr. Jo Stockton, Louise Tee, Dr. Valerie Somani, Claire Bryer, Anetta Ptasinska and Celina Whalley. Thanks to Jo for your limitless patience and help with nanopore and Sanger sequencing. One day I will buy you some nuclease-free water. Thanks to Lou for taking me step by step through an entire GeCKO screen, and splitting my cells when I went to Glastonbury. Without all of you, the results section of this thesis would be threadbare.

Additional thanks to my friend Dr. James Connolly, for providing invaluable advice over the course of the PhD, particularly with primer design and endless questions about thesis formatting.

Acknowledgement is given to Dr. Phillipe Taniere, who provided high quality samples and was generous with advice and support. I am also grateful to the rest of the sarcoma unit for supporting my scientific endeavours.

Funding from the Royal College of Surgeons in the form of a Clinical Research Fellowship is gratefully acknowledged.

Finally, gratitude is extended to all the patients who consented to donating blood and surplus tissue in the course of this research.

Table of Contents

Chapter 1: Introduction 1

1.1 Sarcoma 1

1.2 Liposarcoma 1

1.3 Retroperitoneal Tumours 3

1.4 Retroperitoneal Liposarcoma 4

1.5 Surgery for retroperitoneal liposarcoma 12

1.6 Non-operative interventions for retroperitoneal liposarcoma 14

1.7 Surveillance 19

1.8 Management of recurrent disease 20

1.9 Summary management of RPLS 21

1.10 Genomic alterations in liposarcoma 21

1.11 Novel therapies in retroperitoneal liposarcoma 35

1.12 FOXD4L3 42

1.13 SDHA 42

1.14 CRISPR and Single Cell RNA sequencing 43

1.15 Summary 44

1.16 Hypothesis 45

1.17 Aims and objectives 45

Chapter 2: Materials and Methods 46

2.1 Samples 46

2.2 Nucleic Acid Extraction 47

2.3 Nucleic Acid quantification 51

2.4 Cell culture 52

2.5 DNA sequencing 54

2.6 Protein detection by western blotting 60

2.7 SDHA colorimetric assay 65

2.8 SDHA fusion panel 66

2.9 SdHA fusion in RetroPeritoneal liposarcoma Study (SHARPS) 71

2.10 Gas Chromatography – Mass Spectrometry (GC-MS) 73

2.11 RNA sequencing 76

2.12 Genome Wide CRISPR-Cas9 Knockout screen (GeCKO) 82

2.13 Single cell RNA sequencing 90

2.14 Library Preparation for Next-generation sequencing (NGS) 97

2.15 Graphical Presentation and statistical analysis 99
Chapter 3: The role of FOXD4L3 in WDLS 100
3.1 Introduction 100
3.3 Results 104
3.4 Discussion 115
3.5 Conclusion 118
Chapter 4: The role of SDHA in WDLS 119
4.1 Introduction 119
4.2 Results 123
4.3 Discussion 139
4.4 Conclusion 144
Chapter 5: In vivo assessment of WDLS using metabolic tracing 145
5.1 Introduction 145
5.2 Results 150
5.3 Discussion 168
5.4 Conclusion 174
Chapter 6: Genome Wide CRISPR Knockout Screen 175
6.1 Introduction 175
6.2 Results 177
6.3 Discussion 189
6.4 Conclusion 196
Chapter 7: Transcriptomic landscape of WDLS 197
7.1 Introduction 197
7.2 Results 198
7.3 Discussion 217
7.4 Conclusion 225
Chapter 8 – General Discussion 227
8. 1 Role of FOXD4L3 in WDLS 227
8.2 SDHA in WDLS and TCA cycle dysregulation 228
8.3 Identifying new targets in WDLS 230
8.4 Limitations 233
8.5 Future Work 235
8.6 Conclusion 238
Appendix I : SHARPS patient information sheet and consent form 261
Appendix II 266

Table of Figures

Figure 1.1: Haemotoxylin and Eosin stains of liposarcoma 2
Figure 1.2: Axial cross section of abdominal cavity.. 5
Figure 1.3: Histologic and CT appearances of retroperitoneal liposarcoma. 9
Figure 1.4: Image of fluorescent in situ hybridisation (FISH) performed on WDLS cells 10
Figure 1.5: Pathways active in WDLS/DDLS 32
Figure 2.1: Clustal analysis to identify non-identical sequences in FOXD4L3 54
Figure 2.2: Sanger nested primers for chr9: 68,303, 273 55
Figure 2.3: Schematic overview of SHARPS study 76
Figure 2.4: 96 well plate set-up from 0 to 40,000 cells in triplicate. 83
Figure 2.5: GFP transfected cells at 20x depth (a) and 10x depth (b) 85
Figure 2.6: GeCKO library Day 0 (a) and Day 14 (b) profiles 89
Figure 2.7: WDLS cells after tumour dissociation. 91
Figure 2.8: Tape station profile of cDNA 94
Figure 2.9: Example tape station library prior to NGS 97
Figure 3.1: *FOXD4L3* gene on chromosome 5 (co-ordinates as per GRCh38). 101
Figure 3.2: FOXD4 paralogs 103
Figure 3.3: Electropherogram of reference genome. 110
Figure 3.4: Electropherogram showing a heterozygous SNP (rs7034645). 110
Figure 3.5: Electropherogram showing a homozygous call. 110
Figure 4.1: *FUS-DDIT3.* 121
Figure 4.2: Map of *SDHA* fusion partners 127
Figure 4.3: Western blots for SDHA and Actin in T1/N2 131
Figure 4.4: Western blots for Actin, SDHA and SDHB in one paired WDLS sample (T3/N4) 132
Figure 4.5: Western blots for Actin, SDHA and SDHB in one paired WDLS sample (5/6). Two protein quantities were loaded (20,40 µg). Ponceau stain included. 132
Figure 4.6: Western blots for Actin, SDHA and SDHB in 5 WDL only samples (7-11). 133
Figure 4.7: SDHA and SDHB expression normalised against Actin in tumour and matched normal 134
Figure 4.8: Line graph of *SDHA* activity 135
Figure 4.9: Bar chart of *SDHA* activity 136
Figure 5.1: TCA cycle. 146
Figure 5.2: Labelling patterns of different mass isotopomers 152
Figure 5.3: Normalised total succinate counts in ions/mg wet tissue mass 153
Figure 5.4: Normalised total fumarate counts in ions/mg wet tissue mass 154
Figure 5.5: Normalised total aspartate counts in ions/mg wet tissue mass 154
Figure 5.6: Ratios of Succinate to Fumarate Total Ion Count (TIC) 156
Figure 5.7: Ratios of Succinate to Aspartate Total Ion Count (TIC) 156
Figure 5.8: Blood plasma samples for patients 1-4 159
Figure 5.9: SHARPS samples 1-9 from tumour and normal tissue. Mass isotopomer distribution (MID) of glycolysis metabolites from $^{13}C_6$-glucose (%) 161
Figure 5.10: SHARPS samples 1-9 from tumour and normal tissue. Mass isotopomer distribution (MID) of TCA cycle intermediates and aspartate from $^{13}C_6$-glucose (%) 163
Figure 5.11: Ratio of M+2/M+3 fumarate isotopomers in samples 1-9. 165

Figure 5.12: SHARPS samples 1-9 from tumour and normal tissue. Mass isotopomer distribution (MID) of amino acids from $^{13}C_6$-glucose (%) 167
Figure 6.1: Cell viability assay plotting luminescence in different cell concentrations over time. 177
Figure 6.2: Cell titre glow luminescence at different puromycin concentrations 178
Figure 6.3: Sigmoidal curve of multiplicity of infection by lentiviral supernatant volume. 179
Figure 6.4: Schematic representation of GeCKO screen for WDLS cell line 180
Figure 6.5: Violin plot of all genes distributed by beta score. 183
Figure 6.6: A rank plot of most selected genes by beta score 184
Figure 6.7: Complex enrichment of negatively selected genes 185
Figure 6.8: Scatter plot dividing genes in four different groups based on the largest absolute value of the differential beta score 186
Figure 6.9: Complex enrichment of positively selected genes in Group 2 and negatively selected genes in Group 4. 188
Figure 7.1: t-SNE plot of WDLS sample 200
Figure 7.2: t-SNE plots by immune cell marker 205
Figure 7.3: Violin plots for *CDK4*, *MDM2*, *TSPAN31*, *JUN*, *HMGA2* and *FRS2*. X-axis split by Cluster 1-3 b) Heatmap of key liposarcoma genes separated by cluster 212
Figure 7.4: t-SNE plots per recognised liposarcoma gene 213

Table of Tables

Table 1.1: AJCC TNM staging for retroperitoneal soft tissue sarcoma (32)..... 11
Table 1.2: FNCLCC grading system, table taken from Archives of Pathology and Laboratory Medicine (33). HPF - high powered field.. 11
Table 1.3: Evidence summary of genomic drivers of WD/DDLS subtypes and related clinical trials 41
Table 2.1: Paraffin removal and rehydration settings 50
Table 2.2: List of primers for DNA sequencing 55
Table 2.3: Thermocycling conditions for PCR 56
Table 2.4: End Repair Preparation Master Mix 57
Table 2.5: PCR conditions for Sanger sequencing 60
Table 2.6: Composition of SDS-PAGE 61
Table 2.7: Antibodies used for Western blotting 63
Table 2.8: *SDHA* colorimetric 10 x assay buffer 65
Table 2.9: PCR conditions for SPE target enrichment 69
Table 2.10: PCR conditions for universal PCR amplification 70
Table 2.11: PCR conditions for amplification of adaptor ligated DNA 81
Table 2.12: PCR conditions for sgRNA amplification 88
Table 2.13: PCR conditions for cDNA amplification 93
Table 2.14: Sample Index PCR 96
Table 2.15: Flow cell type by experiment. 98
Table 3.1: Demographics and histology for patient cohort. 105
Table 3.2: Base calls (%) at Chr.9:68,303,273 for samples 1-12 with corresponding sequencing depth. Samples 7-11 were sequenced on a separate run. 106
Table 3.3: rs7034645 frequency as per gnomAD database. 107
Table 3.4: Sanger sequencing results using reverse primer for patient cohort 111
A total of 46 samples underwent Sanger sequencing from unaffected volunteers, the results of which are shown in Table 3.5. 112
Table 3.5: Sanger sequencing results using reverse primer for unaffected cohort 112
Table 3.6: Number and Frequency of rs7034645 genotypes across cohorts 113
Table 3.7: Number and Frequency of rs7034645 genotypes when combined with gnomAD figures 113
Table 3.8: Combined Sanger and nanopore sequencing results for Patients 1-6.. 114
Table 4.1: Summary of recognised fusion genes in soft tissue sarcomas. Adapted from Management of soft tissue sarcoma (188). 120
Table 4.2: Histology for *SDHA* cohort 124
Table 4.3: Significant fusions present in *SDHA* (p-Value <0.05) by sample ID 125
Table 4.4: Significant *SDHA* fusion partners 126
Table 4.5: Significant *SDHA* fusions in tumour samples. All chromosomal co-ordinates are for Chromosome 5. Break Point Location 1 gives the functional region covered by Start 1 to End 1. Break Point Location 2 gives the functional region covered by Start 2 to End 2. 128
Table 4.7: Significant gene fusions per sample 137
Table 5.1: Demographics and histology of SHARPS cohort 150
Table 5.2: Sampling times by SHARPS ID. Times are shown for three blood samples: prior to surgery, from start of operation to tumour removal and from tumour removal to snap frozen. 158
Table 6.1: Multiplicity of infection at different lentiviral volumes 178

Table 6.2: Most significant negatively selected genes in GeCKO screen. Protein function from OMIM database (248) 181
Table 7.1: Most significant genes in Cluster One selected by log fold change when compared to Clusters 2 and 3. 201
Table 7.2: Most highly significant genes in Cluster 2 selected by log fold change when compared to Clusters 1 and 3. 206
Table 7.3: Most highly significant genes in Cluster 3 selected by log fold change when compared to Clusters 1 and 2. Genes denoted with * were also significant in the GeCKO analysis (see 6.5*). Genes denoted ** were included for a discussion point as critical genes implicated in and cancers and liposarcoma respectively..... 207
Table 7.4: Most significantly expressed genes in tumour versus normal by log fold change. Gene function also summarised (248). 216

Abbreviations

ALT	Atypical lipomatous tumour
CDK4	Cyclin-dependent kinase 4
CEBP	CCAAT-enhancer-binding family
CRISPR	Clustered regularly interspaced short palindromic repeats
CT	Computed tomography
DC-PIP	Dichloroindophenol
DDLS	Dedifferentiated liposarcoma
DMSA	Dimercaptosuccinic acid
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EORTC	European Organisation for Research and Treatment of Cancer
FBS	Fetal bovine serum
FDR	False discovery rate
FFPE	Formalin-fixed paraffin-embedded
FISH	Fluorescence in situ hybridisation
FNCLCC	French Federation of Cancer Centers Sarcoma Group
FOXD4L3	Forkhead box D4 like 3
GC-MS	Gas Chromatography Mass Spectrometry
GeCKO	Genome wide CRISPR-Cas9 knockout
GFP	Green fluorescent protein
GIST	Gastrointestinal stromal tumour
HBRC	Human Biomaterials Resource Centre
HMGA2	High mobility group protein AT-hook 2

HPLC	High performance liquid chromatography
JNK	c-JUN N-terminal kinase
MDM2	Mouse double minute 2 homolog
MRI	Magnetic resonance imaging
MSC	Multipotent stem cell
mTOR	Mammalian target of rapamycin
NAC	Neoadjuvant chemotherapy
NES	Normalised enrichment score
NGS	Next generation sequencing
NHEJ	Non-homologous end-joining
PET	Positron Emission Tomography
PORCN	Porcupine
PPAR-γ	Peroxisome proliferator-activated receptor
QEH	Queen Elizabeth Hospital, Birmingham
RET	Rearranged during transfection
RIPA	Radioimmunoprecipitation assay buffer
RPLS	Retroperitoneal Liposarcoma
RPMI	Roswell Park Memorial Institute
RPS	Retroperitoneal Sarcoma
RT-PCR	Real time polymerase chain reaction
scRNA-seq	Single cell RNA sequencing
SDH	Succinate dehydrogenase
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SERCA	Sarco/endoplasmic reticulum (SR/ER) Ca^{2+} ATPase pump
t-SNE	T-distributed stochastic neighbour embedding

TBE	Tris-borate EDTA
TBST	Tris-buffered saline/Tween
TCA	Tricarboxylic acid
TEMED	Tetramethylethylenediamine
TSPAN31	Tetraspanin 31
WDLS	Well differentiated retroperitoneal liposarcoma
Wnt	Wingless-related integration site

Publications arising from this thesis

1. Tyler R, Wanigasooriya K, Taniere P, Almond M, Ford S, Desai A, Beggs A. A review of retroperitoneal liposarcoma genomics. *Cancer Treatment Reviews*. 2020 Jun 1
2. Tyler R, Dilworth M, James J, Blakeway D, Stockton J, Morton, D, Taniere P, Gourevitch D, Desai A, Beggs AD. The Molecular Landscape of Well Differentiated Retroperitoneal Liposarcoma. *The Journal of Pathology*. 2021 Jan 4.
3. Tyler R, Lee M, Ierodiakonou V, Hodson J, Taniere P, Almond M, Ford S, Desai A. Prognostic implications of histological organ involvement in retroperitoneal sarcoma. *BJS Open*. 2021 Oct 5

Chapter 1: Introduction

1.1 Sarcoma

The term sarcoma derives its origin from the Greek word ‘sarc’ meaning flesh, thus a tumour of the flesh. They are termed mesenchymal in origin, relating to the connective tissues and can arise from the mesoderm or ectoderm (1). This is in contrast to carcinomas, which have an epithelial origin and arise from the endoderm. The current World Health Organisation classification lists more than 100 subtypes of sarcoma (2). The majority of sarcomas are assigned a prefix, which describes the tissue of origin e.g., liposarcoma, a sarcoma originating from adipose tissue, or angiosarcoma originating from blood vessels.

For simplicity, sarcomas can be broadly split into soft tissue sarcomas and sarcomas of bone. There are over 5300 new sarcomas diagnosed in the UK every year, accounting for 1.3% of all cancers (3). Of these, 2/3 are soft tissue sarcomas equating to around 9 new diagnoses per day (4). Soft tissue sarcomas are relatively rare, but are more common than well-known cancers, such as testicular cancer. Indeed, soft tissue cancers carry a 10-fold increase in mortality in comparison with testicular cancer (1).

1.2 Liposarcoma

Liposarcoma is the most common soft tissue sarcoma in adults. In children, rhabdomyosarcoma is most common, whereas liposarcomas are very rare (5). They are usually large and are found in the extremities (popliteal fossa/ medial thigh),

retroperitoneum, spermatic cord, groin and shoulder area. Liposarcoma of the extremities accounts for 75% of diagnoses with 20% occurring in the retroperitoneum and the remaining divided between other sites (6). Their peak incidence occurs in the 6th and 7th decades with an equal predilection for men and women (6). There is no known aetiology or risk factor specific to liposarcoma, and exercise, diet and smoking have not known to be associated with their development either.

The prevailing opinion is that liposarcoma develops from an adipocytic progenitor cell halted in its differentiation which then undergoes lipogenic arrest (7). These progenitor cells are thought to be pericapillary adventitial cells that closely resemble fibroblasts with a capacious endoplasmic reticulum. Fat accumulates in the cytoplasm and the cell loses is endoplasmic reticulum and forms a round shape, with the nucleus pushed to the corner of the cell (8). The end result is the lipoblast, a neoplastic cell pathognomonic of liposarcoma (Figure 1.1).

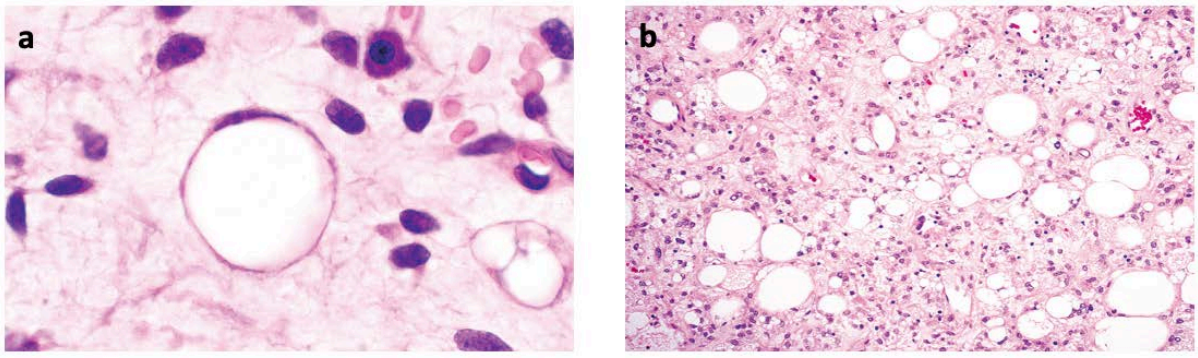

Figure 1.1: Haemotoxylin and Eosin stains of liposarcoma performed by pathologists post-operatively. This shows (a) Lipoblast at late stage of development, (b) Well differentiated liposarcoma (WDLS) with numerous lipoblasts.

Challenges exist in non-specialist practice when differentiating between the benign lipoma and the malignant liposarcoma. Lipomas are benign adipocytic tumours that arise in superficial tissues in contrast to the deep seated liposarcomas. They do not have the ability to metastasise, nor become liposarcomas and are not part of a spectrum of adipocytic malignancies (8).

Liposarcomas can be divided into numerous subgroups, and histology is currently the most significant prognostic indicator. Goldblum et al. modify the World Health Organisation classification of liposarcoma by dividing them into three groups, the first of which is subdivided into well differentiated (WDLS) or atypical lipomatous tumours (ALT) and dedifferentiated (DDLS) liposarcomas (7). The second and third subgroups are myxoid or round cell liposarcoma and pleomorphic liposarcoma.

ALT and WDLS are synonyms and the choice of one over the other is based on location and not on histological differences (9). ALT arise in the superficial soft tissues or the muscles of the extremity and exhibit low-grade behaviour. WDLS is used to describe retroperitoneal lesions, which have the ability to dedifferentiate and become locally invasive. There is no recognised explanation for why the anatomical location, promotes a more aggressive phenotype but it is likely due to the fact that retroperitoneal tumours can present much later than their extremity counterparts.

1.3 Retroperitoneal Tumours

The retroperitoneum represents a complex potential space that can host a large variety of benign and malignant tumours, both primary and metastatic (10). A third of retroperitoneal tumours are sarcomas, with the remaining two thirds being split

between primary epithelial cancers such as adrenal and pancreatic, primary lymphoproliferative tumours such as Hodgkin's lymphoma and metastatic disease from known or unknown primary sites such as testicular cancer (11).

1.3.1 Retroperitoneal Sarcomas

In the United Kingdom there are between 250-300 new cases of retroperitoneal sarcoma (RPS) diagnosed each year, making it the second most common site of sarcoma after the extremities (10). 70% of these are liposarcoma, with 15% leiomyosarcoma and smaller numbers of malignant nerve sheath tumour, solitary fibrous tumour and synovial sarcoma (11).

1.4 Retroperitoneal Liposarcoma

As a general rule, WDLS and DDLS are the most common retroperitoneal liposarcomas, with myxoid and pleomorphic liposarcomas very rarely encountered.

WDLS is a low-grade tumour, composed of proliferating mature adipocytes and accounts for 40-45% of all liposarcomas (12). WDLS are frequently > 30cms in size at diagnosis. They do not metastasise, but have a high propensity for local recurrence with rates of up to 43% at 8 years when not given radiotherapy (13). Importantly, 20% of WDLS will dedifferentiate to a higher grade tumour, on average at seven to eight years (10).

DDLS is a high grade, more aggressive, typically non-lipogenic sarcoma with the ability to metastasise (14). It can either arise de novo, as a recurrence of WDLS or juxtaposed to WDLS. Dedifferentiation is the term used to describe a morphological

regression to a less differentiated state. In an inverse relationship to WDLS, 75% of DDLS are of retroperitoneal origin, with 25% in the extremity (7). Figure 1.2 illustrates the anatomical origin of WDLS/DDLS.

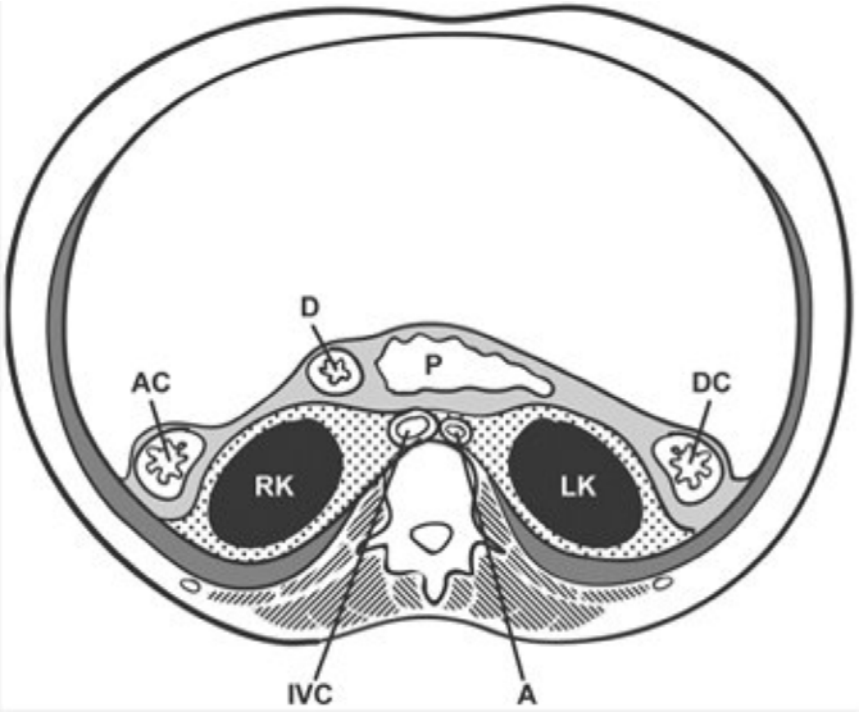

Figure 1.2: Axial cross section of abdominal cavity. Labelled compartment is retroperitoneum. AC (ascending colon), D (duodenum), P (pancreas), DC (descending colon), RK (right kidney), LK (left kidney), IVC (inferior vena cava), A (aorta). Dotted pattern surrounding kidneys represents retroperitoneal fatty tissue, from which WDLS/DDLS arise. Image taken from Journal of Trauma and Acute Care Surgery (15).

Primary retroperitoneal myxoid liposarcomas (MLS) are extremely rare and are usually misdiagnosed, in reality being a metastatic deposit from an undiscovered extremity myxoid liposarcoma. The largest series of MLS documented 5/213 to have a primary retroperitoneal origin (16). Pleomorphic liposarcomas (PLS) are also very rare. Hornick et al. in a series of 57 PLS, found only 4 to arise from the retroperitoneum (17). These tumours are poorly understood and carry a dismal prognosis.

1.4.1 Presentation

The capacious nature of the retroperitoneum allows these tumours to grow to a considerable size before detection and are the largest tumours on record (18). The majority of patients are either asymptomatic or have vague, non-specific complaints at initial presentation. Asymptomatic patients may be diagnosed incidentally on cross sectional imaging done for another purpose. Symptomatic patients may notice an increased abdominal girth, potentially self-palpating an abdominal mass. 50% present with abdominal pain (18). In rarer cases, sensory symptoms may predominate, and some patients can suffer motor symptoms from involvement of the femoral nerve. Physical examination may reveal a palpable abdominal mass, with fullness or asymmetry.

1.4.2 Investigation

National guidelines within the UK state that any patient with a suspected soft tissue sarcoma should be referred to a diagnostic centre for triple assessment with clinical history, imaging and biopsy (19).

1.4.3 Radiological investigation

CT (computed tomography) or MRI (magnetic resonance imaging) are mandatory for the investigation of a suspected RPS and are essential in pre-operative planning. A contrast enhanced CT can differentiate between a WDLS and DDLS. WDLS contain a large volume of mature fat with mottled or ‘streaky’ zones (18). These features have been shown to have high diagnostic sensitivity of up to 100% (20). Due to this

high level of radiological sensitivity surgery can be planned without biopsy when imaging is reviewed by an experienced diagnostician (21).

MRI scanning has a role in delineating neurovascular involvement or pelvic tumours, but CT often suffices and to date no data exists to indicate a superior imaging modality in pre-operative planning (22, 23).

At diagnosis patients should undergo chest imaging, either as a chest x-ray or CT to evaluate for lung metastasis, which is the most common site of distant disease in DDLS (24). A functional assessment of the contralateral kidney is necessary for planning treatment of the ipsilateral RPS. This is usually achieved using differential renal scanning, such as a DMSA (dimercaptosuccinic acid) scan (23).

1.4.4 Biopsy

Image-guided percutaneous coaxial core needle biopsy is strongly recommended unless the imaging is pathognomonic, (e.g. heterogenous dedifferentiated/ well-differentiated liposarcoma (23). If there is uncertainty regarding the diagnosis or histology is needed to plan neoadjuvant or palliative chemotherapy for locally advanced or unresectable disease, then a biopsy is imperative. The safest method of biopsy appears to be a trans-retroperitoneal approach with a co-axial technique (25).

1.4.5 Pathological Assessment of biopsy and final specimen

The pathological assessment of sarcoma specimens should be carried out by a sarcoma specific pathologist due to the rarity and complexity of the microscopic appearances, and the adverse impact on survival of an incorrect diagnosis. Biopsy specimens should be given a histological report and grade.

Final pathological assessment of RPL starts with gross macroscopic appearance. Grossly WDLS appear as deep yellow, multilobar lesions with fibrous bands and gelatinous regions (Figure 1.3). Tumours are subjected to haematoxylin and eosin staining. Diagnostic criteria for WDLS requires mature fat with the presence of atypical cells. Without atypical cells present, the diagnosis cannot be made. These atypical cells tend to be spindled cells with hyperchromatic nuclei (26). When pathological interpretation is difficult due to the subtlety of cytological atypia on routine histology, cytogenetic and molecular testing is needed to aid in classification. Differentiating between benign lipomas, other malignant neoplasms and non-lipomatous neoplasms can be incredibly challenging without these ancillary tests.

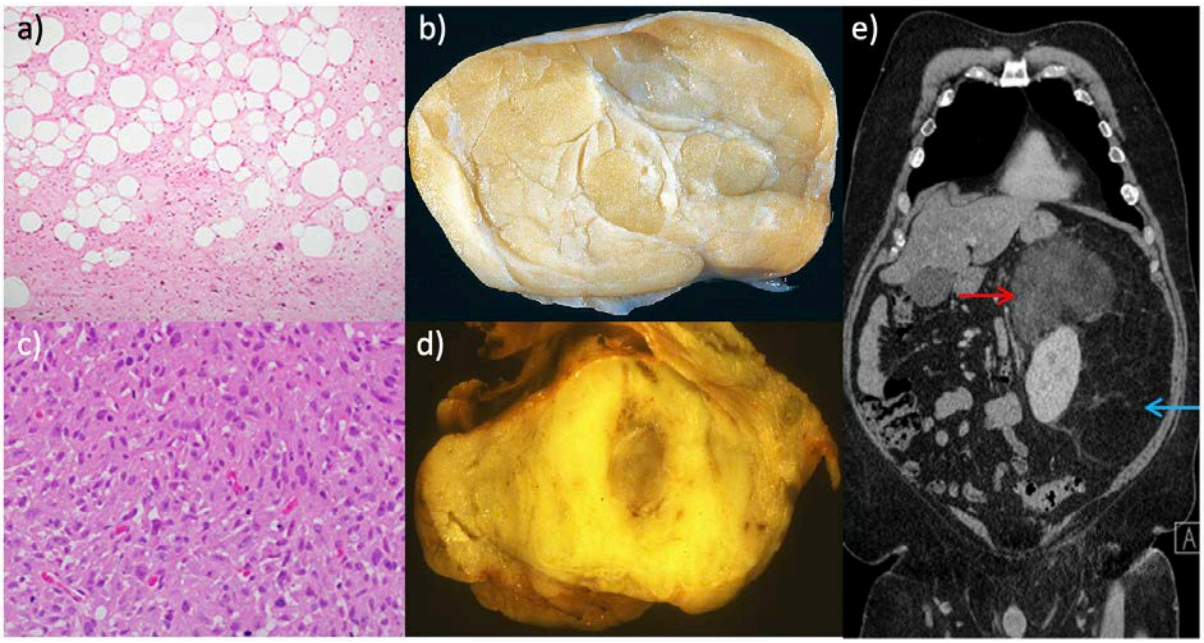

Figure 1.3: Histologic and CT appearances of retroperitoneal liposarcoma. (a) H+E stain of WDLS (X40 magnification). (b) WDLS resembling normal fat, with fibrous septa. (c) H+E stain of DDLS (x200), note the lack of adipocytic differentiation. (d) DDLS. Central darker area represents dedifferentiated tumour, surrounded by yellow/white WDLS component. (e) Coronal CT image of retroperitoneal liposarcoma. Blue arrow denotes a homogenous WDLS component extending inferiorly. Red arrow denotes heterogenous DDLS component superior to left kidney. Images (a), (c) and (e) taken from Journal of Clinical Oncology (27). Image (b) and (d) taken from Enzinger and Weiss's Soft Tissue Tumors (7).

Immunostaining for Mouse double minute 2 (*MDM2)* and Cyclin dependent kinase 4 (*CDK4*) has since been replaced with fluorescent in situ hybridisation (FISH) as a means of separating WDLS from other benign lipomatous lesions which do not overamplify MDM2. It is a far more sensitive and specific method of diagnosing WDLS and works on biopsy material (28). FISH works by counting the *MDM2* copy number in individual cells with a ratio calculated in reference to a standard centromeric probe (CEP12) (Figure 1.4). Thway et al. reported FISH concordance rates in WD/DDLS of 96.5%. However, FISH has its limitations, with the authors noting that it was labour and cost-intensive, and performed less well in a subgroup of microscopically equivocal well-differentiated lipomatous neoplasms (29).

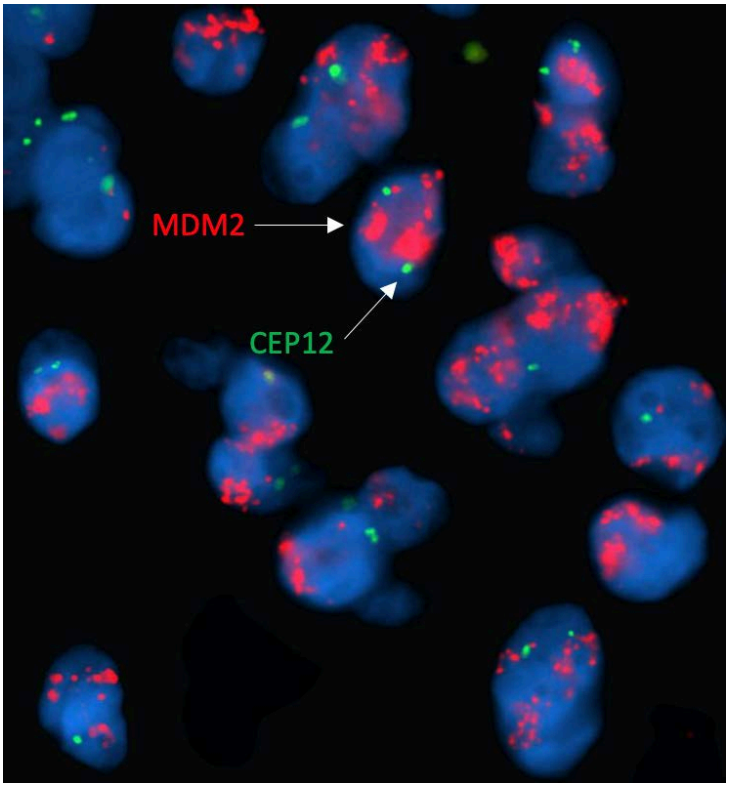

Figure 1.4: Image of fluorescent in situ hybridisation (FISH) performed on WDLS cells. This technique is carried by the pathologist post-operatiely. Amplification of MDM2 (red) compared to reference probe CEP12 (green). Image taken and adapted from Clinical Case Reports (30).

1.4.6 Staging

Staging soft tissue sarcomas uses the American Joint Committee on Cancer TNM staging (23). The majority of tumours are >15 cm (T4) at presentation and in the two main subtypes (LMS and LPS) lymph node metastasis are rare. For this reason, more weight is given to grading (G1-G3) and currently the French Federation of Cancer Centers Sarcoma Group (FNCLCC) system is used. The FNCLCC grade is determined by three parameters: differentiation (histology specific), mitotic activity, and extent of necrosis (31). This grading system is strongly correlated with disease-specific survival (20). The full TNM and grading system is given in Table 1.1 and 1.2.

Table 1.1: AJCC TNM staging for retroperitoneal soft tissue sarcoma (32)

Tumour (T) category	**T Criteria**
TX	Primary tumour cannot be assessed
T0	No evidence of primary tumour
T1	Tumour less than 5 cm in greatest dimension
T2	Tumour more than 5 cm and ≤10 cm in greatest dimension
T3	Tumour more than 10 cm and ≤15 cm in greatest dimension
T4	Tumour more than 15 cm in greatest dimension
Nodal (N) status	**N Criteria**
N0	No regional lymph node metastasis or unknown lymph node status
N1	Regional lymph node metastasis
Metastasis (M) status	**M Criteria**
M0	No distant metastasis
M1	Distant metastasis

Table 1.2: FNCLCC grading system, table taken from Archives of Pathology and Laboratory Medicine (33). HPF - high powered field.

Tumour differentiation	**Definition of parameters**
Score 1	Sarcomas closely resembling normal adult mesenchymal tissue (e.g WDLS)
Score 2	Sarcomas for which histologic typing is certain (e.g MLS/DDLS)
Score 3	Embryonal and undifferentiated sarcomas, sarcomas of unknown origin
Mitotic count	
Score 1	0-9 mitoses per 10 HPF
Score 2	10-19 mitoses per 10 HPF
Score 3	≥20 mitoses per 10 HPF
Tumour necrosis	
Score 0	No necrosis
Score 1	<50% tumour necrosis
Score 2	≥50% tumour necrosis
Histologic grade	
Grade 1	Total score 2,3
Grade 2	Total score 4,5
Grade 3	Total score 6,7,8

Clinicians can also use a nomogram developed by Gronchi et al. that has been shown to predict post-operative disease-free and overall survival (34). This takes into account patient age, tumour size, grade, histologic subtype and the presence of multifocality, defined as two or more tumours.

1.5 Surgery for retroperitoneal liposarcoma

1.5.1 Pre-operative planning

As tumours frequently present late, patients can already be malnourished prior to surgery. Nutritional status, cardiorespiratory fitness and the need for psychological support all need to be considered prior to embarking on such extensive surgery.
If imaging suggests ureteric, vascular or hepatic involvement then the relevant specialities should be involved to ensure a complete and safe resection.

1.5.2 Anatomical considerations

The only current potentially curative option for retroperitoneal liposarcoma is resectional surgery where the primary aim is to leave no macroscopic disease where possible (22). Local recurrence is responsible for 75% of all liposarcoma related deaths (35). Retroperitoneal liposarcomas are often close to vital structures in the retroperitoneum and it is not uncommon to remove ipsilateral kidney and colon, but also spleen, tail of pancreas, diaphragm and liver. Rates of resection of adjacent involved organs ranging from 34-93% have been reported (36).

The majority of UK sarcoma units aim for en bloc resection of the sarcoma and contiguous organs that are macroscopically involved by the sarcoma; an approach called extended compartmental resection. No efforts are made to resect organs that lie adjacent to the tumour but are not involved by it. Conflicting practice on the extended compartmental approach does exist internationally. Judgment needs to be made, often preoperatively when deciding which neurovascular structures can or should be sacrificed. For example, it is not uncommon for retroperitoneal sarcomas to intimately involve the femoral nerve, which requires excision along with the tumour. This decision needs weighed up against the potential for a long-term reduction in mobility.

The principal site of retroperitoneal liposarcoma recurrence is loco-regional. Techniques that reduce intraabdominal recurrence should be pursued. In this regard, the TARPSWG advise complete resection of all retroperitoneal fatty tissue at risk of harbouring tumour (23). It is these sort of techniques which high volume sarcoma centres are able to practice and lead to better treatment outcomes (37).

1.5.3 Debate on extended compartmental resection

A highly debated issue in the surgical management of RPS is the extent of resection needed to achieve complete resection. In extended compartmental resection (ECR) adjacent organs are removed even without gross evidence of tumour invasion at the time of surgery where those organs would allow a 1 cm margin (18). The other option is simple complete resection where adjacent organs such as the kidney are left unresected. Two groups have shown improved locoregional disease control in those who had undergone ECR versus those with a simple complete resection (38,

39). Importantly, Gronchi et al. have demonstrated improved 5-year overall survival (40).

Critics of ECR argue that it a highly morbid technique with significant complication rates ranging between 22-31% reported (41). Others argue that the additional 1 cm margin cannot be guaranteed in every case, due to the proximity to unresectable structures such as the duodenum or aorta. The debate continues and will do so until higher quality randomised evidence exists.

1.6 Non-operative interventions for retroperitoneal liposarcoma

1.6.1 Radiotherapy

Complete surgical excision is the gold standard of treatment for retroperitoneal liposarcoma, but this is not always possible. Incomplete excision is not uncommon, and many patients develop local recurrence regardless of the success of surgery. Importantly it is local recurrence that patients die from, rather than distant metastatic disease (42). As a consequence of this there is great interest in reducing local recurrence via non-surgical means. The success of radiotherapy in extremity sarcoma at lowering local recurrence combined with the theory that residual microscopic disease drives recurrence has formed the rationale for radiotherapy in RPS (43).

For several reasons radiotherapy is normally given neo-adjuvantly. The primary reason is that the tumour displaces radiosensitive organs such as small bowel out of

the field of radiation, limiting the toxicity and allowing a higher dose of radiation (18). Adjuvant radiotherapy causes small bowel enteritis in up to 60% of cases (44). Secondly, pre-operative radiotherapy requires a smaller field, and a lower dose compared to post-operative regimes (45).

In September 2020 the results of the first randomised control trial in retroperitoneal sarcoma – STRASS - were published (46). This trial recruited patients with non-metastatic operable RPS to neo-adjuvant radiotherapy and surgery (RT/S) or surgery alone (S). Overall, preoperative radiotherapy did not show any significant benefit and was not recommended. However, in the liposarcoma subgroup, 3-year abdominal recurrence-free survival was 71.6% and 60.4% in RT/S versus S groups (HR = 0.64, 95%CI 0.40-1.01). The authors conclude that a further prospective clinical trial in RPLS only is required.

1.6.2 Chemotherapy

The role of chemotherapy in WDLS/DDLS is limited by poor response rates and a lack of data concerning molecular targets in sarcoma. The majority of chemotherapy is given in a metastatic and palliative setting. Doxorubicin and ifosfamide have been the agents of choice over the last twenty years, with a growing interest in more targeted therapies such as nutlins and CDK4 antagonists.

1.6.2.1 Neoadjuvant chemotherapy

The effectiveness of neoadjuvant chemotherapy (NAC) for high risk soft tissue sarcomas is a matter of extensive debate (47). There are concerns that by delaying

surgery in a tumour operable at diagnosis with a NAC regime, it would bring minimal benefit with a risk of rendering the tumour inoperable. The primary goal of any NAC in retroperitoneal sarcoma must be to improve rates of histologically negative margins. These concerns have been raised in other cancers, such as colorectal and pancreatic. The FOxTROT trial randomised patients with colorectal cancer to NAC or adjuvant chemotherapy (AC) and found no significant evidence of pre-operative progression, whilst histological margins improved (48). Neoadjuvant FOLFIRINOX in resectable pancreatic cancer has also shown higher negative margin rates in the randomised setting (49).

Nevertheless, in RPS the delivery of NAC regimes is hindered by challenges in identifying high grade tumours pre-operatively and the limited number of available drugs to choose from (50). The evidence for NAC in retroperitoneal sarcomas is extrapolated from extremity sarcoma, where some minor benefits in overall survival are used to justify its implementation (51). There have been no randomised control trials comparing NAC and surgery to surgery alone in RPS, making it difficult to draw any firm conclusions. Miura et al, in a National Cancer Database registry, included 163/ 8653 patients with RPS who received NAC (52). There was a lower overall survival in the chemotherapy group compared with surgery alone (40 months v 68.2 months; $p<0.01$). However, the main criticism was that patients with high-risk tumours were significantly more likely to receive chemotherapy.

In contemporary practice NAC is considered for larger higher-grade lesions such as DDLS and not WDLS, and particularly those that require extensive multivisceral surgery. It is hoped that when more targeted NAC regimes are developed, they will

be able to treat micrometastases from haematogenous spread, e.g DDLS, and potentially downsize a tumour so curative surgery is possible (53).

As a follow-up to the STRASS trial, STRASS 2 is a phase III study, randomising patients with retroperitoneal sarcoma to either three cycles of neoadjuvant chemotherapy plus surgery, versus surgery alone. The experimental arm is divided into high grade LPS and LMS. High grade LPS is to receive doxorubicin and ifosfamide whilst patients with LMS will receive doxorubicin and dacarbazine. It intends to recruit 250 patients over 5 years and starts imminently (54).

1.6.2.2 Adjuvant chemotherapy

Adjuvant chemotherapy (AC) is not considered standard practice in the management of RPLS. A pooled analysis from the two European Organisation for Research and Treatment of Cancer (EORTC) phase 3 clinical trials in soft tissue sarcoma failed to demonstrate a survival advantage for adjuvant doxorubicin-based chemotherapy, specifically in the retroperitoneal subgroup (55). This was further demonstrated by Datta et al. (56) who compared 390 RPS (majority liposarcoma) resections who received AC with 377 who received surgery alone using propensity score-matching methodology. The AC group had significantly worse long-term survival (47.8 v 68.9 months, p=0.017). The likelihood of further specific retroperitoneal sarcoma AC trials seems low in the face of limited interest by the oncology community, a testament to the lack of meaningful results to date.

1.6.2.3 Chemotherapy for recurrent or metastatic retroperitoneal liposarcoma

Given poor rates of response, systemic chemotherapy is generally reserved for unresectable or metastatic disease. Doxorubicin-based chemotherapy, either as a single agent or with ifosfamide, has been a standard of treatment for recurrent soft tissue sarcomas for many years. Italiano et al. in a in the largest series of advanced DDLS and WDLS found the overall objective response rate (using RECIST criteria) to doxorubuicin based chemotherapy was 12%, but this only included 12 DDLS and no WDLS (57). A separate study showed an objective response rate of 25% to high dose ifosfamide in 6 patients with WDLS and 22 with DDLS (58). It is important to note these cytotoxic drugs require good patient performance status something that may not always be present in the metastatic setting (59).

Second line treatments include newer agents such as eribulin and trabectedin are approved for advanced, previously treated soft tissue sarcomas. Eribulin, a microtubule inhibitor showed promise in a phase 2 study with 15/32 patients with advanced DDLS showing progression free survival at 12 weeks (60). Trabectedin, a marine-derived alkylating agent, acts in several ways, but primarily binds to the minor groove of DNA and causes the FUS-DDIT3 fusion protein to be displaced from promoters (61). This FUS-DDIT3 oncoprotein is found in myxoid liposarcoma and therefore trabectedin is a uniquely useful drug in this subtype, with response rates of up to 51% demonstrated (62). However, there is no good evidence for trabectedin use in WDLS/DDLS and is not recommended in current guidelines. The poorer treatment responses observed here are in part due to the paucity of molecular

targets in liposarcoma, a reflection of our limited understanding of the tumour biology.

1.7 Surveillance

Retroperitoneal sarcomas are at high risk for locoregional recurrence post resection. Prompt detection can improve resectability and the development of complications such as bowel obstruction and malignant ascites (23). Rates of local recurrence for WDLS approach 100% and it could be argued that these are essentially incurable lesions (7). WDLS dedifferentiate on average at 7 to 8 years, but this can be up to 20 years after the original diagnosis. Henricks et al. found a 24% local recurrence rate and a 17% metastatic rate at three years, with the majority of metastases to the lungs (74%) and liver (24%) (14).

These high recurrence figures inform the various surveillance guidelines for soft tissue sarcoma, of which two, the National Comprehensive Cancer Network and the French National Cancer Institute are specific for retroperitoneal sarcoma, with the TARPSWG giving general guidance (63). The general consensus is that follow up should be shorter initially as radiological recurrence may antedate symptomatic recurrence. A CT scan with chest x-ray every four months for five years, following this a six-monthly chest x-ray and an annual CT scan. These are by no means didactic and common sense must be exercised. For example, DDLS are far more likely to get lung metastasis and will need more regular chest imaging than for a WDLS, and so the intensity of follow up should be tailored to the risk category.

1.8 Management of recurrent disease

Recurrence is unfortunately a common occurrence in retroperitoneal liposarcoma and is the leading cause of death. For patients who develop recurrence in the absence of metastatic disease surgery is still the primary treatment option but is not always the correct one. The decision to operate is more complicated due to the lower complete resection rates than primary resection – 50-60% versus 70-80% respectively (64). The loss of tissue planes make surgery far more challenging with a higher risk of complications such as enterotomy or bleeding.

Factors associated with a longer survival after salvage surgery for RPS are a long recurrence-free interval, unifocal disease and complete resection (42, 65, 66). Park et al. also showed that tumour growth of less than 0.9 cm/month after recurrence was predictive of longer disease free survival when the local recurrence was resected (67). Ikoma et al. hypothesised that watchful waiting to characterise the clinical behaviour of the disease would improve patient selection. They found that a recurrence to salvage interval of greater than six months was associated with an increased survival (64). All of these findings have to be considered when discussing surgery for local recurrence in retroperitoneal liposarcoma. In the right patient, salvage surgery can extend life expectancy and improve quality of life. But patients with early, fast growing or multifocal local recurrences should be considered for systemic therapies, with operative intervention reserved for palliation only.

1.9 Summary management of RPLS

Surgery with negative margins remains the only curative option in RPLS, and yet recurrence rates remain high. There is some evidence of a small benefit from pre-operative radiotherapy but no convincing evidence for chemotherapy at any time point. More recently, efforts into understanding the genome of sarcomas have paved the way for advancements in other subtypes, such as trabectedin in myxoid liposarcoma, and imatinib in gastrointestinal stromal tumour (GIST). Understanding the molecular landscape of WDLS is in its infancy and represents an area of genuine opportunity.

1.10 Genomic alterations in liposarcoma

1.10.1 Introduction to sarcoma genomics

Cancer is largely a genetically encoded disease. The developments in molecular genetics, epigenetics and genomics has vastly changed cancer diagnosis, classification, prognostication and treatment (68). These advances have formed the basis of personalised medicine.

Sarcomas can be cytogenetically divided into two distinct groups: sarcomas with specific genetic alterations and sarcomas with complex karyotypic abnormalities with no recurrent genetic mutations (68). In the first category is Ewing's sarcoma which expresses the EWSR1-FLI1 fusion (69). Oncogenic fusion transcripts account for 15-20% of all sarcoma cases (70). Secondly, specific mutations in the proto-oncogene *C-KIT* and *PDGFRA* are responsible for GISTs, and molecular therapies have been very successful in targeting these. Pleomorphic liposarcomas are a good

example of the second category with complex karyotypic abnormalities and no recurrent genetic mutations, other examples are leiomyosarcoma and undifferentiated pleomorphic sarcoma.

1.10.2 Introduction to liposarcoma genomics

This subchapter aims to focus predominantly on WDLS and DDLS. Efforts are made to concentrate on each separately, but due to the ability of WDLS to dedifferentiate, there are shared genomic drivers. A large proportion of **1.10.2** to **1.12** is taken from a review article I had published entitled "A review of retroperitoneal liposarcoma genomics" (71).

WDLS are characterised by the presence of supernumerary ring and/ or giant marker chromosomes. The mutational mechanisms that underpin the formation of these structures are fundamental to liposarcoma genomics. Early work by Garsed et al. proposed the formation of ring and marker chromosomes through a process called chromothripsis (72). Chromothripsis is characterised by a massive genomic fragmentation and rearrangement, generated in a single event localised to an isolated chromosomal region; which in WD/DDLS is chromosome 12q (73). An amplification phase follows in which repetitive break-fusion bridge cycles allow incorporation of additional chromosomal regions to the ring neochromosomes and concludes with neochromosomal stabilisation (74). A core group of oncogenes located on chromosome 12q are almost always amplified, including MDM2, likely as a result of active selection (72).

More recent work by Cortes-Ciriano on behalf of the Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium analysed patterns of chromothripsis across 2658 tumours from 38 cancer types using whole genome sequencing data (73). The most common cancer type to exhibit chromothripsis was liposarcoma, in 100% of samples. It is increasingly recognised as a major driver of cancer evolution, but the events that initiate chromothripsis are more poorly understood. The 'initiation event' could be a random event which then becomes fixed through natural selection in a precursor cell, or a separate mechanism yet to be discovered (75).

This amplified region of chromosome 12 houses multiple genes, with the most compelling evidence to date demonstrating an oncogenic role for *MDM2*, *CDK4*, High mobility group protein AT-hook 2 *(HMGA2)* and Tetraspanin 31 (*TSPAN31)* (76). In general amplification of this region results in increased cell proliferation and decreased apoptosis (77). Outside of this amplicon are the c-Jun N-terminal kinase (JNK) and Fibroblast growth factor (FGF) pathways, whilst Peroxisome proliferator-activated receptor (*PPAR-γ*) and Rearranged during transfection (*RET)* are implicated as well (8). Some data exists on epigenetic regulation in liposarcoma methylomes with significant interest in the CCAAT-enhancer-binding (*CEBP)* family expression (78).

1.10.2.1 MDM2

MDM2 is the most studied of all genomic aberrations in RPLS. Its amplification is considered to represent one of the earliest events in the formation of WD/DDLS, although whether it is the primary event remains to be elucidated. Arguing against this is the requirement for extrachromosomal material to be formed, which is unlikely

to be driven independently by one single gene. Furthermore, Cortes-Ciriano observed patterns of mutations in chromothripsis regions with high-level copy number amplifications in liposarcomas that suggested that these mutations occurred at the late stage of tumour development, likely after chromothripsis (73).

MDM2 amplification has been found to correlate with elevated *MDM2* expression (79). *MDM2* encodes a negative regulator of the tumour suppressor p53. It acts by binding to the transcription activation domain of p53, blocking p-53 dependent transcription (80). *MDM2* also targets p53 for proteasomal degradation, acting as ubiquitin ligase (81). With the occurrence of DNA damage *MDM2* is able to auto-degrade, and allow p53 levels to return to normal, enabling it to actively induce cell death in response (82). It therefore maintains a tight control on p53 levels in healthy cells, something which is disrupted in MDM2 overamplification.

The current data from human tumours collectively supports an oncological role for MDM2 amplification by abrogating the tumour suppressive p53 pathway. By allowing cells to progress through the cell cycle under conditions that could generate or perpetuate DNA damage tumourigenesis is initiated (83).

Importantly this amplification represents an alternative mechanism to p53 mutation to inactivate the p53 -signalling pathway and promote tumour progression in sarcoma. Vassilev et al. were able to induce apoptosis in *MDM2* amplified sarcoma cell lines through MDM2 antagonists, highlighting that the downstream p53 pathway is structurally intact. This mutual exclusivity and the presumed functional p53 pathway is why MDM2 is such an attractive therapeutic target.

MDM2 is amplified in 7% of human cancers without concomitant p53 mutation and a third of all sarcomas, whereas in WD/DDLS rates greater than 90% are reported (84). Somaiah et al. performed exome sequencing on 17 WD/DDLS; *MDM2* and *CDK4* amplification were the only two gene amplifications universally identified in every sample (85).

In the largest study of its kind Bill et al. correlated the degree of *MDM2* amplification with clinical outcomes. Elevated *MDM2*, as measured by genomic amplification and mRNA expression, was associated with a poorer recurrence-free survival (86). Other studies have confirmed a higher *MDM2* copy number is associated with poorer outcomes, poorer response to cytotoxic therapy and interestingly a higher propensity for a retroperitoneal location (16, 87).

MDM2 amplification has become a diagnostic tool for differentiating WD/DDLS from benign lipomatous lesions and other high grade sarcomas, and in comparison to other genomic markers has a higher sensitivity (88).

1.10.2.2 CDK4

CDK4 is amplified in over 90% of WD/DDLS cases (89). The *CDK4* gene encodes a 33-kD protein that forms complexes within cyclin D. This complex phosphorylates protein Rb (pRB) dissociating it from the pRB-E2F complex (8). E2F binds DNA to upregulate transcription of genes required for G1-S transition (90). Alterations in the CDK4/RB1/E2F pathway play a crucial role in the pathogenesis of many tumour

types. Knock-in mice that express a CD4 mutant originally found in melanoma develop several tumours in variety of tissues (91).

Italiano et al. attempted to clarify the role of *CDK4* amplification in WD/DDLS. They compared the clinicopathological characteristics of 45 *CDK4*-ve/MDM2+ve cases with 143 *CDK4*+ve/*MDM2*+ve. *CDK4*-ve tumours were more likely to be low-grade lipoma like lesions, occur in the extremities and have a lower local recurrence and mortality rate (90). *CDK4* negativity likely confers a survival benefit. In addition, *CDK4* is often co-amplified with *MDM2* and is likely to result in increased proliferation through combined effects on p53 and the cell cycle checkpoints (92).

Lee et al. performed Real-time polymerase chain reaction (RT-PCR) on 48 patients who had undergone compete resection of an intraabdominal liposarcoma (31 WDLS/ 17 DDLS) (93). WDLS that developed a local recurrence had significantly higher levels of *CDK4* amplification (P=0.041). High levels of *CDK4* amplification were associated with poorer recurrence-free survival compared to low *CDK4* amplification at multivariate analysis.

These findings have led the way for pre-clinical research with *CDK4* antagonists. Helias-Rodzewicz et al. treated liposarcoma cell lines with a *CDK4* inhibitor NSC625987 and observed a significant increase in adipocytic differentiation in cells with eliminated copies of *CDK4* (94). Furthermore, Perez et al. induced senescence in both liposarcoma cell lines and mice xenografts using palbociclib, a selective *CDK4*/6 inhibitor (95).

Like MDM2, CDK4 provides a useful diagnostic marker, predominantly through FISH.

1.10.2.3 HMGA2

The 12q13-15 region contains *MDM2* and *CDK4* but also 164 other genes, which have been studied to varying degrees and of which *HMGA2* is one (96). *HMGA2* encodes for an architectural transcription factor and in sarcomas (liposarcomas, uterine leiomyosarcomas and salivary gland pleomorphic adenomas) is found to be rearranged, amplified and overexpressed (97). Of note, *HMGA2* is rearranged in up to 70% of benign lipomata, resulting in a truncated protein, and is implicated in their development (98).

In a series of 38 liposarcomas (7/38 WDLS), FISH analysis was used to flank the 12q13-15 region (96). *HMGA2* was always amplified and rearranged, at a rate similar to that of *MDM2*. More recently, a separate group used similar techniques showing amplification of the proximal parts of *HMGA2* (5'-untranslated region (UTR) and exons 1-3) was associated with WDLS and a good prognosis, whereas *CDK4* and *JUN* amplifications were associated with DDLS and a poorer prognosis (99).

Xi et al. demonstrated that *HMGA2* is required for *C/EBPβ*-mediated expression of *PPAR-γ*, promoting adipogenic differentiation. When *HMGA1* was knocked down, it impaired adipocyte growth and when overexpressed promoted the formation of mature adipocytes (100). Furthermore, Arlotta et al. hypothesised that the *HMGA2* protein specifically promoted the growth of adipocytes. This was confirmed by the very low fat phenotype of a homologous *HMGA2* knock out mouse. In transgenic

mice with a truncated *HMGA2* there was a high incidence of lipomatosis further linking *HMGA2* to a role in adipogenesis (101).

Finally, Narita et al. demonstrated that *HMGA2* proteins cooperate with the p16 tumour suppressor to promote cellular senescence, but this antiproliferative activity is negated by co-expression of *MDM2* and *CDK4* (102). Evidently, *HMGA2* is likely to play a role in the pathogenesis of liposarcoma but its direct role in adipocytic malignant transformation has not been elucidated. Italiano et al. believe that *HMGA2* alone will only lead to benign lipomata, but in combination with *MDM2*/*CDK4* amplification, a malignant phenotype will be induced (96).

1.10.2.4 TSPAN31

An additional gene in the 12q13-15 region is Tetraspanin 31 (*TSPAN31)* a transmembrane superfamily gene. It was originally identified from an amplified area in malignant fibrous histiocytoma and has a higher rate of amplification in intrabdominal liposarcomas than in extremity lesions (103, 104). Forus et al. analysed a panel of 98 sarcomas: *MDM2* was amplified in 9 samples and *TSPAN31* in 10; in 8 of these samples they were co-amplified (105).

Interestingly there appears to be a different expression profile of *TSPAN31* in WDLS compared to DDLS. In multiplex ligation-dependent probe amplification of 77 lipomatous tumours the amplification profile of *TSPAN31*, *MDM2*, *YEATS4*, *CDK4* and *HMGA2* was frequently found to have a higher level status in DDLS compared to WDLS. The authors concluded that amplification status could be related to the process of dedifferentiation in liposarcomas and potentially used as a predictive tool

in the future (106). Nevertheless, the exact pathogenic mechanism of *TSPAN31* in liposarcoma has yet to be illustrated and more work needs to be done before any useful clinical translation takes place.

1.10.2.4 FRS2

There has been recent interest in identifying other novel genes present on the 12q13-15 amplicon of which Fibroblast growth factor receptor substrate 2 (*FRS2*) is one. *FRS2* codes for a signal transducing protein that links receptor tyrosine kinases (RTKs) to downstream signalling pathways, such as MAPK/ERK and PI3K/AKT/mTOR (107).

Wang et al. documented *FRS2* amplification in all 57 liposarcomas and mRNA transcriptional upregulation in 19 WDLS samples, but not in lipomata or normal fat (108). Jing et al. more recently evaluated the frequency of *FRS2* amplification and its relationship with clinical features. In their series, 92.1% of WDLS were *FRS2* amplified, and retroperitoneal tumours were found to have a higher *FRS2/CEP12* ratio than those in the extremity (109).

Importantly, Zhang et al. have shown that the FGFR selective inhibitor NVP-BGJ-398 inhibited the growth of liposarcoma cell lines with concomitant suppression of FGFR signal transduction (110). This pathway serves as an additional potential therapeutic target.

1.10.2.5 YEATS 4

The last of the well documented genes on amplicon 12q13-15 is *YEATS4*. It is suggested that *YEATS4* is required for repression of p53 tumour suppressor pathway during cell proliferation, and when overamplified is able to overly supress the pathway leading to tumourigenesis (111). Barretina et al. used an shRNA knock down of *YEATS4* in a dedifferentiated liposarcoma cell line to reduce cell proliferation, lending support for *YEATS4* playing a key role in liposarcoma (112).

Apart from this study, the majority of data regarding *YEATS4* in liposarcoma is confined to amplification and overexpression with limited functional work having been undertaken. Outside of sarcoma, *YEATS4* is a recognised oncogene in non-small cell lung cancer and its knockdown in colorectal cancer cell lines induces apoptosis (113, 114).

1.10.2.6 Outside of the 12q13-15 amplicon

As sequencing technologies have improved, a deeper interrogation of the genome has become possible. Egan et al. were the first group to perform whole genome sequencing on a WDLS, in order to search for therapeutic targets outside of the 12q13-15 amplicon. They found 7 damaging single nucleotide variants, amplification across multiple chromosomes and 11 gene fusions, albeit in only one sample (115). Of note, they identified a gene fusion in amplified Discoidin domain-containing receptor 2 (*DDR2*), a gene involved in multiple cellular processes, present on 1q23.3. Importantly, the kinase domain was predicted to remain intact, which is clinically relevant, as *DDR2* activity can be curtailed by kinase inhibitors such as imatinib, nilotinib and dasatanib.

1.10.2.7 Methylation patterns

Many sarcomas contain alterations in genes that encode chromatin regulators, which in turn can cause epigenetic dysregulation. Epigenetic mechanisms modify gene expression without causing any change in cellular DNA. They do so through four mechanisms; methylation, histone modification, microRNAs and other noncoding RNAs such as long noncoding RNAs (116).

Inactivating mutations and deletions in *PTEN* (a tumour suppressor) are rare in sarcoma, so it's lack of expression may be explained by epigenetic regulation. Indeed, the promoter region of *PTEN* has found to be hypermethylated at CpG sites in soft tissue sarcomas, resulting in gene silencing (116). Taylor et al. reported concurrent sequencing of tumour genomes, exomes and transcriptomes to delineate the molecular landscape of primary and locally recurrent DDLS. Interestingly, 24% of DDLS methylomes revealed alterations in the differentiation pathway gene *C/*EBPα. Treatment with demethylating agents restored *CEBPA* expression, was anti-proliferative and pro-apoptotic in vitro and reduced tumour growth in vivo (78). This was the first illustration of a potential role for demethylating agents in liposarcoma.

Figure 1.5 illustrates pathways active in both WDLS and DDLS.

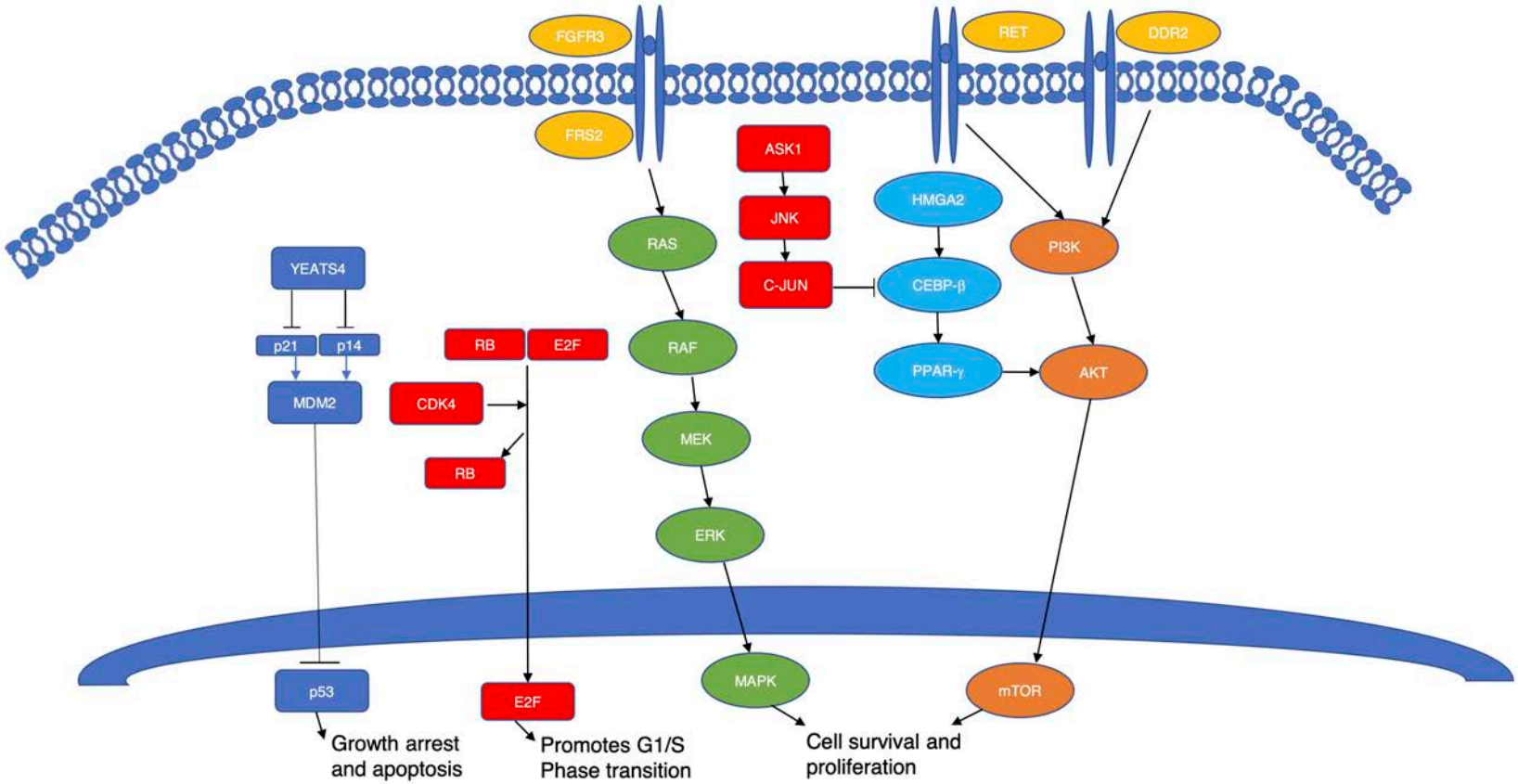

Figure 1.5: Pathways active in WDLS/DDLS. Pointed arrows denote upregulation, whilst flat-headed arrows denote down regulation. *MDM2* is overamplified in WDLS/DDLS and blocks transcription of p53. *YEATS4* increases levels of *MDM2* by supression of p21 and p14. *CDK4* is commonly found to be co-amplified with MDM2, and phosphorylates Rb, dissociating it from the pRb-E2F complex, allowing E2F to drive G1/S transition. *HMGA2* is amplified in both WDLS and DDLS and drives CEBP-β mediated expression of PPAR-γ promoting adipogenic differentiation. The MAPK/ERK pathway is activated in WDLS and DDLS by *FRS2* and *FGFR3* amplification respectively. The PI3K pathway is also active; in WDLS increased levels of RET drive it, whilst *DDR2*

amplification in DDL achieves the same. Both pathways drive cell survival, proliferation and angiogenesis.

1.10.3 Genomic drivers of dedifferentiation

The genomic changes associated with progression from WDLS to DDLS are complex and poorly understood. Both subtypes harbour the neochromosomes and oncogenes previously explored. Nevertheless, DDLS tumours must possess additional mechanisms to become high grade, more cellular and aggressive. In general, DDLS exhibit more complex chromosomal aberrations (106).

A major element of dedifferentiation is the loss or downregulation of adipogenesis, which results in a non-lipogenic tumour mass which can be difficult to distinguish histologically. Genes involved in adipocyte metabolism such as *LIPE*, *PLIN* and *PLIN2* are amongst those uniquely absent in DDLS, suggesting a loss of ability to act like fat in these tumour cells (117, 118). DDLS also tends to have a more rearranged genome, especially the 12q13-15 amplicon which contains genes which control adipocyte differentiation such as *CPM* and *HMGA2.*

Mariani et al. analysed the expression levels of key genes in adipogenesis in 16 liposarcomas, of which 14 were retroperitoneal. By comparing expression levels through RT-PCR, in tumours that overexpressed *JUN* with those that didn't, they found that the *C/EBP*β transcriptional network was impaired in the group that overexpressed *JUN* (119). They concluded that dedifferentiated

tumours are committed to differentiate into adipocytes, but failure to achieve this is driven by *JUN* overexpression, which would be in keeping with the non-adipogenic nature of DDLS. High amplification levels of *JUN* (>16 copies) have also been correlated with decreased disease free survival, corroborating the oncogenic role of *JUN* (87).

Beird et al. compared the genomic landscape of synchronous WDLS and DDLS in 17 paired tumours. There were three main findings: firstly, a low somatic mutational burden across both tumour types. Secondly, the presence of shared somatic mutations, albeit in low number. Finally, they identified a significantly larger number of gene fusions and copy number alterations in DDLS when compared to WDLS (120). The higher copy number alterations and gene fusions in DDLS were attributed to the result of increased break-fusion bridge cycles. They further suggested that the low number of shared somatic mutations meant tumours derived from a common ancestral clone, but diverged early in their development. They hypothesised that the capacity to dedifferentiate is determined early in the disease.

Horvel et al. compared 29 paired WD/DDLS tumours by array-based comparative genomic hybridisation. The analysis segregated all but one pair together, and the genotypic similarities between the components implied that genetic changes preceded phenotypic progression *(121)*. This further supports the argument that the background genotype of some WDLS already possess the capacity to dedifferentiate.

1.11 Novel therapies in retroperitoneal liposarcoma

Due to the poor rates of response to conventional chemotherapy in RPLS there is great interest in finding targeted therapies by understanding the genetic drivers of tumourigenesis. An example of the successful targeting of a specific molecular pathway in sarcoma is the use of imatinib, a tyrosine kinase inhibitor used in gastrointestinal tumours (GISTs) with a mutation in the C-KIT gene. This significantly improved overall survival in the metastatic setting and is now used both neo-adjuvantly and adjuvantly (122).

Nutlins are small molecule antagonists that inhibit the binding of *MDM2* to p53 and restore p53 activity and apoptosis, making them an attractive therapeutic target. Muller et al. showed downstream p53 dependent transcription and apoptosis in liposarcoma cell lines treated with Nutlin-3a (123). The transition from in vitro to in vivo success has proved difficult.

Only two studies have specifically reported on *MDM2* antagonists in patients with RPLS. Ray-Coquard et al. reported a proof of mechanism study for MDM2-antagonist RG7112 in 20 patients with WD/DDLS. Post treatment biopsies had increased concentrations of p53 compared with baseline, and every patient had stable disease, but each of the patients in the trial had at least one adverse side effect, of which twelve were serious (124). Wagner et al. in 9 patients with advanced WDLS found 4 patients had prolonged stable disease with one prolonged partial response. 1/3 of the patients experienced serious adverse effects (125) .

In pre-clinical work, Jung et al. performed liquid biopsies on patients being treated with MDM2 inhibitors for DDLS. They demonstrated *TP53* mutations appearing in circulating cell-free DNA, with the mutation burden increasing over time correlating with a change in tumour size (126). This was the first clinical demonstration of a *TP53* mutation in response to an MDM2 antagonist, representing an escape mechanism for *MDM2* amplified cells. Resistance mechanisms combined with side effects have precluded the long-term use of MDM2 antagonists in WD/DDLS. Reflecting on this, it is possible that single agent use, i.e monotherapy is not enough to sustain tumour suppression over time. This should encourage researchers to search for new driver genes in liposarcoma, which could be targeted alongside MDM2.

A phase II trial of Palbocicilib, a CDK4 antagonist has been reported in advanced WD/DDLS. 30 patients (25 DDLS, 5 WDLS) were deemed eligible with a progression free survival of 66% at 12 weeks (127). Although encouraging, the results may be biased by the more indolent behaviour of WDLS in comparison to

DDLS (128). Ongoing research is investigating predictive markers to Palbociclib (21).

Four trials using CDK4 antagonists have been reported in retroperitoneal liposarcoma, two using palbociclib, one using ribociclib and the last a pan-*CDK* inhibitor – flavopiridol – in combination with doxorubicin (127, 129-131). The largest of these studies included 15 patients with WDLS (129). The 12-week progression-free survival rate was 57.2% with one complete response, although haematological adverse events were common. Ongoing research is investigating predictive markers to palbociclib (21).

Peroxisome proliferator activator receptors (PPAR) are critical regulators of adipocytic differentiation. *PPAR-*γ has a critical role in the terminal differentiation of adipocytes, and *PPAR-*γ agonists have been shown to induce adipocytic differentiation in liposarcoma cell lines and demonstrate anti-tumour activity (132). Despite this results in patients have been disappointing. Debrock et al. reported

results from a phase II trial of rosiglitazone in 12 patients with DDLS and myxoid liposarcoma. No clinical response was seen, with a mean time to disease progression of six months (133). The use of the antiviral Nelfinavir, tyrosine kinase receptors and immunotherapy are also being studied, but are very much in their infancy (134).

It would be fair to say that drug therapies in WD/DDLS are uninspiring at best and for that reason surgery remains the gold standard for the foreseeable future. Nevertheless, translation of targeted molecular therapeutics has been successfully demonstrated in GIST. With emerging technologies such as next-generation sequencing, CRISPR, nanopore and single cell sequencing anticipated to generate new targets - something this body of work aims to utilise – improvements in care may be possible.

To drive this improvement in care, researchers have looked to other genes outside the 12q13-15 amplicon, in an attempt to avoid the toxicity and resistance caused by MDM2 and CDK4 inhibitors. In 2015, work began in the Begg's laboratory to

perform comprehensive molecular characterisation of WDLS as part of the pre-pilot for the 100,000 Genomes Project. This generated two genes of interest (*FOXD4L3* and *SDHA*) which are briefly discussed, before a more in-depth discussion in chapters 2-4. In addition to these genes, we also adopted a broader approach using CRISPR-cas9 technology and single cell sequencing in the hope of identifying new targets.

Table 1.3 provides an up-to-date evidence summary of genomic drivers and related clinical trials.

Table 1.3: Evidence summary of genomic drivers of WD/DDLS subtypes and related clinical trials

Histology	Genes	Associated Pathways	Genomic Aberration	Targeted Drug	Trials
WDLS/DDLS	MDM2	MAPK/ERK + PI3K/AKT/mTOR	Amplified 20/20 (100%)	MDM2 Antagonist RG7112	EudraCT 2009-015522-10
	CDK4		Amplified 20/20 (100%)	CDK4/6 Inhibitor - Palbociclib	NCT01209598
	HMGA2		Amplified and rearranged 38/38 (100%)	HMGA2 Inhibitor - Netroposin	Not in RPL
	FRS2		Amplified 98/101 (97%)	FGFR selective inhibitor NVP-BGJ-398	Not in RPL
	DDR2	PI3K/AKT/mTOR	Amplified 6/56 (11%), fused in 1 sample	Kinase inhibitors - Imatinib/Nilotinib	Not in RPL
	PPAR-γ		Expressed in 44/54 (81.5%)	PPARγ ligand - Pioglitazone	NCT00004180
DDLS	YEATS 4	P14/P21/MDM2	Amplified 30/40 (75%)	NA	NA
	FGFR3	PI3K/AKT/mTOR + RAS	Amplified in 2/56 (4%)	Pan-FGFR Inhibitors	Not in RPL
	RB		Loss of Heterozygosity 16/27 (60%)	NA	NA

1.12 FOXD4L3

Preliminary work by Professor Andrew Beggs and Dr. Jo Stockton performed whole genome sequencing, RNA sequencing and methylation array analysis on 8 WDLS samples (135). A recurrent mutation within exon 1 of Forkhead Box D4 Like 3 (*FOXD4L3*) was observed (chr9: 68,303, 273A>T); c.322A>T;pLys108Ter.). *FOXD4L3* is a forkhead box associated transcription factor with the observed mutation (chr9: 68,303, 273A>T) lying in the DNA binding region for Forkhead transcription factors (136). Protein level expression of FOXD4L3 was confirmed in normal tissue via IHC but was absent in tumour tissue. Furthermore, the Gene Expression Omnibus suggested there is mRNA expression for FOXD4L3 unlike the other paralogs of the genes (137).

These findings required further investigation for two reasons. Firstly, the observed mutation occurs in a single exon gene, in the functional forkhead box. Secondly, somatic mutations are rare in soft-tissue sarcoma and this gene could represent a driver of tumourigenesis in WDLS.

1.13 SDHA

The aforementioned preliminary work also found a recurrent gene fusion (5/8 samples) between Succinate Dehydrogenase A (*SDHA)*. A third of soft tissue sarcomas carry recurrent, chromosomal translocations (138). Many of these generate fusion proteins, which act as transcription factors, driving cancer initiation

and growth. *SDHA* plays a critical role in the tricarboxylic acid (TCA) cycle, and loss of function may prevent the normal function of the cycle. The suggestion that *SDHA* fusions may play a role in WDLS required further exploration both in vitro and in vivo.

1.14 CRISPR and Single Cell RNA sequencing

FOXD4L3 and *SDHA* are unexplored genes in liposarcoma, and warrant investigation. Nevertheless, they could easily be passengers in a wider genomic rearrangement and have no direct role in tumourigenesis. Newer methods of interrogating the liposarcoma genome are therefore encouraged.

CRISPR/Cas9 is a simple two-component system used for targeted gene editing. The first component is the Cas9 protein, which contains an endonuclease which generates double-stranded breaks in the target DNA (139). The second component of effective targeted gene editing is a single guide RNA (sgRNA) that enables its anchoring to Cas9 and a 20 base pair spacer sequence complementary to the target gene.

The two-component system then relies on either of two DNA repair pathways: non-homologous end-joining (NHEJ) or homology-directed repair (HDR). NHEJ occurs much more frequently and involves random insertion and deletion of base pairs at the cut site. This error-prone mechanism usually results in frameshift mutations, often creating a non-functional polypeptide (140).

This pathway has been particularly useful in genetic knock-out experiments and functional genomic CRISPR screens.

Genome-wide CRISPR/Cas9 knockout screens offer a broader approach, where every gene is knocked out using a pool of sgRNA's, and a selection pressure applied – either positive or negative. This allows for the unbiased identification of multiple genes, essential for survival and proliferation.

To date, no CRISPR screens in WDLS have been published. By performing a CRISPR screen, we further explored the liposarcoma genome in an attempt to uncover new targets.

Another new technology that can yield new targets, is single cell RNA sequencing (scRNA-seq). Compared with traditional sequencing this allows the detection of heterogeneity among individual cells. It also allows the mapping of the tumour microenvironment, which is poorly understood in RPL. Combining this with a genome-wide CRISPR/Cas9 knockout screen has the potential to lead to new avenues of discovery.

1.15 Summary

Since the advent of next generation sequencing, the number of studies exploring the molecular landscape of retroperitoneal liposarcoma has increased, albeit at a slower rate than epithelial counterparts. This introduction has summarised the hallmark genes of WDLS/DDLS such as *MDM2*/*CDK4*. As toxicity and resistance have

hampered the progress of drugs inhibiting these genes, it is critical the wider genome is interrogated for progress to be made. It is possible that *FOXD4L3* and *SDHA* play a role in liposarcomagenesis, but require a basic interrogation before therapeutics can be contemplated. New approaches through technologies such as CRISPR can afford a much wider appreciation of tumour biology, in this rare cancer, and will hopefully generate meaningful data.

1.16 Hypothesis

1) FOXD4L3 is recurrently mutated in WDLS
2) *SDHA* fusions are present in WDLS
3) *SDHA* fusions lead to a loss of function and a truncated TCA cycle

1.17 Aims and objectives

1) Validate the presence of the somatic mutation (chr9: 68,303, 273A>T) >T) in FOXD4L3 in a larger patient cohort.
2) Demonstrate *SDHA* fusion in vitro.
3) Explore *SDHA* functionality in vivo, using metabolic tracing.
4) Identify novel driver genes in WDLS through the combination of CRISPR and single-cell RNA sequencing.

Chapter 2: Materials and Methods

2.1 Samples

2.1.1 Fresh tissue samples

Liposarcoma tissue samples were collected prospectively from patients treated at the Queen Elizabeth Hospital, Birmingham (QEH) between 2017 and 2019. Storage of tissue samples was in a registered tissue bank - the Human Biomaterials Resource Centre (HBRC), and access to it was granted via existing ethics (Beggs: 13-121.) Samples were taken from macroscopically well differentiated tumour by the pathologist immediately after surgery. In certain cases, macroscopically well differentiated samples were taken from a tumour with a separate dedifferentiated component, and this is stated where relevant. Histologically confirmed tumour-free matched adjacent tissues taken from healthy fat were also available for paired analysis. Samples were stored in a -80° freezer.

2.1.2 FFPE samples

Historical formalin-fixed paraffin-embedded (FFPE) samples were available via existing ethics and found using a prospectively maintained Sarcoma Unit database. 32 x 8 μm FFPE scrolls of WDLS and matched normal (8 paired, 10 normal only and 5 tumour only) were obtained for RNA extraction.

2.1.3 Blood samples

In Chapter 3 experimental work is undertaken which involves a comparison with healthy controls. DNA samples of healthy controls were supplied by Professor Ian Tomlinson as Primary Investigator of the Colorectal Tumour Gene Identification (CORGI) study. DNA had been extracted from blood and stored at -80 ˚C previously. The ethics for this was covered by UK National Cancer Research Network Multi-Research Ethics Committee (17/SC/0079). Blood samples taken prospectively from WDLS patients was done so under existing ethics, and with written permission from the patients.

2.2 Nucleic Acid Extraction

2.2.1 DNA extraction from fresh tissue and cell line

Extracting nucleic acids from liposarcoma specimens was found to be challenging for several reasons. Even though the raw amount of tumour available was always high, the tumour itself is made up of a low number of large cells, which limits the total amount of extractable DNA and RNA. The fatty nature of the tumours also meant that spin columns were more likely to become blocked and reduce the quality of any eluent. Lastly, lipid-rich tissues can interact with the chemistry of DNA isolation buffers – likely a result of protocols being designed for epithelial tissues. Where specific adipose extraction kits were available these were used. The overall consequences were low nucleic acid yields, which tended to be of poorer quality (i.e fragmented).

DNA extraction from fresh tissues and the cell line was performed using an AllPrep DNA/RNA Mini Kit (Qiagen, Germany). 1 x 10^7 cells were harvested as a cell pellet. 10 μL of β-mercaptoethanol (PanReac, Barcelona) and 650 μL of Buffer RLT Plus were added to 25 μL of tissue sample or cell pellet. A 5mm stainless steel bead was added and the samples disrupted using a Tissue Lyser II (Qiagen, Germany) at 20 Hz for 2 minutes. The lysate was then centrifuged for 3 minutes at maximum speed.

Homogenised lysate was transferred to an AllPrep DNA spin column placed in a 2 ml collection tube. This was then centrifuged for 30 s at 8000 x g. 500 μL of Buffer AW1 was added to the spin column and centrifuged for 15 s at 8000 x g to wash the spin column membrane. The flow through was then discarded. 500 μL of Buffer AW2 was then added to the spin column and then centrifuged at full speed for two minutes. The spin column was placed in a new 1.5 mL collection tube. 100 μL of Buffer EB was added directly to the spin column membrane and then centrifuged for one minute at 8000 x g to elute the DNA. DNA was then quantified and stored at - 20° until required.

2.2.2 RNA extraction from fresh tissue

RNA extraction was performed using the RNeasy Lipid Tissue Mini Kit (Qiagen, Germany). Fresh tissue underwent scalpel dissection. 75 mg of tissue was added to 1 mL of QIAzol lysis Reagent in a collection tube before disruption at 20 Hz for 3 minutes using the TissueLyser II. 200 μL of chloroform was added to each sample and then centrifuged at 12,000 x g for 15 minutes at 4˚. This created an upper

aqueous phase away from the lower red phenol phase. The aqueous phase was removed, and 1 volume of 70% ethanol added.

700 µL of the sample was then transferred to an RNeasy Mini spin column in a 2 mL collection tube. This was centrifuged for 15 s at 8000 x g, and the flow-through discarded. This was repeated with buffer RW1. 500 µL of Buffer RPE was added to the RNeasy column, and centrifuged for 2 mins at 8000 x g. The RNeasy column was placed in a new 1.5 mL tube, 50 µL of RNase-free water added, and centrifuged for 1 minute at 8000 x g.

The eluted volume of RNA then underwent a DNAse treatment using TURBO DNA-free™ Kit (ThermoFisher, USA) as per manufacturer's instructions (141). The final eluted RNA samples were then quantified and stored at -80˚ C.

2.2.3 RNA extraction from FFPE

RNA extraction from FFPE was performed using the truXTRAC™ FFPE microTUBE RNA Kit (Covaris, USA).

A screw-cap microtube was opened and 110 µL of RNA Lysis Buffer and an 8 µm FFPE scroll added. The samples were processed using the settings provided in Table 2.1 to dissociate the paraffin and rehydrate the tissue.

Table 2.1: Paraffin removal and rehydration settings

System	Duty Factor	Peak Incident Power	Cycles per burst	Treatment Time	Temperature
E220	10%	175 Watts	200	5 minutes	20˚C

The microtube was opened and 10 µL of PK solution added. The sample was then processed as before, except for 10 seconds. The sample was incubated at 56˚C for 15 minutes to facilitate proteinase K digestion. A further incubation at 80˚C for 60 minutes was undertaken to reverse formaldehyde crosslinks. The sample was transferred to a clean 1.5 mL tube (Eppendorf, Germany) and centrifuged at 15,000 x g for 15 minutes. The supernatant was transferred to a new microcentrifuge tube to undergo residual DNA removal.

A master mix was created consisting of 13 µL of MnCl2 solution, 7 µL Dnase buffer and 10 µL of Dnase I. This was added to each sample, mixed gently by pipetting and then incubated at room temperature for 15 minutes. The samples then underwent purification using a spin column protocol as per manufacturer's instructions (142). Eluted RNA was quantified, then stored at -80˚C for long-term storage.

2.2.4 DNA extraction from blood

Blood samples were taken in a 10 mL EDTA-treated Vacutainer (BD, USA) collection tube. Collection tubes were centrifuged for 10 minutes at 2,500 x g to create a buffy coat interface between the lower red blood cell layer and upper plasma layer. The

buffy coat was extracted and added to Well 1 of the Maxwell 16 Blood DNA Cartridge (Promega, USA). A plunger was placed into Well 7. 300 μL of Elution Buffer was added to an elution tube. DNA-Buffy Coat protocol was then ran on the Maxwell RSC Instrument. The eluted sample was removed, underwent DNA quantification and then stored at -20° until required.

2.3 Nucleic Acid quantification

2.3.1 QuBIT fluorometer

DNA and RNA quantification were carried out by the QuBit Fluorometer 3.0 (Thermofisher, USA). A 1:200 quantification assay was made by adding 199 µL buffer to 1 µL dye. Two standards were run first, with a 190:10 ratio of quantification assay to standard. Samples were run with a 199:1 ratio. This was read by the instrument and a ng/ µL sample concentration given.

2.3.2 Tapestation

DNA and RNA quality and fragment length were assessed using the Agilent Tapestation 2200 (Agilent, USA). Screen tapes were chosen depending on nucleic acid concentration. For DNA analysis, a ladder was prepared by mixing 3 µL D1000 Sample Buffer with 1 µl of D1000 ladder. Each sample was prepared by mixing 3 µL D1000 Sample Buffer with 1 µL DNA sample. For RNA analysis, an electronic ladder was run. Each sample was prepared by mixing 5 µL RNA sample buffer and 1 µL

RNA. This was heated at 72 ˚C for three minutes and placed on ice for two minutes. The samples were then loaded into the instrument and analysed.

2.3.3 Gel electrophoresis

DNA gel electrophoresis was performed using a 2% agarose gel to separate the DNA fragments by size and ensure a successful PCR reaction. A solution was prepared using 10 X Tris-borate-ethylenediaminetetraacetic acid (TBE) buffer (Geneflow, UK), 2% w/v agarose (Lonza, USA) and 1:10000 Gel Red Nucleic Acid Stain (Biotium, USA) dissolved in deionised water. This was warmed to solubilise, poured into a gel cassette and left for 15 minutes to set. Once set, the gel was added to the gel box, and covered with 10 X TBE. 3 μL of sample was mixed with 3 μL of loading buffer (Invitrogen, USA) and loaded onto the gel. Samples were run alongside a 1Kb Plus DNA Ladder (Thermoscientific, USA). Electrophoresis was performed for 30 minutes at 120 V. The resultant gel was read on a UV reader (U:genius 3, Syngene,UK).

2.4 Cell culture

A well differentiated retroperitoneal liposarcoma cell line (93T449, 68, female, Caucasian), which had previously undergone confirmation for MDM2 amplification, was purchased from ATCC. Cells were grown in T75 flasks (Corning, UK) in a humidified incubator at 37˚C supplemented with 5% CO2. Cells were maintained with media changes every three days and split at 80% confluence. Cells were grown as adherent monolayers. Cell lines never reached more than 15 passages. Media was RPMI 1640 with L-Glutamine (Life technologies, UK) supplemented with 10%

fetal bovine serum (FBS) (Sigma-Aldrich, USA) and 1% Penicillin-Streptomycin (Life Technologies, USA). The cell line was shown to be free from mycoplasma infection using an EZ-PCR Mycoplasma Test Kit (Biological Industries, Israel).

To split cells media was removed and cells washed with Phosphate Buffered Saline (PBS, Sigma-Aldrich, USA). Following this, 2 mL 0.25% Trypsin-EDTA solution (Life-technologies, UK) was added and incubated at 37° for five minutes. Cell detachment was confirmed under the microscope. 4 mL of maintenance media was added to trypsinised cells and mixed prior to centrifuging at 200 x g for five minutes. Cell pellets were re-suspended in media and diluted as appropriate. They were then added to a fresh flask and returned to the humidified incubator.

To perform cell counts and assess number of viable cells, 15 μL suspended cells were mixed with 15 μL of Trypan Blue (Sigma-Aldrich, UK) and counted using a TC20 ™ Automated Cell Counter (Bio-Rad, USA) taking the average of two alive values.

2.5 DNA sequencing

2.5.1 Primer design

Custom primers were designed using Primer3Plus software (A. Untergasser, Germany) and tested using the In-Silico PCR tool on UCSC Genome Browser (143). Primers were obtained from Sigma-Aldrich, USA.

Difficulties in primer design existed because FOXD4L3 sits in an ancestral duplication site. It has one paralog in FOXD4L6, from which it differs by three base pairs. This did not affect nanopore sequencing as long-range primers generate an amplicon long enough to distinguish between paralogs. As Sanger sequencing can only produce a much shorter amplicon, Clustal Omega (EMBL-EBI, UK) sequence alignment was used to align FOXD4L3 and FOXD4L6 and identify non-identical sequences (see Figure 2.1).

chr9:68,302,889-68,302948

cagcccgcccaggccaggtgatcggccgccacatcccctgcgactgaagcacctgctcc 300

chr9:41,128,387-41,128,446

cagcccgccccaggctaggtgatcggtcgccacatcccctgcgactgaagcacctgctcc 300

********** ********** **********

* Identical residues

Figure 2.1: Clusta Omega analysis to identify non-identical sequences in FOXD4L3 (Top) and FOXD4L6 (Bottom). Non-identical residues are shown through a lack of asterisk. The forwards primer was then designed to incorporate these residues. Clustal forwards primer shown in yellow.

Sequencing primers were then designed that would only amplify FOXD4L3. A separate pair of nested primers were designed to flank chr9:68,303, 273 (see Figure 2.2). Primers are shown in Table 2.2.

```
1    CATTCATCCCAGGCTTCCAGCTCAGCCCGCCCCAGGCCAGGTGATCGGCCGCCACATCCCCTGCGACTGA
71   AGCACCTGCTCCTCCATGAACCTGCCAAGAGCTGAGCGCCTTCGCTCCACACCGCAGCGCAGCCTCCGGG
141  ACTCCGATGGGGAAGACGGTAAAATCGATGTCCTGGGAGAGGAGGAAGATGAAGACGAGGTGGAAGACGA
211  GGAGGAGGCGGCGAGCCAGCAGTTCCTAGAGCAGTCGCTCCAGCCGGGGCTGCAGGTGGCCCGGTGGGGC
281  GGGGTTGCGCTTCCCCGAGAGCACATCGAGGGCGGCGGCGGCCCGAGCGACCCCTCAGAGTTTGGCACCA
351  AGTTCAGGGCACCGCCAAGGTCTGCGGCGGCCTCTGAAGATGCCCGGCAGCCGGCAAAGCCCCCCTACTC
421  GTACATCGCGCTCATCACCATGGCCATCCTGCAAAACCCGCACAAGCGCCTCACGCTCAGCGGCATCTGC
491  GCCTTCATTAGTGGCCGCTTCCCCTACTACCGCCGCAAGTTCCCCGCCTGGCAGAACAGCATCCGCCACA
```

Figure 2.2: Sanger nested primers for chr9: 68,303, 273 with 166 base pair amplicon (Forwards Primer: Yellow, Somatic mutation/ SNP: Red, Reverse Primer: Blue). As Sanger sequencing has a lag before starting to read, primers are designed more than 50 base pairs from the base pair of interest.

Table 2.2: List of primers for DNA sequencing

Primer Type	Primer Direction	Sequence	Annealing Temp (˚C)	Amplicon Size (bp)	GRCh38 Coordinates
Nanopore sequencing	Forwards	GCCCAGAAGACCATCATCCC	60	2550	chr9: 68,302,649 -
	Reverse	ATGGTTTTCCAGCTTGTTGAGG	45.45		chr9: 68,305,198
Sanger sequencing	Forwards	CCAGGCCAGGTGATCGG	67.9	2301	chr9: 68,302,898 -
	Reverse	ATGGTTTTCCAGCTTGTTGAGG	66		chr9: 68,305,198
Sanger nested	Forwards	AGTTTGGCACCAAGTTCAGG	60.1	166	chr9: 68,303,205 -
	Reverse	CCACTAATGAAGGCGCAGAT	60.2		chr9: 68,303,370

2.5.2 Nanopore sequencing

DNA Samples were quantified using a Qubit fluorometer. Samples were then normalised to the lowest samples concentration. As amplicon sequencing was being undertaken, the first stage required a PCR reaction using the nanopore sequencing

primer from Table 2.2. A master mix was prepared containing template DNA, 2 µL of forwards and reverse primers at 0.4 µM, 25 µL LongAmp Taq 2X Master Mix (New England Biolabs, USA) and nuclease-free water up to a total of 50 µL. It was then quickly transferred to a preheated thermocycler (94°C) where it underwent PCR as per Table 2.3.

Table 2.3: Thermocycling conditions for PCR

Step	Temp (°C)	Time
Initial Denaturation	94	30 s
30 Cycles	94 45-65 6	30 s 60 s 50 s per kb (2.5)
Final Extension	65	10 min
Hold	4-10	

Products of the PCR reaction were then run on an 2% agarose gel to separate the DNA fragments by size and ensure a successful PCR. The PCR products were then quantified.

2.5.2.1 DNA repair and end-prep

PCR products were standardised to the sample with the lowest quantity (6.7 ng/ µL) by a relative addition of nuclease-free water (Omega Bio-tek, USA). The volumes were further adjusted to 48 µL with nuclease-free water. DNA was then prepared for

nanopore sequencing via steps summarised in the Oxford Nanopore Ligation sequencing protocol (144). An abridged summary of the steps is given here. Reagents were supplied by Oxford Nanopore Technology, UK unless stated.

In a 0.2ml thin-walled PCR tube the following mix was prepared (Table 2.4).

Table 2.4: End Repair Preparation Master Mix

Reagent	Volume
DNA	48 μL
NEBNext FFPE DNA Repair Buffer	3.5 μL
NEBNext FFPE DNA Repair Mix	2 μL
Ultra II End-prep reaction buffer	3.5 μL
Nuclease free water	3 μL
Total	60 μL

The end repaired preparation was incubated using a Tetrad 2 thermal cycler (MJ Research, USA) at 20°C for 5 minutes and 65°C for 5 minutes.
60 µL of resuspended AMPure XP beads (Beckman Coulter, USA) were added to the end-prep reaction and pelleted on a magnetic rack. Two bead washes were performed with 70% ethanol, before resuspension in 25 µL of nuclease-free water. This was then pelleted on a magnet and 25 µL of eluate removed. 1 µL of end-prepped DNA was then quantified.

2.5.2.2 Native barcode ligation

A unique barcode was assigned to each sample. 500 ng of each end-prepped sample in 22.5 µL of nuclease-free water was then added to 2.5 µL of native barcode and 25 µL of Blunt/TA Ligase Master Mix.

50 µL of resuspended AMPure XP beads were added to the reaction and mixed by pipetting. Samples were then pelleted and washed as previously, before resuspension in 26 µL of nuclease-free water. After further pelleting, 26 µL of eluate was removed, and 1 µL quantified as previously. Next 700 ng of pooled sample was diluted in 50 µL of nuclease-free water.

2.5.2.3 Adapter ligation and clean-up

Adapter ligation was performed by adding 20µL of barcode adapter mix, 20 µL of NEBNext Quick Ligation Reaction Buffer and 10 µL of Quick T4 DNA Ligase to the 700 ng pooled sample.

40 µL of resuspended AMPure XP beads were added to the adapter ligation reaction and pipetted. Samples were placed on a magnetic rack, beads pelleted, and the supernatant removed. The beads were washed twice with 250 µL of S Fragment Buffer before the supernatant was discarded. The sample was removed from the magnet and the pellet resuspended in 15 µL Elution Buffer (EB). The beads were

pelleted one final time, before 15 μL of eluate was removed. 1 μL of adapter ligated DNA was then quantified, with a result of 6.76 ng/μL.

12 μL of DNA library was prepared and loaded in the priming port of the MinION flow cell as per protocol, before the sequencing run was commenced. Sequencing outputs were in FASTQ format and analysis performed by Professor Andrew Beggs. Raw sequence FAST5 files were demultiplexed and converted to FASTQ using the Guppy Basecaller. Called FASTQ files underwent two rounds of polishing with Canu followed by a round of polishing with Medaka. Sequence data was aligned using minimap2 followed by variant calling using FreeBayes (command line freebayes -C 2 -0 -O -q 20 -z 0.10 -E 0 -X -u -p 2 -F 0.6). Raw allele counts at the desired sites were outputted with the samtools command.

2.5.3 Sanger sequencing

DNA samples were diluted to 20 ng in 3 μL using nuclease-free water, then mixed well and spun down. A master mix of 4 μL of 5x Phusion HFB, 1 μL of 100 μM forwards and reverse primers, 0.4 μL dNTP, 0.2 μL of Tac Polymerase and 10.4 μL of nuclease-free water was made up and added to samples in a PCR strip tube. Except for primers, reagents were from New England Biolabs, USA. Forwards and reverse primer information is contained in Table 2.2

Samples were transferred to a thermocycler and incubated using the conditions in Table 2.5. The annealing temperature was set at 63˚C based on the average annealing temperature of the two primers minus 3˚C. The extension time was set at

69 seconds based on 30 seconds multiplied by the size of the product in kilobases (2.301 x 30). PCR products were run on a 2% agarose gel to confirm a successful PCR and then quantified.

Table 2.5: PCR conditions for Sanger sequencing

Step	Cycles	Incubation temperature (˚C)	Incubation time
Initial denaturation	1	98	30s
Denaturation	25	98	10s
Annealing	25	58	20s
Extension	25	72	69s
Final Extension	1	72	5 mins

The nested primer was diluted to 3.2 pmol by adding 1.6 µL of primer (100 µM) to 48.4 µL of nuclease-free water. 7 ng of PCR product, 1µL of 3.2 pmol sequencing primer and 8 µL of nuclease-free water was then mixed and added to 10 µL of Big Dye®Terminator Mix (Thermofisher, USA). Samples were then loaded on to the Sanger sequencing platform and sequencing commenced.

2.6 Protein detection by western blotting

2.6.1 Protein extraction

100 mg tissue was dissected in Radioimmunoprecipitation assay buffer (RIPA buffer, Thermo-Scientific, USA) with 100 x protease inhibitor (Sigma-Aldrich, USA) 100 mg/100µL for five minutes, and then left for thirty minutes at room temperature. The

sample was centrifuged at 10,000 x g for 10 minutes creating a fatty layer on top of the solution. The solution below this was removed and centrifuged twice more to generate a lysate for each sample. Lysates were then boiled for five minutes at 100°C. Protein was quantified using the BCA Protein Assay Kit (Thermo-Scientific, USA) as per manufacturer's instructions (145).

2.6.2 Western Blotting and densitometry

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to separate proteins using the Mini Trans-Blot Cell system (Bio-Rad, USA). A 12% loading gel and 5% stacking gel was prepared using the reagents in Table 2.6.

Table 2.6: Composition of SDS-PAGE

	12% Loading gel (mL)	5% Stacking gel (mL)
H_20	6.6	5.5
1.5M TRIS (pH 8.8)	5	1
10% SDS	0.2	0.08
10% Ammonium Persulfate	0.2	0.02
TEMED	0.008	0.008
30% acrylamide mix	8	1.3

The loading gel was poured between the two glass plates, leaving enough space from the top of the plates for the stacking gel and a comb. The gel was left for ten minutes to set and then the stacking gel was added. Either one or two ten-well combs were added to the top depending on the number of samples being run.

Quantified samples were normalised on each gel for protein quantity and volume and mixed with 1:6 loading buffer volume (Pierce Lane Marker Reducing Sample Buffer, Thermo Scientific USA.) Any remaining volume was made up with nuclease-free water. Samples were then boiled at 100°C for two minutes, briefly centrifuged and loaded onto the gel. MagicMark XP Western Protein Standard (Thermo-Scientific, USA) was run alongside samples to confirm molecular weight. Electrophoresis was performed at 80 V for ten minutes and then 106 V until the protein was sufficiently resolved. Transfer onto an Amersham PVDF membrane (GE Healthcare, Germany) was performed at 90 V for ninety minutes.

The efficiency of transfer was determined by post-transfer staining of the gel. Ponceau S Staining Solution (P7170, Sigma-Aldrich Ltd, UK) was added to the gel and incubated at room temperature for an hour with gentle agitation using a rocking platform shaker (VWR, UK). The membrane was then rinsed in distilled water, dried and imaged using a V700 Photo Scanner (Epson, Japan).

2.6.2.1 Blocking, addition of primary/secondary antibody

Once membrane transfer was complete, membranes were blocked to minimise background staining due to non-specific antibody binding. Blocking was performed with 5% Dried Skimmed Milk powder (Marvel, UK) in 1x Tris-buffered saline/Tween (TBST) (Sigma-Aldrich, USA) for one hour with gentle agitation. Proteins were normalised to a loading control protein expression (β-Actin) from the same membrane. As the proteins of interest (SDHA and SDHB) differed in sizes to the loading control protein the membrane was cut after staining and each piece was analysed separately with the desired antibody. All antibodies are listed in Table 2.7 with dilutions stated.

Table 2.7: Antibodies used for Western blotting

Target	Isotype	Manufacturer	Diluent	Dilution
β-Actin	Mouse	Sigma	0.05% Tween/5% Milk	1:5000
SDHA	Mouse	Abcam	0.05% Tween/5% Milk	1: 1000
SDHB	Mouse	Abcam	0.05% Tween/5% Milk	1: 1000

Due to the relative sizes of proteins, actin (42 kDa) was tested first, then SDHA (73 kDa) and finally SDHB (32 kDa). Primary antibodies incubated overnight at 4˚C with gentle agitation. The milk solution containing the primary antibody was poured off the

membrane, and then rinsed three times with 50 ml PBS + 0.02 % Tween for 5 minutes. The membrane was transferred to a secondary antibody conjugated to horseradish peroxidase and left for an hour at room temperature. This was followed by three washes as described previously.

The membrane was covered with enhanced-chemiluminescence substrate EZ-ECL (Biological industries, Israel) for one minute, before removing and covering the membrane. 18 x 24 cm X-Ray films (SLS, UK) were placed inside cassettes with the covered membrane for a defined time period. They were then developed using the Protec ECOMAX X Ray Film Processor (Raytech Diagnostics, Canada).

2.6.2.2 Band densitometry

For western blot quantification, densitometry is the first choice and is considered the gold standard. The band density is proportional to the protein concentration and is therefore semi-quantitative. Band densitometry was calculated using ImageJ (Version 1.53a, NIH, USA) where scanned films were converted to 32 bit. Bands were individually selected with the rectangular region of interest tool and 'Gels' function, before quantification of the peak area. This generated arbitrary area values for each of the genes of interest. Differences in band density between WDLS and paired normal were tested for statistical significance using the Wilcoxon signed-rank test.

2.7 SDHA colorimetric assay

200 mg of paired tissue samples were diced in 300 µL of 25 mmol phosphate buffer (pH 7.4) using a scalpel. Samples were snap frozen and thawed in a 37˚C water bath to promote protein lysis. Samples were centrifuged at 10,000 x g, the lower liquid phase extracted, and 4 x 50 µL lysates generated for each sample.

A reaction buffer was prepared as per table 2.8. 50 µM Decylubiquinone, 5 mM 2,6-Dichloroindophenol and 40 mM Malonate were prepared separately. Decylubiquinone acts as a succinate dehydrogenase (SDH) electron acceptor which enables electron transfer from SDH to Dichloroindophenol (DC-PIP). DC-PIP acts as the final electron acceptor, which decolourises on reduction. Malonate is a competitive inhibitor of SDH by binding to the SDH subunit where succinate normally would. It was used as a negative control.

Table 2.8: *SDHA* colorimetric 10 x assay buffer

Reagent	Final Concentration	Volume
Potassium phosphate	200 mM	25 µL
Triton	10%	2 µL
Succinate	1 M	4 µL
Rotenone	1 mM	10 µL
Antimycin A	5 mM	0.8 µL
Sodium Azide	25 M	8 µL

50 µL of homogenate was added to 120 µL of 10x Reaction buffer and 130 µL of water in a cuvette. 1 µL of decylubiquinone was added and incubated for 6 minutes at room temperature. The cuvettes were added to the FLUOstar Omega Microplate Reader (BMG Labtech, UK) and kinetic measurement started at 600 nM. At one minute 6 µL of DC-PIP was added to initiate the reaction and mixed well.

Measurement continued for another three minutes and 1 µL of malonate was added to one of the duplicate samples. Measurement continued until 40 minutes and then terminated. The absorbance values at 600 nm were plotted. Linear regression was performed using GraphPad (Prism, USA) to quantify the degree of slope as this defined SDHA activity. The slope of each sample with malonate, was subtracted from the paired sample without malonate.

2.8 SDHA fusion panel

An SDHA fusion panel was performed following the QIAseq™ Targeted RNAscan Panel (Qiagen, Germany) protocol (146). An abridged summary of the steps is given here. Reagents were supplied by Qiagen unless stated. The basic premise is highly efficient RNA conversion to cDNA, gene specific single primer enrichment (SPE), and molecular barcoding.

2.8.1 Single primer enrichment

Single primer enrichment (SPE) requires a priori genomic information for primer design, which in this case are the co-ordinates of *SDHA*. SPE in contrast to traditional PCR, only targets one of the template strands at a locus-of-interest (*SDHA*) (147). Extension occurs along the template until it reaches a universal primer at one end. This imposes no predefined target length and allows the discovery of structural variants such as gene fusions.

2.8.2 First strand cDNA synthesis

5 µL of total RNA (100 ng) per sample was added to each well of a 96-well PCR plate (Qiagen, Germany) on ice. 1 µL of RP primer was added to reach a total volume of 6 µL. The plate was transferred to a thermocycler and incubated at 65˚C for 5 minutes. The plate was removed and placed on ice for two minutes. 2 µL of BC3 buffer, 1 µL of Rnase inhibitor and 1 µL of EZ reverse transcriptase was added to each well. The plate was placed in the thermocycler and incubated at 25˚C for 10 minutes, 42˚C for 30 minutes and 70˚C for 15 minutes.

2.8.3 Second strand cDNA synthesis

A second strand synthesis mix was prepared consisting of 5µL of nuclease-free water, 2µL of XC buffer, 1 µL of RH RNase, 1 µL of dNTP and 1 µL of BX enzyme. This 10 µL mastermix was added to each well of the previous plate on ice. The plate

was placed in a thermocycler and incubated at 37˚C for 7 minutes, 65˚C for 10 minutes and 80˚C for 10 minutes.

2.8.4 End repair/ dA tailing

A master mix consisting of 5 µL ERA Buffer and 15 µL Nuclease-free water per sample was made in a LoBind tube on ice and added to each reaction. 10 µL of ERA enzyme was then added and mixed gently. The plate was then added to a thermocycler pre-chilled at 4˚C and incubated at 4˚C for one minute, 20˚C for 30 minutes and 65˚C for 30 minutes.

2.8.5 Adapter ligation

20 µL of 5 x Ligation Buffer, 10 µL of DNA Ligase and 15 µL of nuclease-free water was added to each well of the plate, along with a 5 µL IL-N7 adapter, to a total volume of 100 µL. The plate was incubated at 20˚C for 15 minutes using a thermocycler with the lid open.

2.8.6 Sample clean-up 1

The 100 µL reactions were transferred to a 300 µL 96-well PCR plate for sample clean up. 90 µL of QIAseq beads were added to the reaction and mixed ten times. The plate was placed on a magnetic rack and once the solution was clear the supernatant was removed. Three separate washes were performed with 260 µL of 80% ethanol. DNA was eluted from the beads with 52 µL of nuclease-free water. 50

µL was transferred to a new tube and 65 µL of QIAseq beads added. The plate was placed back on the magnet, and a beads wash performed as before except with a final elution volume of 10.4 µL. This completed bead-based clean up.

2.8.7 SPE target enrichment

A reaction mix consisting of 4 µL QIAseq RNA Buffer, 4 µL QIAseq Targeted RNAscan panel, 0.8 µL QIAseq IL-F and 0.8 µL HotStarTaq DNA polymerase was added to the newly eluted 10.4 µL sample to create a total volume of 20 µL. The plate was sealed and added to a thermocycler with the following conditions (Table 2.9).

Table 2.9: PCR conditions for SPE target enrichment

Step	Cycles	Incubation temperature (˚C)	Incubation time
1	1	95	15 min
2	8	95	15s
		68	10 min
3	1	72	5 min
	1	4	Hold

Once completed the reactions were placed on ice.

2.8.8 Sample clean-up 2

30 µL of nuclease-free water was added to each 20 µL reaction to bring the volume to 50 µL and transferred to a 96-well PCR plate for purification. A beads wash was performed as previously stated, except using a final elution volume of 13.4 µL.

2.8.9 Universal PCR amplification

A universal PCR reaction mix was prepared consisting of 4 µL of QIAseq RNA buffer, 0.8 µL of QIAseq IL-U primer, 0.8 µL of QIAseq Index Primer (paired with a specific adapter in the adapter ligation step) and 1 µL of HotStarTaq DNA polymerase. This was added to the previous reaction and incubated as per Table 2.10.

Table 2.10: PCR conditions for universal PCR amplification

Step	Cycles	Incubation temperature (˚C)	Incubation time
1	1	95	15 min
2	25	95	15s
		60	2 min
3	1	72	5 min
	1	4	Hold

2.8.10 Sample clean-up 3

30 µL of nuclease-free water was added to each 20 µL reaction to bring the volume to 50 µL. A beads wash was performed again, with a final elution volume of 21 µL of

sterile water which constituted the library. Libraries were then prepared for next generation sequencing as per 2.14.

2.9 SdHA fusion in RetroPeritoneal liposarcoma Study (SHARPS)

2.9.1 Ethics

To assess the role of SDHA in vivo, a basic science study was designed utilising metabolic tracing experiments in patients undergoing surgery to remove a WDLS. The metabolic tracer used in the study was $^{13}C_6$ glucose (Cambridge Isotopes, USA). Using fluid enriched with $^{13}C_6$ glucose poses no risk to patients. $^{13}C_6$ glucose is used to increase the proportion of ^{13}C to the more abundant ^{12}C, and allow metabolic tracing by gas chromatography mass spectrometry (GC-MS).

An ethics application was made to run the study at the QEH. This was made to the Health Research Authority via the Integrated Research Application System (IRAS Project ID - 238251) on the 9th October 2018. The project was entitled 'The consequences of SDHA fusion gene on tumourigenesis and tumour metabolics in well differentiated retroperitoneal liposarcoma.' This was shortened to SHARPS (SdHA fusion in RetroPeritoneal liposarcoma Study) as a working title and final ethical approval was granted 18th March 2019. A separate application was made to the University Hospitals Birmingham for internal Research and Development approval (RRK 6457). Patient consent form (V 1.3), patient information sheet (V 1.6) and ethical approval are attached in Appendix I.

2.9.2 SHARPS protocol

Potential patients for enrolment in the study were identified at the weekly sarcoma multidisciplinary team meeting. Eligible patients had a diagnosis of WDLS made radiologically or by tissue biopsy. It was accepted that the final histology may change as radiological analysis and tissue biopsies are not 100% sensitive. Diabetic patients were excluded from the study on the basis that they may be given intravenous insulin the night before surgery. Patients were invited to clinic and fully informed of the risks and benefits. A patient information leaflet was issued, and a study specific consent form signed. Patients could withdraw from the study at any point.

Study consent was reconfirmed the night before surgery. Following this, two venous blood samples were taken into Vacutainer® EDTA collection tubes (BD, USA). Blood samples were immediately spun in a 4°C centrifuge for 15 minutes at 3000 x g. After spinning, the supernatant (plasma) was extracted and added in 200 µL aliquots to cryovials. These were then flash frozen in liquid nitrogen before being stored at -80°C.

On the morning of surgery, a team brief was performed. This primarily dealt with the surgical and anaesthetic concerns for the case, but allowed time for the chief investigator to detail what the study would entail, and what was required of the relevant staff members.

16 g of $^{13}C_6$ glucose, supplied as a sterile powder, was made up by the chief investigator, in 50 mL 0.9% Sodium Chloride (Baxter, USA) to give a concentration of 0.32 g/ mL. This was then labelled and witnessed by the anaesthetist. Once the operation began, a bolus of 25 mL (8g) was given as a loading dose. This was followed by a further 8g over 2 hours, at a rate of 12.5 mL/ hour.

Two further venous blood samples were collected intra-operatively. This was either done as close to 1 and 2 hours from the loading dose as possible, or if the operation was likely to last less than two hours, completed at 30 and 90 minutes. Blood samples were processed and stored as before.

Once the tumour was removed from the abdomen it was taken to the pathology department immediately, where a histopathologist took samples of tumour and healthy paired normal. The samples were placed in cryovials and snap frozen in liquid nitrogen, before being taken by the HBRC in a dry ice container to store at -80°C.

2.10 Gas Chromatography – Mass Spectrometry (GC-MS)

2.10.1 Polar metabolite extraction from tissue

Liquid nitrogen was added to a polystyrene box. A pestle and mortar were added, so the bottom of the mortar was bathed in liquid nitrogen. A microspatula for transferring powdered samples was precooled in the liquid nitrogen. Samples were stored on dry ice during the experiment when not being processed.

Electronic scales were tared to a cold sample tube. Each sample was crushed by a pestle into a fine powder and 60 mg transferred to a sample tube. Samples were stored in the tube rack bathed in liquid nitrogen once processed.

A cold metal bead and 1 mL chilled high performance liquid chromatography (HPLC) grade methanol was added to each sample. Samples were transferred to wet ice. 200 µL of d_6-pentanedioic acid (2.5 µg/ mL) was added as an internal standard. Tubes were lysed in a tissue lyser (Qiagen, Germany) at 20/ s frequency for 3 minutes. The sample was then spun at maximum speed for 5 minutes to pellet the debris. The supernatant was removed and transferred to a glass tube. 1 mL of water and 1 mL of chloroform was added using a glass pipette and vortexed vigorously for 20 seconds. The samples were then spun on a large centrifuge for 5 minutes at maximum speed.

The top phase was transferred to a new glass tube and 1 mL of chloroform was added, vortexed and spun down as before. The top phase containing polar metabolites was then withdrawn and added to a labelled 2 mL tube being careful not to remove the non-polar or protein phase. The top fractions were placed in a speed dryer for 2 hours, then left in a fume cupboard overnight to dry out completely. The samples were then stored at -20˚C.

2.10.2 Polar metabolite extraction from plasma

200 µL of blood plasma and 200 µL of d_6-pentanedioic acid (2.5 µg/ mL) was added to a 2 mL tube on ice. 1 mL of ice-cold methanol was added and the sample

vortexed for 15 seconds. This was spun at maximum speed for 10 minutes, before the top polar phase was removed and placed in a speed dryer for 2 hours.

2.10.3 Derivitisation

Polar metabolites from either tissue or plasma were then derivitised. 60 µL of 2% Methoxyamine hydrochloride (MeOX) in pyridine was added to each sample. This was incubated at 60˚C for 1 hour and spun for thirty seconds. To this 40 µL of N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA) was added and vortexed for 15 seconds. This was then incubated at 60˚C for 1 hour, spun for thirty seconds before being transferred to a glass GC-MS vial.

2.10.4 GC-MS analysis

GC-MS analysis utilised both the Agilent 6890 gas chromatograph (Agilent, USA) together with an Agilent 5975C mass spectrometer. Electron impact was 70 eV with source and quadrupole held at 230˚C and 150˚C respectively. Helium was used as the carrier gas at a flow rate of 1 mL/ min. 1 µL of the derivatised sample was then loaded with an inlet temperature of 280˚C on to the column. Oven temperature remained at 100˚C for 1 minute before increasing at a rate of 5˚C/ min to a maximum of 330˚C. Selected ion monitoring (SIM) mode was used for ion detection. The MetaboliteDetector software package corrected for natural isotope distribution and determined the mass isotopomer distribution (MID) (148).

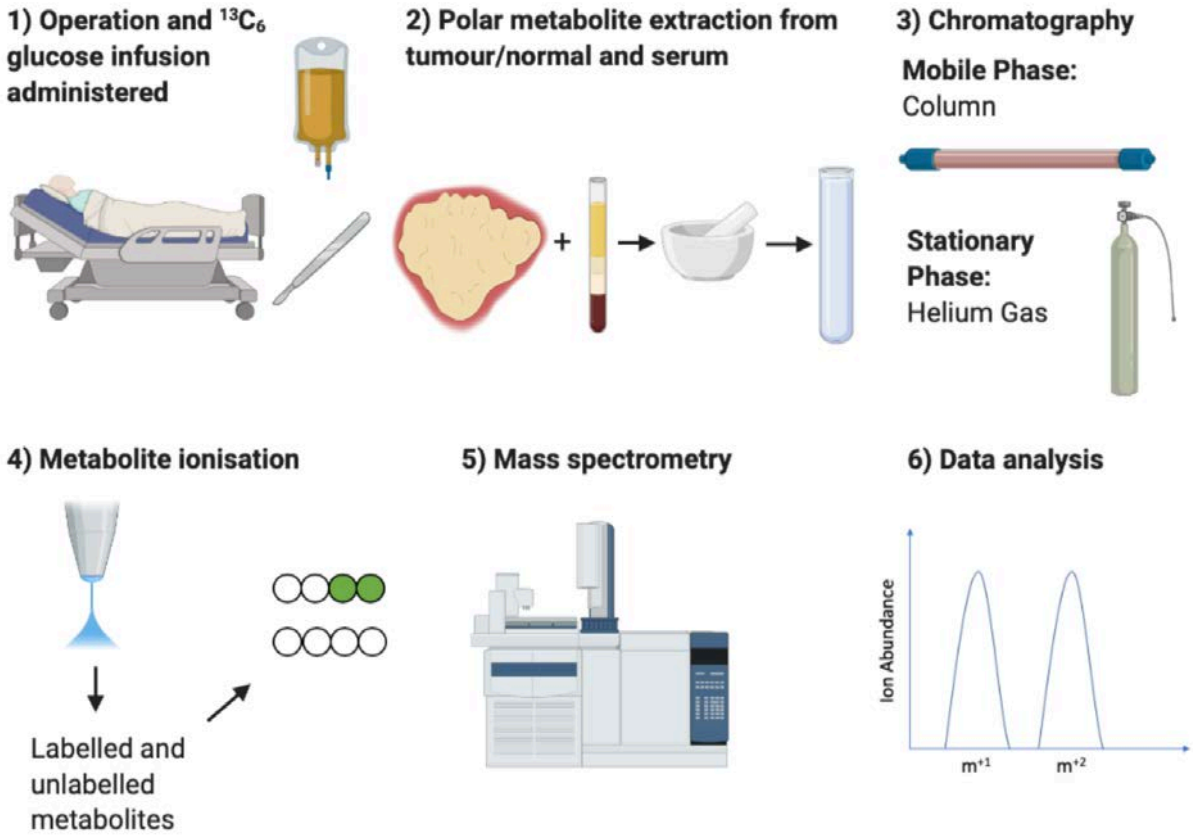

Figure 2.3: Schematic overview of SHARPS study. The patient is given a bolus of glucose tracer at the very beginning of sugery (8 grams). 4 g/hour is then infused for two hours. The tumour is taken to pathology quickly, where tumour and normal tissue are snap frozen in liquid nitrogen. Polar metabolites are extracted from both tumour and paired normal tissue, and prepared for GC-MS. The readouts from this are analysed by Metabolite Detector software, which gives ratios of isotopomers and allows us to infer which metabolic pathways are preferentially used.

2.11 RNA sequencing

To assess samples 1-9 for the presence of an *SDHA* fusion, total RNA sequencing was performed using the NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs, USA). The lowest starting sample was 51.6 ng/ µL.

2.11.1 Probe hybridisation to ribosomal RNA

A total of 619 ng from each sample was added to a PCR tube in up to 12 µL of nuclease-free water. 1 µL of NEBNext rRNA depletion solution and 2 µL of probe hybridisation buffer was added to each sample. Samples were placed in a thermocycler, incubated at 95˚C for 2 minutes before ramping down to 22˚C at 0.1˚C/ second to remain at 22˚C for 5 minutes.

A mastermix of 2 µL NEBNext RNase H, 2 µL NEBNext RNase H Reaction Buffer and 1 µL Nuclease-free Water was made up and 5 µL added to each sample, resulting in a total volume of 20 µL. This was mixed thoroughly before being incubated for 30 minutes at 37˚C.

2.11.2 Dnase I digestion

A DNase master mix was prepared on ice containing 5 µL of Dnase I Reaction Buffer, 2.5 µL of DNAse I and 22.5 µL of nuclease-free water. 30 µL of Dnase I master mix was added to 20 µL of RNA sample resulting in a total volume of 50 µL. This was incubated in a thermocycler for 30 minutes at 37˚C.

2.11.3 RNA purification, fragmentation and priming

110 µL of resuspended RNA Sample Purification Beads were added to the previous sample and incubated for 10 minutes on ice to before being placed on a magnetic rack. Once the solution was clear, the supernatant was discarded. 200 µL of freshly prepared 80% ethanol was then added to the tube whilst on the magnetic rack,

incubated at room temperature for 30 seconds and then removed. A second wash was then repeated.

The samples were removed from the magnet, and the RNA eluted from the beads in 7 µL of nuclease-free water. This was incubated for two minutes at room temperature, before being placed on a magnet until the solution was clear and 5 µL of supernatant containing RNA removed. To each 5 µL sample, 4 µL of NEBNext First Strand Synthesis Reaction Buffer and 1 µL of Random Primer was added. This was incubated for 2 minutes at 94˚C before being transferred on ice to undergo first strand cDNA synthesis.

2.11.4 cDNA synthesis

To each 10 µL sample, 8 µL of NEBNext Strand Specificity Reagent and 2 µL of NEBNext First Strand Synthesis Enzyme Mix were added. The final volume of 20 µL was mixed thoroughly before incubation in a thermocycler at 25˚C for 10 minutes, 42˚C for 15 minutes and 70˚C for 15 minutes.

A second strand cDNA synthesis mix was prepared consisting of 10 µL first strand synthesis product, 8 µL NEBNext Second Strand Synthesis Reaction Buffer, 4 µL NEBNext Second Strand Synthesis Enzyme Mix and 48 µL of nuclease-free water. This was incubated in a thermocycler for 1 hour at 16˚C.

2.11.5 Purification of double-stranded cDNA

144 µL of resuspended SPRIselect beads were added to the second strand synthesis mix (80 µL) and incubated at room temperature for 5 minutes. The samples placed on a magnet and after the solution was clear, the supernatant was removed and discarded. The beads then underwent two ethanol washes as previously, before air dried for 5 minutes with the lids open.

The samples were removed from the magnetic rack before the DNA was eluted in 53 µL of 0.1X TBE buffer and incubated for 2 minutes at room temperature. The samples were placed back on the magnet until the solution was clear, and 50 µL of supernatant removed and transferred to a clean nuclease-free PCR tube.

2.11.6 End prep of cDNA Library

7 µL of NEBNext Ultra II End Prep Reaction Buffer and 3 µL of NEBNext Ultra II End Prep Enzyme Mix were added to the 50 µL cDNA libraries. Samples were mixed thoroughly before incubation in a thermocycler for 30 minutes at 20˚C and 30 minutes at 30˚C.

2.11.7 Adapter ligation

The NEBNext Adaptor was diluted 5-fold in the Adaptor Dilution Buffer and kept on ice. A ligation reaction mix was prepared on ice consisting of 60 µL end-prepped DNA, 2.5 µL diluted adapter,1 µL NEBNext Ligation Enhancer and 30 µL NEBNext Ultra II

Ligation Master Mix. This was mixed thoroughly and incubated for 15 minutes at 20˚C on a thermocycler. 3 µL of USER Enzyme was then added to the ligation mixture, resulting in a total volume of 96.5 µL. This was mixed well and incubated at 37˚C for 15 minutes before proceeding to purification.

2.11.8 Purification of the ligation reaction

87 µL resuspended SPRIselect beads were added to each sample, mixed thoroughly and incubated at room temperature for ten minutes. The tube was placed on a magnetic rack and two ethanol washes were performed as before. The samples were removed from the magnetic rack, and DNA eluted from the beads by adding 17 µL of 0.1X TE buffer. The tube was placed on a magnet until the solution was clear. 15 µL of supernatant was transferred to a clean PCR tube prior to PCR enrichment.

2.11.9 PCR enrichment of adaptor ligated DNA

A PCR reaction was set up for each individual sample using 15 µL of adaptor ligated DNA, 25 µL NEBNext Ultra II Master Mix, 5 µL Universal PCR primer and 5 µL index primer. The reactions underwent PCR amplification using the PCR conditions in Table 2.11.

Table 2.11: PCR conditions for amplification of adaptor ligated DNA

Cycle Step	Cycles	Temperature (˚C)	Time
Initial Denaturation	1	98	30 seconds
Denaturation Annealing/ Extension	12	98 65	10 seconds 75 seconds
Final Extension	1	65	5 minutes
Hold		4	

2.11.10 Purification of PCR Reaction using SPRIselect beads

45 µL of resuspended SPRIselect Beads were added to the PCR reaction before the samples were placed on a magnetic rack and two ethanol washes performed.

The samples were removed from the magnetic rack, and the DNA eluted by adding 23 µL of 0.1X TE to the beads. The samples were mixed well, spun down in a microcentrifuge and incubated for two minutes at room temperature. They were then placed in the magnetic rack until the solution was clear. 20 µL of supernatant was then added to a clean PCR tube. This constituted the final library which was then prepared for next generation sequencing as per 2.14.

2.12 Genome Wide CRISPR-Cas9 Knockout screen (GeCKO)

2.12.1 Lentiviral production

Lentiviral production and library amplification was kindly performed for a previous experiment and gifted by Ms. Louise Tee as per manufacturers protocol (149). A brief summary of this is given here.

Bacterial plasmids containing lentiviral envelopes and green fluorescent protein (GFP) were grown on Agar plates. GeCKO pooled sgRNA libraries were purchased from Addgene, USA. Electrocompetent cells were transformed through electroporation with GeCKO pooled library A + B. Transformed bacteria were grown on bioassay plates before plasmid DNA was extracted. Pooled GeCKO libraries underwent PCR amplification using primers that include Illumina adapters. They were then sequenced on a Mi-Seq (Illumina, USA) to check coverage and library quality. Lentivirus was produced by transfecting low passage HEK293FT cells with lentiviral plasmids using lipofectamine.

2.12.2 Lentiviral titer

To calculate the multiplicity of infection required for optimal lentiviral transfection two preliminary experiments were performed. Firstly, a cell viability assay was performed to determine the optimum number of cells per well for continuous growth over 72 hours. Secondly, a puromycin screen was performed to generate a kill curve and

derive the lowest concentration at which puromycin was toxic to the cells. Puromycin acts as a selection antibiotic in the GeCKO screen.

2.12.3 Cell Viability Assay

250 µL of PBS was added around the edge of a 96-well clear bottom opaque walled culture plate (Corning, USA) to act as a humidity barrier. 1 mL of 4 x 10^5 cells was made up with fresh media and 200 µL added in triplicate to the 96 well-plate. Serial dilutions were performed down the plate to the last row where a negative control of media only was added. This is illustrated in Figure 2.4.

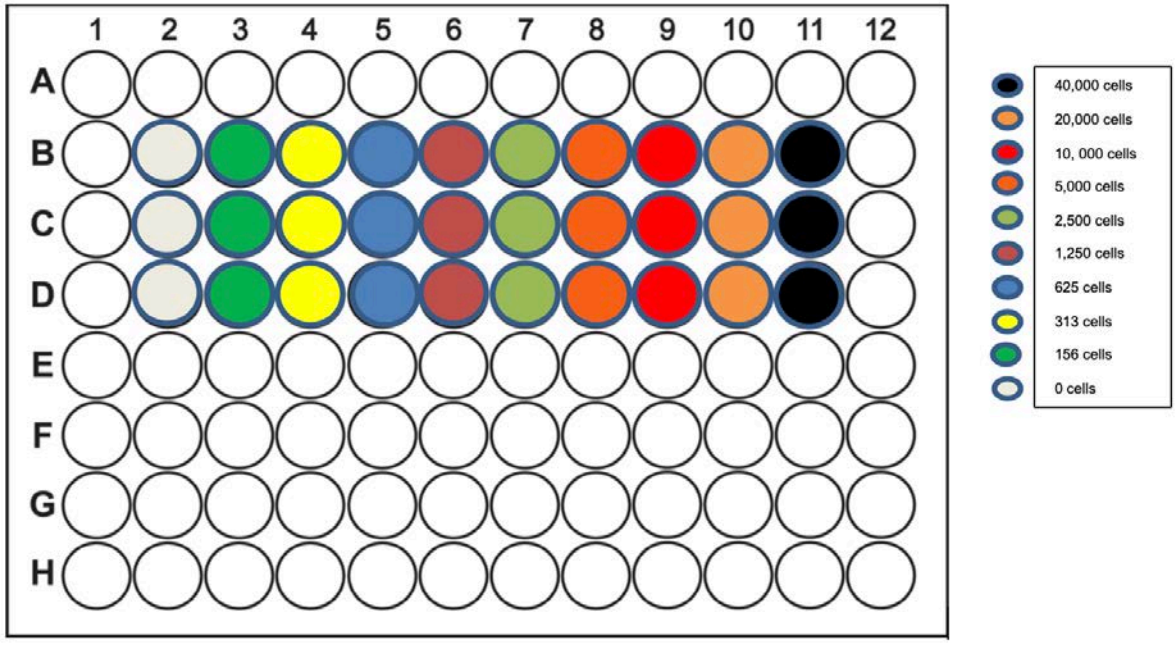

Figure 2.4: 96 well plate set-up from 0 to 40,000 cells in triplicate.

The plate was left in the incubator overnight to allow the cells time to adhere. On the morning of day 1 a solution of CellTiter-Glo®2.0 (Promega, USA) was prepared. 4 mL of substrate and 4 mL of enzyme was added to 1992 mL media as a working solution. 100 µL of this was added to each well of the plate. On day 2, the plate was

read using a EnSpire Multimode Plate Reader (Perkin Elmer, USA) at 24 hours. Further measurements were taken at 32, 48, 50, 72 and 80 hours. The average reading per time point was calculated for each of the cell concentrations. 5000 cells showed the closest to linear growth pattern.

2.12.4 Antibiotic Kill Curve

A puromycin kill curve was performed as this would be used as the selection antibiotic. 5000 cells were seeded in 100 µL to wells of a 96-well culture plate. On day 1 100 µL of puromycin was added to the cells, in duplicate, at the following total concentrations – 0, 1, 2, 3, 4, 5, 6 µg/ mL. The cells were incubated for 24 hours. 1 mL of 2x RealTime-Glo™ was prepared by adding 2 µL of MT Cell Viability Substrate and 2 µL of NanoLuc®Enzyme to 996 µL of cell culture medium. 100 µL was removed from each active well in the plate, and 100 µL of 2x RealTime-Glo™ added. The plate was placed on a shaker, and shaken carefully for 5 minutes. It was then covered with tin foil, left to incubate at room temperature for 10 minutes before being read using the plate reader. At a concentration of 0.6 µg/mL there was no cell activity.

2.12.5 Determination of the lentiviral titer through transduction

A clear bottom opaque walled culture plate (Corning, USA) was seeded at a density of 5×10^5 cells in 2 mL culture media per well with 8 µg/ mL of polybrene (Merck KGaA, Germany). 5×10^5 cells were chosen after based on the cell viability assay. To separate wells, 0, 2.5, 5, 10, 20, 40 µL of lentiviral supernatant was added. To ensure no lentiviral spillage and efficient centrifugation, parafilm was wrapped around

the plate, and the centrifuge buckets were lubricated prior to use. The plate was centrifuged at 1000 x g for 2 hours at 33˚C.

24 hours after the end of spinfection, cells were dissociated with 500 µL of TrypLE, before resuspension in 1 mL of media. The cell concentration of the no virus well was then measured as 377,000 cells/ mL. GFP transfected cells are show in Figure 2.5.

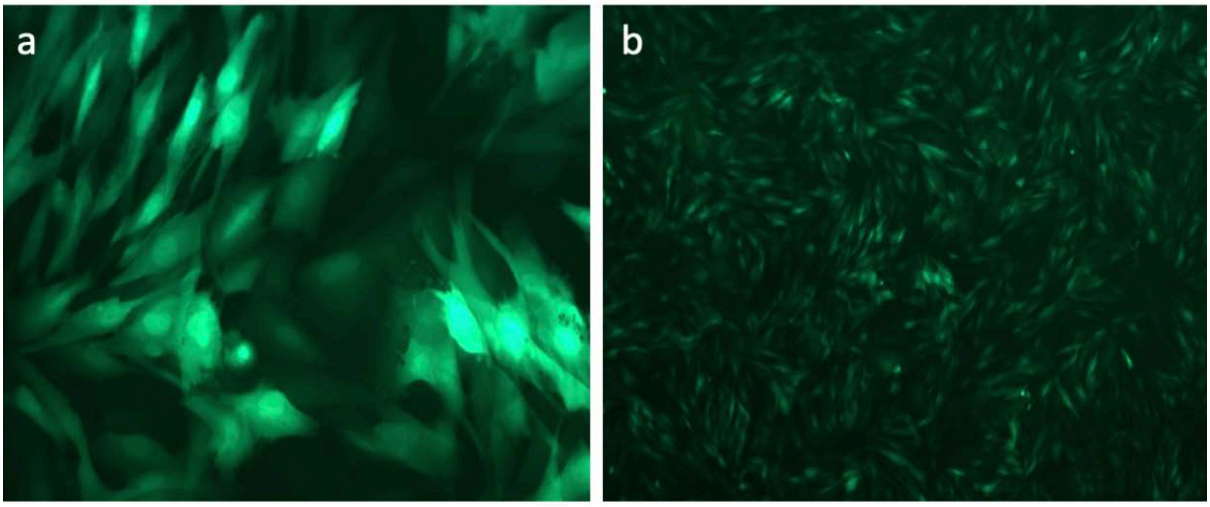

Figure 2.5: GFP transfected cells at 20x depth (a) and 10x depth (b). To be confident of successful lentiviral transfection, the sgRNA's deliver the coding region for green fluorescent protein. This can be directly visualised by fluoresence microscopy.

For each virus condition, four wells of a 96-well culture plate were seeded with 4 x 10^3 cells in 100 µL media. 100 µL of media with puromycin at was added to two wells out of each four, with 100 µL of media added to the remaining two. The plate was then incubated for 72-96 hours until the no virus condition contained no viable cells, and the no antibiotic selection condition were 80-90% confluent.

A 50:50 mix of 2x RealTime-Glo™ and media was added, and the cell viability quantified using a plate reader. For each virus condition, the multiplicity of infection (MOI) was calculated as the average luminescence of the 2 wells in the condition with antibiotic selection divided by the average luminescence of the 2 wells in the condition without antibiotic selection. Using GraphPad, the lentiviral supernatant volume (log_{10}) was plotted against MOI. The curve was interpolated to determine what logarithmic concentration of lentivirus equalled an MOI of 0.3. This then underwent logarithmic back transformation to give 8.97 µL lentiviral supernatant.

2.12.6 Lentiviral screening

The sgRNA library was transduced at an MOI < 0.3 to ensure that the majority of cells receive only one genetic perturbation. The transduction was scaled up so that the sgRNA library had a coverage of > 250 cells expressing each sgRNA. For this library size of 122,000 unique sgRNAs, 1.016×10^8 cells were transduced at a MOI of 0.3.

Cells were seeded into 6-well plates (Corning, USA) in 2 mL of media per well. 24 hours later, 8.97 µL lentiviral supernatant with polybrene at 8 µg/mL was added and spun as per the previous conditions. On day 3 of the screen, the virus was removed from each of the plates and washed with 2 mL PBS. 500 µL of TrypLE and 1 mL media was added to each well, incubated for ten minutes, combined into a 50 mL Falcon tubes (Corning, USA) and spun for 5 minutes to generate a cell pellet. The supernatant was removed, and cells were resuspended in media. 18 x T150 culture flasks (Corning, USA) were seeded with 6 million cells to give at least 1.016×10^8

cells, and incubated. On day 5 of the screen puromycin was added to each flask at a concentration of 0.6 µg/ mL. The cells underwent this selection for a further 7 days.

Once selection was complete 4.36 x 10^7 cells were present. Ideally, we aimed for a total of 6 x 10^7 cells. The pellet was split, with 2.18 x 10^7 cells frozen down as day 0 genomic DNA. The remaining 2.18 x 10^7 cells were then seeded in media into T150 flasks and incubated. Cells were grown for 14 days in normal conditions and passaged as required. 6.25 x 10^7 cells were alive at the end and 3 x 10^7 cells were frozen down in cell pellets and stored at -20˚C in preparation for harvesting the genomic DNA.

2.12.7 Harvesting genomic DNA

Genomic DNA was harvested using a Zymo Midi Prep Plus Kit (Zymo, USA) according to the manufacturer's protocol resulting in 200 µL of eluted DNA (150). The eluted DNA was then quantified by the Qubit fluorometer (Day 0: 16,320 ng/ Day 14: 23,986 ng).

2.12.8 Preparation of the gDNA for next generation sequencing

For next generation sequencing (NGS) analysis PCR was performed to amplify the sgRNA target region with Illumina adapter sequences. For Day 0, four reactions were set up using 1-4 of the NGS-Lib-Fwd primers combined with 1 Lib-KO-Rev barcode primer. For day 14, six reactions were set up using 5-10 of the NGS-Lib-Fwd primers combined with a different Lib-KO-Rev barcode primer. Reactions

consisted of 2.5 µL of forwards and reverse primers, 25 µL NEB Q5 High Fidelity Mastermix, 4.4 µL of nuclease-free water and 15.6 µL of genomic DNA. A positive control of genomic DNA and negative control of nuclease-free water was also used. Reactions were transferred to a thermocycler and incubated using the conditions below (Table 2.12).

Table 2.12: PCR conditions for sgRNA amplification

Step	Cycles	Incubation temperature (°C)	Incubation time
1	1	98	3 min
2	24	98	10 sec
		53	10 sec
		72	25 sec
3	14	72	2 min

Each reaction product was run on a 2% agarose gel to confirm successful PCR. Day 0 and Day 14 PCR products were pooled separately. The amplified screening NGS library was then purified using a Zymo-Spin V with Reservoir as per protocol (149). Purified PCR reactions were then pooled, and the final eluted DNA was quantified, and quality controlled as previously stated (Day 0 12.6 ng/ µL/Day 14 14.5 ng/ µL). The tape station profiles are shown in Figure 2.6.

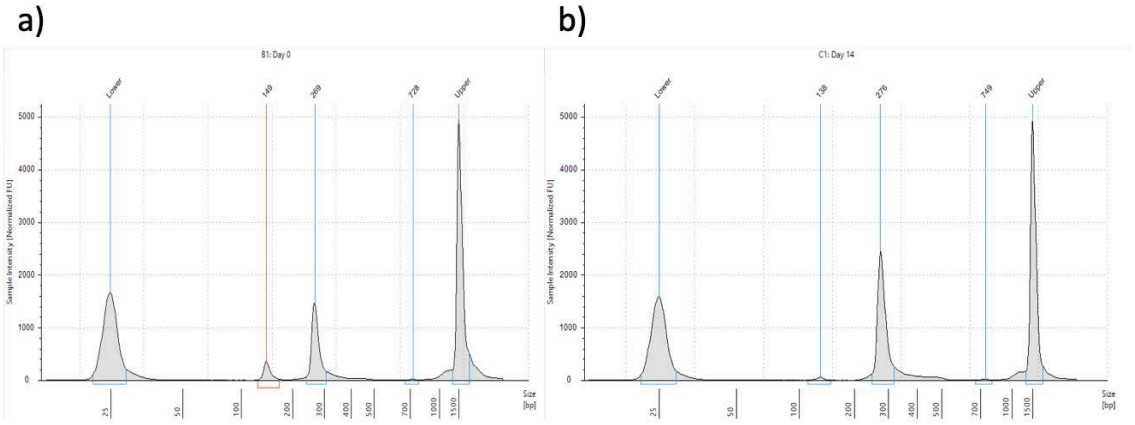

Figure 2.6: GeCKO library Day 0 (a) and Day 14 (b) profiles. Peaks of 149 base pairs (red line) in (a) and 138 in (b) indicate the libraries for sequencing.

2.12.9 DNA gel extraction

40 µL of the purified PCR product was run on a 2% agarose gel using a wide comb to allow loading of the PCR product and 15µL of a loading buffer. The gel was ran at 100V for 60 minutes. The product was present on the gel at 260bp and cut out using a scalpel under UV light. The gel fragments were carefully trimmed of excess agarose, weighed and then DNA extraction performed using the Monarch® DNA Gel Extraction Kit (NEB, USA) as per manufacturer's instructions (151). DNA was quantified to give final concentrations of Day 0: 8.34 ng/ µL and Day 14: 11.1 ng/ µL. The two libraries were then prepared for next generation sequencing as per 2.14.

2.13 Single cell RNA sequencing

A WD component of a retroperitoneal liposarcoma containing both WD and DD areas was stored in a culture medium overnight before being transferred on ice to undergo tumour dissociation. Single cell RNA sequencing was carried out using reagents and equipment from Chromium Single Cell 3' Reagent Kits v3 (10x Genomics, USA).

2.13.1 Single cell extraction

The tumour was classed as 'tough' for the protocol, based on the authors previous experience at dissecting it with a scalpel. 0.3 g of tumour was weighed out and cut into 2-4 mm pieces. This was transferred to a gentleMACS^{TM} C Tube which contained a pre-prepared enzyme mix (2.2 mL RPMI, 100 µL Enzyme H, 50 µL Enzyme R and 12.5 µL Enzyme A). The C-Tube was closed, attached upside down onto the sleeve of the gentleMACS Dissociator and the 'h_tumor_01' gentleMACS programme ran. Once finished, the C-Tube was detached and incubated for 30 minutes at 37°C under continuous rotation using the MACSmix Tube Rotator. The C-Tube was then reattached to the gentleMACS Dissociator and the h_tumor_01 programme ran once more.

The cell suspension was added to a 70 µM MACS SmartStrainer placed on a 50 mL centrifuge tube. The cell MACS SmartStrainer was then washed with 20 mL of RPMI. The cell suspension was centrifuged at 300 x g for 7 minutes. The supernatant was removed completely and resuspended in RPMI. The % of viable cells was calculated to be 25% using the cell counter.

2.13.2 Dead Cell Removal

The sample was spun at 300 x g for 10 minutes and the supernatant removed. 100 µL of Dead Cell Removal MicroBeads were added and the cells resuspended with a wide-bore pipette tip, before being incubated for 20 minutes at room temperature. The cell suspension was diluted with 500 µL of 1 x Binding buffer and added to a prepared column. Live cells passed through which were collected in a 15 mL falcon that underwent centrifugation at 300 x g for 10 minutes. The supernatant was removed, and 1 mL of PBS containing 0.04% BSA added, and the cells resuspended. After a further centrifugation and wash, the cell count was calculated at 6.33 X 10^5 with 32% alive (see Figure 2.7). The cells were then placed on ice before being undergoing GEM generation and barcoding.

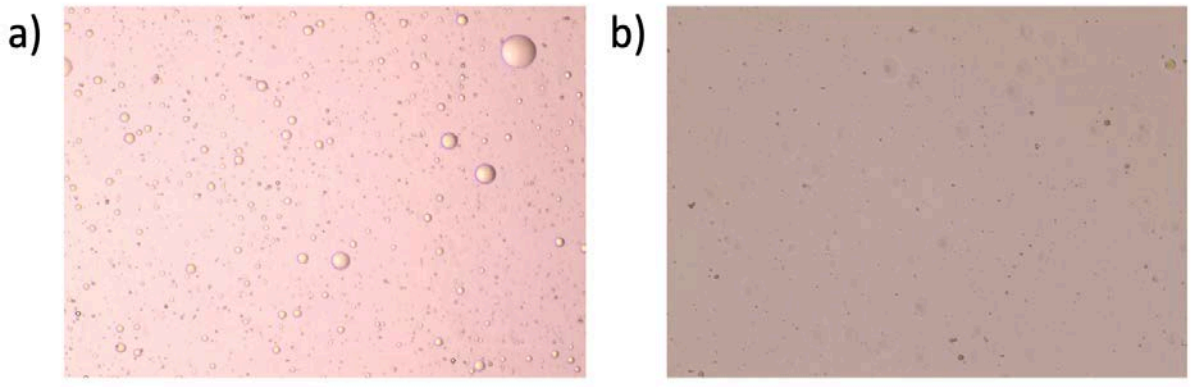

Figure 2.7: WDLS cells after tumour dissociation (a) and (b) prior to dead cell removal (10 x microscopy). 2.7 (a) contains lipid laden cells, with remnant dead cells.

2.13.3 GEM Generation and Barcoding

33.4 µL of a reverse transcription mastermix and 19.9 µL of nuclease-free water was added to 26.7 µL of labelled cell suspension stock. A Chromium Chip B in a 10x

Chip Holder was assembled. 75 µL of the master mix/ cell suspension was added into row one. 40 µL of gel beads were added to row two, and 280 µL of partitioning oil added to row three. The 10 x gasket was then attached to the controller and the chip placed in the Chromium controller. The Chromium Single Cell B program was then ran for 8.5 minutes. Once finished, 100 µL of GEMs were aspirated from the recovery wells, and transferred to a tube strip on ice. The recovered GEM's were then incubated on a thermal cycler at 53˚C for 45 minutes and 85˚C for 5 minutes.

2.13.4 cDNA Library Construction

125 µL of recovery agent was added to each sample at room temperature and left for one minute. This was then removed along with the partitioning oil from the tube. 200 µL of vortexed Dynabeads were added to each sample, pipetted ten times and incubated for ten minutes. The sample was added to a 10x magnetic separator in high position until the solution cleared. The supernatant was removed and 300 µL of 80% ethanol added to the pellet before being removed after thirty seconds. A further 200 µL of ethanol was added to the pellet, and removed after thirty seconds. The sample was then placed on the 10x magnetic separator for one minute before being removed. 35.5 µL of Elution Solution I was added, and the sample incubated for two minutes. It was placed on the magnetic separator at low until the solution cleared, and 35 µL of sample added to a new tube strip.

65 µL of cDNA amplification reaction mix was added to the 35 µL samples. This was pipetted 15 times, centrifuged and then incubated as per the conditions in table 2.13.

Table 2.13: PCR conditions for cDNA amplification

Step	Cycles	Incubation temperature (°C)	Incubation time
1	1	98	3 min
2	11	98	15 sec
		63	20 sec
		72	1 min
3	1	72	1 min
4	1	4	Hold

60 µL of SPRIselect reagent was added and the samples placed on the magnetic rack until the solution cleared. 80 µL of supernatant was added to a new strip tube whilst the pellet was taken through for pellet clean-up.

Three ethanol washes were performed with 200 µL of 80% ethanol before the samples removed from the magnet. 40.5 µL of Buffer EB was added and the mixture pipetted 15 times. The sample was incubated at room temperature for 2 minutes before being added to the magnet until the solution cleared. 40 µL of sample was then added to a new tube strip. 1 µL of the pellet clean-up was ran on the Agilent TapeStation High Sensitivity D1000 Chip (Figure 2.8)

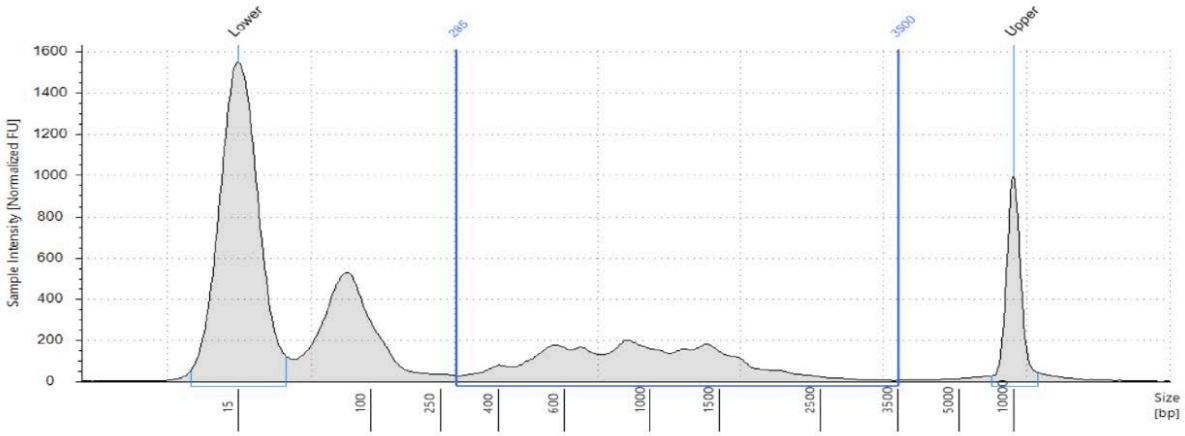

Figure 2.8: Tape station profile of cDNA. Base pair size runs along x-axis. Sample intensity runs up y-axis. The region of 285 bp to 3500 bp is highlighted in blue, representing the cDNA library. The software calculates a concentration of 496 pg/μL. This gave a total cDNA yield of 198.4 ng.

To calculate cDNA yield the cDNA concentration (496 pg/μL) is multiplied by the elution volume (40 μL) and TapeStation dilution factor (x10), then divided by 1000. This gave a total cDNA yield of 198.4 ng. 25% of this (4.96 ng) was taken forward into the 3' Gene Expression Library Construction.

2.13.5 Fragmentation, end repair and a-tailing

10 μL (25%) of purified cDNA sample was transferred to a new tube. 25 μL of Buffer EB and 15 μL of Fragmentation Mix was added. This was pipetted 15 times and centrifuged briefly. The tube was transferred to a pre-cooled thermal cycler (4˚C) and then incubated at 32˚C for 5 minutes and 65˚C for 30 minutes. 30 μL of SPRIselect reagent was added to each sample and pipetted 15 times. It was then incubated at 5 minutes before being placed on a magnet until the solution cleared. 80 μL of supernatant was removed, and the pellet washed twice with 125 μL of 80% ethanol. The tube was replaced on a magnet at a low setting until the solution cleared. 50.5 μL of Buffer EB was then added to each sample and mixed 15 times,

before incubating at room temperature for two minutes. The tube was placed back on the magnet, until the solution cleared and 50 µL was transferred to a new tube.

2.13.6 Adaptor ligation

50 µL of adaptor ligation mix was added to the sample, mixed 15 times and briefly centrifuged. It was then incubated at 20˚C for 15 minutes. 80 µL of SPRIselect reagent was then added to the sample and placed on a magnetic rack at high, until clear, and the supernatant removed. The pellet was washed twice as before with 80% ethanol and air dried for two minutes. It was then removed from the magnet, before 30.5 µL of Buffer EB added and mixed well. After two minutes of incubation at room temperature, it was placed on the magnet until the solution cleared and 30 µL transferred back to a new tube.

2.13.7 Sample index PCR

The appropriate sample index was chosen and recorded. 60 µL of sample index PCR mix and 10 µL of a sample index was added to the 30 µL sample. This was mixed gently and centrifuged before undergoing incubation as per the conditions in Table 2.14.

Table 2.14: Sample Index PCR

Step	**Cycles**	**Incubation temperature** (°C)	**Incubation time**
1	1	98	45 sec
2	12	98	20 sec
		54	30 sec
		72	20 sec
3	1	72	1 min
4	1	4	Hold

60 µL of SPRIselect Reagent was added to the sample, mixed well and incubated at room temperature for 5 minutes. It was then placed on a magnet until the solution cleared, and 150 µL of supernatant added to a new tube. 20 µL of SPRIselect Reagent was added to the sample, mixed well and incubated as before. The tube was placed on the magnet, 165 µL supernatant removed once clear, and the beads washed twice 80% ethanol. The tube was removed from the magnet, 35.5 µL of Buffer EB added and incubated at room temperature for 2 minutes. It was then placed back on the magnet, and 35 µL transferred to a new tube. The 3' Gene Expression library was now fully constructed. The library was then prepared for next generation sequencing as per 2.14.

2.14 Library Preparation for Next-generation sequencing (NGS)

2.14.1 Library Quantification

Libraries were quantified using the D1000 ScreenTape System (Agilent, USA) which calculated the average sized library peak (Figure 2.9). This was used to determine the molarity of the library in nM using the following calculation: (Concentration of library (ng/µL) / (660x Average sized peak in base pairs)) X 1000000.

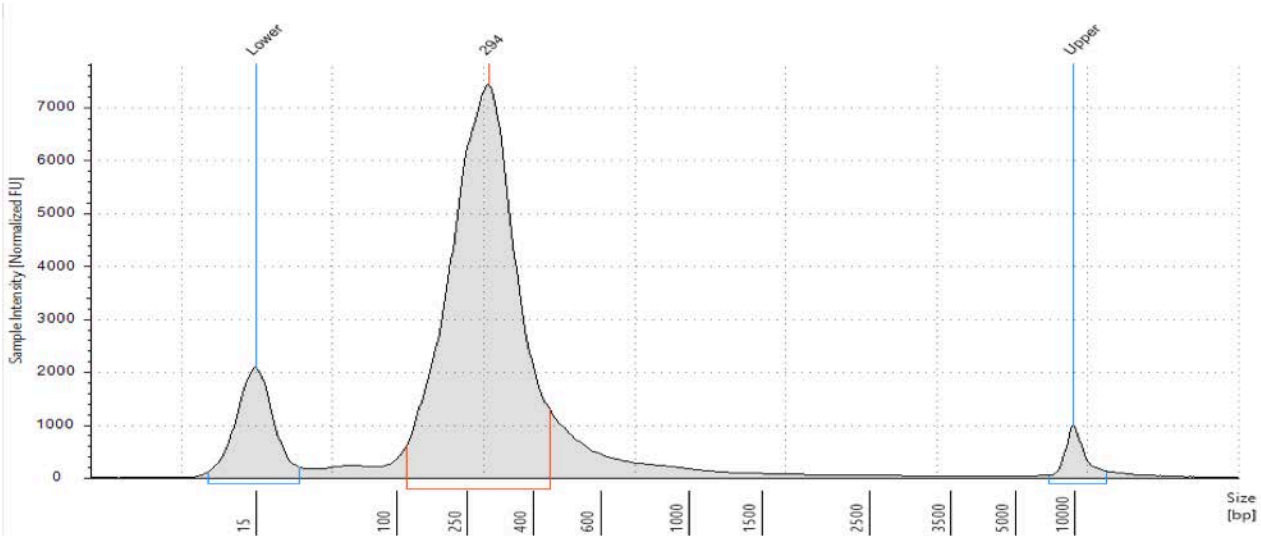

Figure 2.9: Example tape station library prior to NGS. Base pair size runs along x-axis. Sample intensity runs up y-axis. Lower and upper markers shown in blue are used by the analyser to normalise against. The peak highlighted in red shows an average base pair length of 294.

2.14.2 Library preparation and loading

Each library was diluted to 4 nM by the addition of nuclease-free water. The libraries were then pooled and mixed gently. To denature the library 5 µL of the pooled library was added to 5 µL of 0.2N NaOH. It was vortexed briefly, centrifuged at 280 x g for

one minute before being incubated at room temperature for 5 minutes. 5 µL of 200 mM Tris-HCl was added and further centrifuged at 280 x g for one minute. The denatured libraries were diluted to 20 pM by adding 985 µL prechilled HT1. The result was a 20 pM library in 1 ml.

The library was diluted to 1.6 pM in 1300 µL by adding 78 µL of the library to 1222 µL of HT1 solution. This was mixed well and pulsed on a centrifuge. 1.3 mL of the 1.6 pM library was loaded onto the NextSeq 500 reagent cartridge. 1994 µL HT1 was added to 6 µL of QIAseq A Read Primer 1 to a final concentration of 0.3 µM. 2 mL was then loaded on to the NextSeq reagent cartridge. Depending on the type of experiment a different flow cell was chosen, which is shown in Table 2.15.
The flow cell was loaded into the imaging compartment of the NextSeq 500.

Table 2.15: Flow cell type by experiment.

Chapter	Cycle Number	Output Type	Instrument
3	300	Mid	NextSeq 500
4	300	High	NextSeq 500
5	150	High	NextSeq 500
6	150	High	NextSeq 500

Standalone sequencing selecting paired end reads was then ran. The run progress was monitored on Illumina BaseSpace Software. Output files were trimmed using Trimmomatic, aligned with STAR and gene counted with DESeq2. Bioinformatics analysis was performed using Partek Flow using a default RNA sequencing pipeline.

2.15 Graphical Presentation and statistical analysis

Data was presented in graphical format using Prism Version 8 (GraphPad Software, USA). Figures with illustration were created with BioRender.com and Microsoft PowerPoint (Version 16.32, USA). Statistical analysis was performed using Prism Version 8 or SPSS version 22 (IBM Corp, USA). Statistical advice was obtained from Mr. James Hodson. Associations between exposure and outcome were calculated using an odds ratio. Odds ratios were reported with a chi-squared distribution, 95% confidence intervals and corresponding p-value. Binary logarithms were used unless stated.

Chapter 3: The role of FOXD4L3 in WDLS

3.1 Introduction

Preliminary work in the Beggs laboratory carried out whole genome sequencing on WDLS samples. It identified a possible somatic mutation in *FOXD4L3* in 8/8 samples - chr9:68,303,273A>T; c.322A>T;pLys108Ter (135).

FOXD4L3 is a single-exon protein coding gene located on chromosome 9 (9q 21.11). It encodes for a protein comprised of 417 amino acids, of which about 100 make up a highly conserved DNA-binding domain or forkhead box (152) . According to Gene Ontology database, annotations related to this gene include DNA-binding transcription factor activity (153) . Using the Human Protein Atlas, there is evidence of RNA expression in brain, reproductive and haematopoietic tissues and skin (154).

FOXD4L3 is a member of the forkhead box family. This superfamily of transcriptional regulators are comprised of 50 FOX proteins, encompassing 19 subfamilies (FOXA to FOXS), all related through their 'forkhead box' (155). FOX proteins play important roles in regulating the expression of genes involved in cell growth, proliferation and differentiation (156).

Depending on the family and cell type their role can be either as an oncogene or tumour suppressor (157). For example in humans, loss of *FOXO1* expression as a

result of chromosomal deletion (13q14) promotes androgen independent tumourigenesis whilst several key members of the FOXA, C, D, M and P subfamilies are strongly implicated in driving initiation, maintenance and progression in cancers (158).

In 2015 when preliminary experiments were carried out, this genomic aberration in *FOXD4L3* was not published in any body of work, and felt to be of interest for several reasons. As Figure 3.1 shows, it lies at the start of the forkhead box. This normally encodes a forkhead box protein, which has been shown to act in other FOX proteins as a multifaceted transcription factor (155) . The putative somatic mutation is predicted to produce a stop-gain, and therefore a premature stop codon. A premature stop codon would typically produce nonsense mediated decay of the gene product after translation, which could be significant if the forkhead box was not translated and transcriptional activity lost.

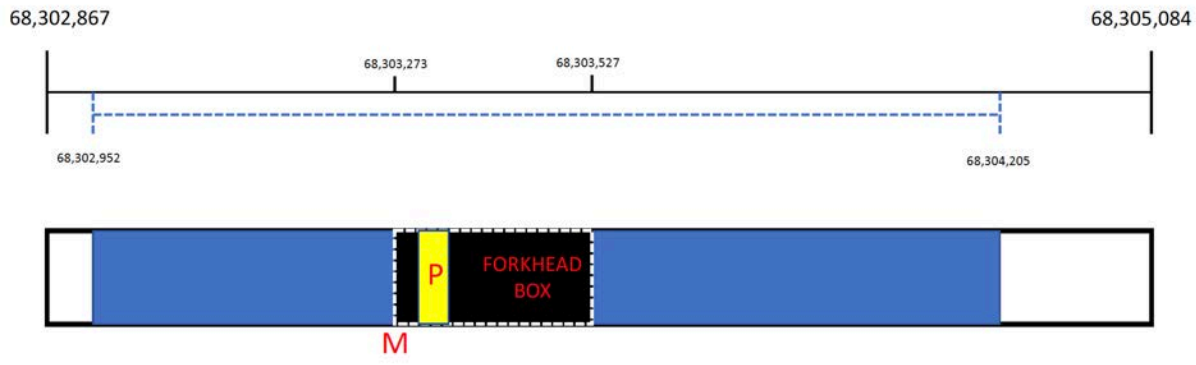

Figure 3.1: *FOXD4L3* gene on chromosome 5 (co-ordinates as per GRCh38). Blue dashed line indicates exonic boundary. M is the observed mutation. P is promoter region. Forkhead box indicated by white dashed line. Mutation and forkhead box begin at identical genomic loci.

However, in principle, a mutation can only produce nonsense-mediated decay in a gene with two exons or more, and as this is a single-exon gene, this created some doubt (159). Furthermore, although somatic mutations are clearly drivers in some cancers, in sarcomas they are rare, and the majority of somatic mutations are 'passengers' – i.e they are biologically neutral – rather than drivers which are responsible for oncogenesis (160, 161). We therefore intended to validate that this mutation was recurrent and somatic in a larger cohort of samples, using an orthogonal sequencing technique.

To interrogate *FOXD4L3* further, the complexity of the gene must be addressed. *FOXD4L3* sits at a site of an ancestral gene duplication event, which has given rise to FOXD4 paralogs, i.e homologous genes, arisen by duplication and code for a protein with similar but not identical functions (162). In mice, there is a single *FOXD4* orthologue, whereas in the human gene a duplication event has led to a total of 7 (*FOXD4*, *FOXD4L1-L6*) (157). The similarity at the nucleotide level between these seven genes ranges from 97.8% - 98.9%, with FOXD4L3 and FOXD4L6 being most closely matched (163). The FOXD4 paralogs are shown in Figure 3.2.

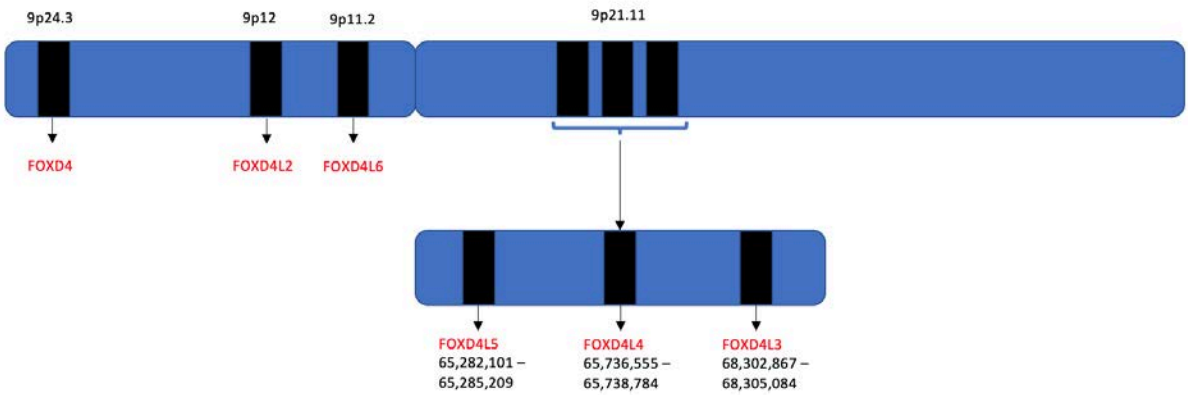

Figure 3.2: The genomic location of FOXD4 paralogs on chromosome 9. *FOXD4L3*,4 and 5 all lie within 9p21.11. *FOXD4, FOXD4L2* and *FOXD4L6* are present on the short arm 9(p). *FOXD4L1* is present on Chromosome 2.

The occurrence of near-identical paralogs is critical when choosing how to effectively sequence this part of the genome. A weakness of short-read sequencing is achieving resolution in complex areas of the genome. In paralogous regions it unable to successfully characterise the transcriptome with enough certainty (164). In the current study, this would mean being unable to distinguish *FOXD4L3* from *FOXD4L6*. For this reason, nanopore sequencing was chosen, a recent technology which generates long reads.

Nanopore technology involves a single-molecule sequencing approach where DNA is passed through a protein pore. With average read lengths of 20 kilobases, it would easily cover the 2,725 bases of *FOXD4L3* (165). In addition, this new technology is less expensive than previous NGS techniques and has reduced sample preparation time (166).

3.3 Results

A total of 19 samples underwent nanopore sequencing from 11 patients. 11/19 samples were solid tumours, of which there were 7 x 100% WDLS, 2 x 100% DDLS, 1 x 90% WDLS/10% DDLS and an ALT. A commercially available WDLS cell line (93T449) was also sequenced. 7 matched normal samples were sequenced, which were always adjacent fatty tissue. There were six paired tumour: normal samples, with one matched blood.

The average overall read depth of two runs was 91,021 with an average read length of 13 kilobases. Patient cohort demographics and histology are summarised in Table 3.1, whilst results of sequencing at the position of chr9:68,303,273 are given in Table 3.2. Patient ID is consistent throughout this chapter.

Table 3.1: Demographics and histology for patient cohort.

Patient ID	Sample Type	Gender	Age	Tumour Histology	Primary/Recurrence	MDM2 Amplified
1	Tumour	M	64	90% WDLS/10% DDLS	Primary	Yes
	Normal			Adjacent fatty tissue		
2	Tumour	F	62	100% WDLS	Primary	Yes
	Normal			Adjacent fatty tissue		
3	Tumour	M	63	100% WDLS	Recurrence	Yes
	Normal			Adjacent fatty tissue		
4	Tumour	F	60	Atypical lipomatous tumour	Primary	No
	Normal			Adjacent fatty tissue		
5	Tumour	M	68	DDLS	Recurrence	Yes
	Normal			Adjacent fatty tissue		
6	Tumour	F	72	100% WDLS	Primary	Yes
7	Tumour	F	81	100% WDLS	Primary	MD
	Normal			Adjacent fatty tissue		
8	Tumour	F	62	100% WDLS	Primary	MD
9	Tumour	F	69	100% WDLS	Primary	MD
10	Tumour	F	71	100% WDLS	Primary	MD
11	Tumour	F	73	100% DDLS	Primary	MD
12	Tumour	F	70	100% WDLS	Primary	Yes
13	Tumour	F	59	100% WDLS	Primary	No
	Normal			Adjacent fatty tissue		
	Blood			Blood		
14	Tumour	M	48	90% WDLS/10% DDLS	Primary	Yes
	Normal			Adjacent fatty tissue		
	Blood			Blood		
15	Blood	F	72	80% WDLS/20% DDLS	Primary	Yes
16	Blood	M	64	Lipoma	Primary	No

*MD – Missing data (No FISH testing carried out for MDM2 on these samples)

3.3.1 Nanopore sequencing results

Table 3.2: Base calls (%) at Chr.9:68,303,273 for samples 1-12 with corresponding sequencing depth. Samples 7-11 were sequenced on a separate run.

Patient	Specimen Type	Depth	A %	T%	C%	G%
1	Tumour	135472	53%	44%	0%	1%
1	Normal	104795	52%	46%	1%	1%
2	Tumour	51926	48%	50%	1%	1%
2	Normal	40817	57%	41%	1%	1%
3	Tumour	140004	76%	23%	1%	1%
3	Normal	83389	76%	22%	0%	1%
4	Tumour	39300	45%	51%	3%	1%
4	Normal	351837	53%	43%	2%	1%
5	Tumour	99352	78%	20%	0%	1%
5	Normal	108672	70%	28%	0%	1%
5	Blood	158893	76%	22%	1%	0%
6	Tumour	209548	77%	21%	1%	1%
7	Tumour	2369	97%	0%	1%	2%
7	Normal	8641	97%	1%	1%	2%
8	Tumour	7033	59%	34%	4%	2%
9	Tumour	2387	80%	16%	2%	2%
10	Tumour	2055	54%	40%	4%	1%
11	Tumour	3510	57%	37%	4%	2%
12	Cell line (93T449)	179410	76%	20%	2%	2%

Thymine (T) was present at this chromosomal locus in 10/11 (91%) of patients' tumour samples, indicating the previously observed chr9:68,303,273A>T was present. However, this observed mutation was also identified in 5/6 (83%) normal samples which raised the possibility of adjacent normal tissue having contaminated the tumour samples.

Efforts were made to observe the histopathological handling of specimens, which gave no evidence of contamination. Critically, the solitary blood sample showed evidence of thymine at this locus, indicating a potential germline origin rather than a somatic mutation.

At this point efforts were made to assess genomic databases to see if more recent evidence had appeared of this being described as a germline variant. The largest of these was the Genome Aggregation Database (gnomAD) (167). This identified chr9:68,303,273A>T; c.322A>T to be a single nucleotide polymorphism (SNP) present in 41.7% of the total sequenced exomes and genomes. It was given the rsID 7034645 to identify it as a specific SNP. Table 3.3 details the respective allele counts (168).

Table 3.3: rs7034645 frequency as per gnomAD database.

	Exomes	**Genomes**	**Total**
Allele Count	19841	4657	24498
Allele Number	47554	11150	58704
Allele Frequency	0.4172	0.4177	0.4173

SNPs are one of the most common types of genetic variation in the genome, and depending on their location can either have no functional consequences, or directly increase cancer susceptibility (169).

In light of this newly available data, results were reinterpreted. The nanopore samples (Table 3.3), allowing for sequencing errors found three distinct categories.

1) Heterozygous (Patient 1, 2, 4, 8,10,11)
2) 75/25 split for Adenine/Thymine (Patient 3, 5, 6, 9, 12)
3) Homozygous (Patient 7)

There was no pattern within the results based on histology or MDM2 status. There was now evidence this SNP was present at a higher rate than the quoted GnoMAD data in WDLS in both preliminary WGS and in the nanopore sequencing (91% v 41.7%). Furthermore, based on its genomic locus, it could affect gene expression. Therefore, further investigation was required, by comparing the SNP frequency in affected patients to the healthy population.

3.2.2 Comparison with a healthy, unaffected population

To achieve this aim, samples that underwent nanopore sequencing would also undergo Sanger sequencing, along with a cohort of healthy, unaffected patients. Although sequencing technologies are rapidly advancing, confirmation of SNPs by Sanger sequencing is still effective (170). It is cheaper than whole genome and nanopore sequencing, and easily scaled up for larger numbers of samples. Although challenging to sanger sequence this SNP in a complex area of the genome, enough

genomic diversity exists between FOXD4L3 and its paralogs to allow a unique primer design. A further assessment of the results when combined with the Sanger sequencing results is discussed later.

3.3.2 Sanger sequencing results

Even though the annealing temperatures and G/C content of forwards and reverse primers were very similar, the forwards primer performed less efficiently when testing on controls. For this reason, the reverse primer was chosen. As previously stated, at these co-ordinates, Hg38 calls an adenine base, which in rs7034645 is replaced by a thymine. Therefore, when reading the reverse primer Sanger chromatogram, a thymine base call identifies the base as per the reference genome, whilst either a double peak (shown by a W) or an adenine base call would represent the SNP (rs7034645). Examples of the observed peaks are show in Figure 3.3/3.4.

When trying to detect SNPs using Sanger sequencing, it is standard practice to use a secondary peak caller to prevent any missed double peaks (171). For this reason, a second highest peak threshold of 25% was set. An example of a missed peak is shown in figure 3.5. Sanger base calls for the patient cohort are shown in Table 3.4

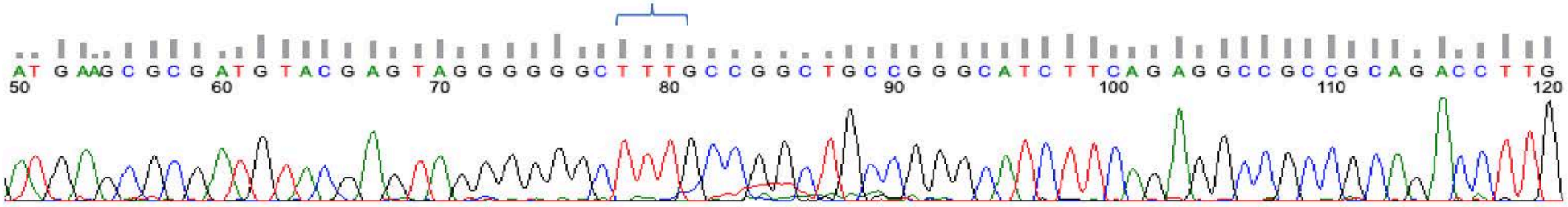

Figure 3.3: Electropherogram produced by Sanger sequencing showing a thymine (T) base call at nucleotide 79 (centre of blue bracket), using the reverse primer. This matches the reference genome.

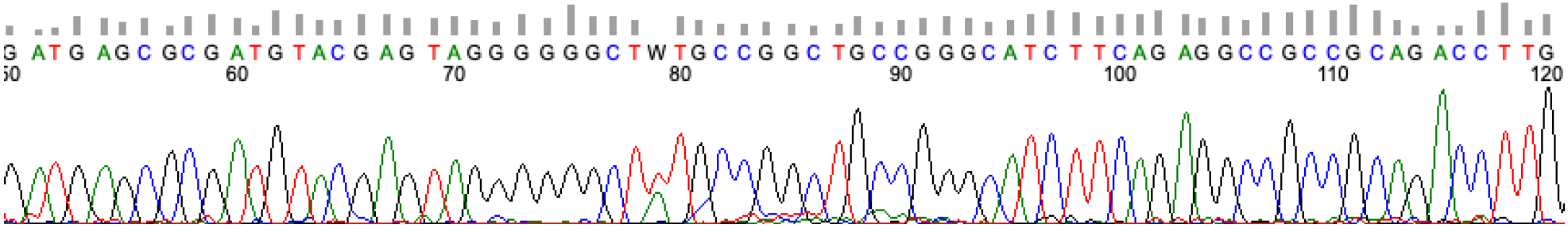

Figure 3.4: Electropherogram produced by Sanger sequencing showing a heterozygous SNP (rs7034645) at nucleotide 79, represented by a 'W'.

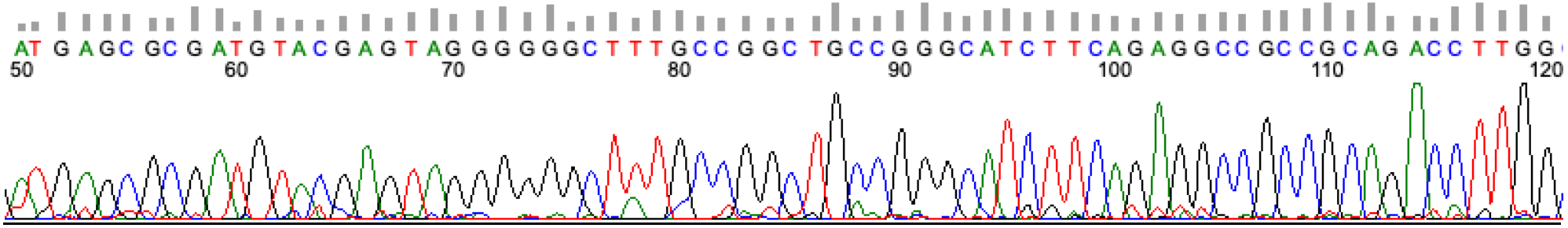

Figure 3.5: Electropherogram produced by Sanger sequencing showing a homozygous call for thymine at nucleotide 79, with smaller adenine peak missed. Secondary peak calling (set to 25%) on Sequencher recalls this as heterozygous "W"- (rs7034645).

Table 3.4: Sanger sequencing results using reverse primer for patient cohort

Patient ID	Sample Type	Sanger Base Call	Zygosity	rs7034645
1	Tumour	W (A or T)	Heterozygous	Present
1	Normal	W (A or T)	Heterozygous	Present
2	Tumour	W (A or T)	Heterozygous	Present
2	Normal	W (A or T)	Heterozygous	Present
3	Tumour	T	Homozygous	Absent
3	Normal	W (A or T)	Heterozygous	Present
4	Tumour	W (A or T)	Heterozygous	Present
4	Normal	W (A or T)	Heterozygous	Present
5	Tumour	T	Homozygous	Absent
5	Normal	W (A or T)	Heterozygous	Present
6	Tumour	T	Homozygous	Absent
12	Tumour	W (A or T)	Heterozygous	Present
13	Tumour	W (A or T)	Heterozygous	Present
13	Normal	W (A or T)	Heterozygous	Present
13	Blood	W (A or T)	Heterozygous	Present
14	Tumour	W (A or T)	Heterozygous	Present
14	Normal	T	Homozygous	Absent
14	Blood	W (A or T)	Heterozygous	Present
15	Blood	W (A or T)	Heterozygous	Present
16	Blood	W (A or T)	Heterozygous	Present

A total of 20 samples underwent Sanger sequencing from 11 patients. 9/20 were solid tumours, of which there were 5 x 100% WDLS, 2 x 90% WDLS/10% DDLS, 1 x 100% DDLS and 1 x ALT. 7 matched normal samples were sequenced, which was always adjacent fatty tissue. There were 7 paired tumour: normal samples, two with blood samples. Two additional blood samples were sequenced, one from a patient with an 80% WDLS/20% DDLS, and one from a patient with a retroperitoneal lipoma.

The SNP was detected in 16/20 (80%) samples. All samples that contained the SNP showed a double peak in the chromatogram, i.e., heterozygous. Normal and blood samples always contained the SNP, whereas four tumour samples did not. 4/7 paired samples (1,2,4 and 13) contained the SNP, along with unmatched tumour 12. In only 1/11 patient samples was the SNP not present (Patient 6).

A total of 46 samples underwent Sanger sequencing from unaffected volunteers, the results of which are shown in Table 3.5.

Table 3.5: Sanger sequencing results using reverse primer for unaffected cohort

Patient ID	Specimen Type	Sanger Base Call	Zygosity	rs7034645
1	Blood	W (A or T)	Heterozygous	Present
2	Blood	T	Homozygous	Absent
3	Blood	T	Homozygous	Absent
4	Blood	W (A or T)	Heterozygous	Present
5	Blood	W (A or T)	Heterozygous	Present
6	Blood	T	Homozygous	Absent
7	Blood	W (A or T)	Heterozygous	Present
8	Blood	W (A or T)	Heterozygous	Present
9	Blood	T	Homozygous	Absent
10	Blood	T	Homozygous	Absent
11	Blood	W (A or T)	Heterozygous	Present
12	Blood	W (A or T)	Heterozygous	Present
13	Blood	W (A or T)	Heterozygous	Present
14	Blood	W (A or T)	Heterozygous	Present
15	Blood	W (A or T)	Heterozygous	Present
16	Blood	T	Homozygous	Absent
17	Blood	W (A or T)	Heterozygous	Present
18	Blood	W (A or T)	Heterozygous	Present
19	Blood	T	Homozygous	Absent
20	Blood	W (A or T)	Heterozygous	Present
21	Blood	W (A or T)	Heterozygous	Present
22	Blood	W (A or T)	Heterozygous	Present
23	Blood	W (A or T)	Heterozygous	Present
24	Blood	W (A or T)	Heterozygous	Present
25	Blood	W (A or T)	Heterozygous	Present
26	Blood	T	Homozygous	Absent
27	Blood	T	Homozygous	Absent
28	Blood	T	Homozygous	Absent
29	Blood	T	Homozygous	Absent
30	Blood	W (A or T)	Heterozygous	Present
31	Blood	W (A or T)	Heterozygous	Present
32	Blood	W (A or T)	Heterozygous	Present
33	Blood	W (A or T)	Heterozygous	Present
34	Blood	W (A or T)	Heterozygous	Present
35	Blood	W (A or T)	Heterozygous	Present
36	Blood	T	Homozygous	Absent
37	Blood	T	Homozygous	Absent
38	Blood	T	Homozygous	Absent
39	Blood	W (A or T)	Heterozygous	Present
40	Blood	W (A or T)	Heterozygous	Present
41	Blood	T	Homozygous	Absent
42	Blood	T	Homozygous	Absent
43	Blood	W (A or T)	Heterozygous	Present
44	Blood	W (A or T)	Heterozygous	Present
45	Blood	W (A or T)	Heterozygous	Present
46	Blood	T	Homozygous	Absent

The SNP was detected in 29/46 (63%) samples.

To measure the association between the exposure (rs7034645) and the outcome (retroperitoneal liposarcoma) an odds ratio was performed comparing the genotypes between blood samples (Table 3.6). Tissue and normal samples were not compared with blood both to allow a fair comparison, and that this mutation was felt to be germline in origin.

Table 3.6: Number and Frequency of rs7034645 genotypes across cohorts

	AA	AT	p-Value	Odds Ratio	95% Confidence Interval
Cases	0	4	0.27	5.34	0.27 to 105.22
Controls	29	17			

The AT genotype of rs7034645 was associated with an increased risk compared with the AA genotype (AT vs. AA: odds ratio (OR) = 5.34) although this was not significant (p=0.27). The wide confidence intervals (0.27-105.22) could indicate that sample size was too small to detect significance.

This data was combined with the frequencies reported by the gnomAD database (168) (Table 3.7).

Table 3.7: Number and Frequency of rs7034645 genotypes when combined with gnomAD figures

	AA	AT	p-Value	Odds Ratio	95% Confidence Interval
Cases	0	4	0.089	12.57	0.67-233.54
Controls	10240	14304			

The AT genotype of rs7034645 was associated with an increased risk compared with the AA genotype (AT vs. AA: odds ratio (OR) = 12.57). This trended towards significance (p=0.089). Again, the wide confidence intervals (0.67-233.54) could indicate that sample size was too small to detect significance and more blood samples are required to achieve this.

For a more detailed interrogation of rs7034645 the nanopore and Sanger data have been combined in Table 3.8 for patients that underwent both types of sequencing.

Table 3.8: Combined Sanger and nanopore sequencing results for Patients 1-6.

Patient ID	Specimen Type	Sanger Base Call	Zygosity	rsID 7034645	A %	T%
1	Tumour	W (A or T)	Heterozygous	Present	53%	44%
1	Normal	W (A or T)	Heterozygous	Present	52%	46%
2	Tumour	W (A or T)	Heterozygous	Present	48%	50%
2	Normal	W (A or T)	Heterozygous	Present	57%	41%
3	Tumour	T	Homozygous	Absent	76%	23%
3	Normal	W (A or T)	Heterozygous	Present	76%	22%
4	Tumour	W (A or T)	Heterozygous	Present	45%	51%
4	Normal	W (A or T)	Heterozygous	Present	53%	43%
5	Tumour	T	Homozygous	Absent	78%	20%
5	Normal	W (A or T)	Heterozygous	Present	70%	28%
6	Tumour	T	Homozygous	Absent	77%	21%

When combined this table highlights the presence of thymine bases in the nanopore sequencing for samples 3T, 5T and 6T where the Sanger sequencing has not called heterozygosity. The low percentage of T calls (23, 20 and 21%) has not been picked up by the minor peak threshold of 25%. However, sample 3N has a T% of 22 which has been picked up by the sequencer.

3.4 Discussion

3.4.1 Exploration of rs7034645 genotypes in Retroperitoneal Liposarcoma

Nanopore sequencing identified 9 heterozygous (50% A/50% T) and 2 homozygous samples (100% A). This pattern was assumed to be rs7034645 in the heterozygotes, which was present in tumour and normal. In theory, a somatic mutation could exist in the tumour, and also in the neighbouring 'normal' tissue as field change patterns are speculated to occur in retroperitoneal liposarcoma (172).

To check for the presence in the germline, a blood sample underwent nanopore sequencing, which showed 76% A/ 22% T. This was one of eight samples with a similar pattern. Interestingly, whenever this pattern occurred, it was always present in its matched partner. The 75/25 ratio was also observed across tumour, normal and blood samples. One possible explanation for this pattern is copy number variation in the germline, which has been described in multiple cancers (173).

Why this pattern is present in some of the tumours and not others is difficult to explain. Gazdar et al. described 55% of lung tumours as being stromal or inflammatory, and we have shown the presence of these cell types in Chapter 7 (174). It is possible that that contamination by non-malignant cells may mask this pattern in these samples, although this is unlikely. Interestingly, the WDLS cell line showed this pattern which offers a pure representation of the cancer. Efforts should be made to compare these findings with any publicly available sequencing data for this cell line (175).

Unfortunately, we were unable to further interrogate this 75/25 pattern in the healthy cohort of patients. Even with the use of a computer programme to detect polymorphic positions, Sanger sequencing is not designed for complex copy number variations. This could be resolved by nanopore sequencing of the healthy cohort, which due to prohibitive costs and a lack of time was not performed.

3.4.2 Increased risk associated with rs7034645

The sanger sequencing identified rs7034645 in 4/4 blood samples (100%) compared to 29/46 (63%) in the unaffected cohort. This gave an odds ratio of 5.34 (95% CI 0.271-105.2), although this was not significant (p= 0.27). If the gnomAD figures for the unaffected cohort are added to our figures, the odds ratio climbs to 12.57, with a trend towards significance (p=0.089).

According to the literature SNPs in genes classed as conferring high risk have an odds ratio of greater than 4. Therefore, this particular SNP could be classed as high risk for WDLS. If this was the case, as it is such a rare disease in the general population combined with the fact rs703465 has at least a 45% prevalence (using gnomAD data) it would be a low penetrance allele.

The presence of a positive odds ratio doesn't necessarily mean the association between exposure and outcome is significant, as our data shows (176). The presence of a wide confidence interval indicates there is an insufficient sample size, and that to detect a significant difference between the two groups you would need a much larger effect size. The key limiting factor with this dataset is the number of blood samples available from the affected cohort, due in part to the infrequent prospective blood banking of WDLS patients and also late ethical approval. Liang et al. used 471 familial breast cancer patients, and compared with 471 healthy controls to find two significant familial breast cancer alleles in the gene *SPP1* indicating the numbers required (177).

The high rate of rs7034645 (63%) in the unaffected samples is of note, as the gnomAD database has a figure of 41.7% (168). There are several possible explanations for this. One is that the secondary peak caller overestimates the number of SNP's by having a lower threshold. If the data were reanalysed without using the secondary peak caller, the affected rate would drop to 12/20 (60%) with the unaffected dropping to 21/46 (46%). This correlates much more closely with the published figures. The forwards primer was also not used, so results are based on one direction of sequencing. It is possible that with forwards and reverse primers,

fewer SNPs would have been called which could drop the rate further. Lastly, although the unaffected patients were screened for a family history of cancer, it is possible that population differences between gnomAD data and ours can contribute to this difference in SNP rate.

SNPs are implicated in cancer through several mechanisms, with the location of the SNP arguably the most important. Rs703465 is located at the start of the only exon in *FOXD4L3*. It is predicted to replace lysine with a premature stop codon (chr9: 68,303,273A>T; c.322A>T;pLys108Ter). This could supress gene transcription and translation of the forkhead domain. Although we have not assessed protein expression, this could prevent any functional protein being produced, which could abrogate pathways it was part of, with phenotypic consequences. Miao et al. analysed SNPs in patients with retroperitoneal liposarcoma with a control group of 131 patients. They genotyped 8 SNPs in 6 candidate genes, including MDM2 and CDK4. 3 SNPs in MDM2 showed significant differences in single loci genotypes and allele frequencies, but these were all intronic and did not affect protein function (178).

It is important to consider why this SNP could produce effects in retroperitoneal tissue – either adipocytic or fibroblastic in nature – but not in other tissue types. The loss of the forkhead binding domain may have functional consequences for adipogenesis but further work is required. Expression quantitative trait loci examines the association of each SNP with expression of relevant mRNA transcripts (179). If this SNP did produce functional effects, reduced FOXD4L3 mRNA expression would be observed.

Finally, one could further postulate that homozygosity for the minor allele may produce a more penetrant phenotype, although this was not encountered in our dataset. Other possibilities are that this SNP has no functional consequences, or that it only has phenotypic effects when combined with other SNPs or somatic mutations.

3.5 Conclusion

Through a combination of nanopore and sanger sequencing we have identified the originally hypothesised somatic mutation (chr9:68,303,273A>T; c.322A>T) to be the single nucleotide polymorphism rs7034645. This was present in a higher percentage of patients with WDLS. At this point, the functional consequences, if any, can only be speculated. A distinction must be made between this as a genetic marker, rather than a causal factor with phenotypic effect. Even if this was a risk allele, it would not meet the screening requirements. As per Wilson and Jungner's principles of screening for disease, WDLS does not represent an important health problem due to its rarity, and secondly there is no acceptable test to detect it, as the risks of serial CT scanning far outweigh the benefits (180).

SNP's in liposarcoma are not well researched. Further work must look at protein expression of *FOXD4L3* and also copy number variation. Large databases such as the 100,000 genomes project could be analysed to check rs7034645 in other cancer subtypes. Rs7034645 deserves more attention within this field, due to the higher prevalence in liposarcoma patients, and also because of its position at the very start of the forkhead box domain.

Chapter 4: The role of SDHA in WDLS

4.1 Introduction

Previous work at the University of Birmingham performed whole genome and RNA transcriptomic sequencing on 8 WDLS. A recurrent fusion gene was found in 5/8 samples between *SDHA* and other genes, as the result of an unstable break point. Additional work was required to ascertain whether these changes were only observed at the DNA/mRNA level, and whether this was translated directly into changes in the protein expression at the tissue level.

A fusion gene is defined as two genes that are joined so that they are transcribed and translated as a single unit (181). Several mechanisms exist to produce fusion genes (182). The first is translocation where DNA is exchanged between two different chromosomes. The second type is insertion, where a fragment of DNA moves to another region, either within the same chromosome, or to a different one. The third mechanism is the deletion of a fragment of DNA, resulting in juxtaposed genes which are transcribed and translated as a single unit. The fourth type is inversion, where chromosomal segments flip in relation to the centromere. The last, and potentially most relevant to WDLS is chromothripsis, as discussed earlier, where chromosomes shatter and reassemble incorrectly.

The first fusion gene was discovered in chronic myeloid leukaemia (CML). 95% of CML contained the *BCR-ABL1* fusion gene which exerted it's oncogenic phenotype through a constitutively active ABL1 kinase, which was effectively targeted with the drug imatinib (183). Gene fusions were then discovered in solid tumours, with many shared by several cancers, e.g *FGFR3-TACCS* in lung, bladder and brain cancers (184). It is now accepted fusion genes account for around 20% of cancer morbidity (185).

Many sarcomas contain gene fusions that can be pathogenetic mechanisms and diagnostic markers (Table 4.1). Alveolar rhabdomyosarcoma (*FOX01-PAX3*), myxoid liposarcoma (*FUS-CHOP*) and Ewing's sarcoma (*EWS-FLI1*) are all characterised by specific gene fusions (186). These tend to fuse a transcriptional regulatory domain to a DNA binding domain, which activates the expression of numerous target genes that promote proliferation, survival and prevent mesenchymal differentiation (187). These would be classed as 'gain of function' fusions.

Table 4.1: Summary of recognised fusion genes in soft tissue sarcomas. Adapted from Management of soft tissue sarcoma (188).

Sarcoma subtype	**Chromosomal Translocation**	**Fusion gene**
Myxoid Liposarcoma	t(12;16)(q13;q37) t(12;22)(q13;q12)	*FUS-DDIT3* *EWSR1-DDIT3*
Clear cell sarcoma	t(12;22)(q13;12)	*EWSR1-ATF1*
Dermatofibrosarcoma protuberans	t(17;22)(q22;q13)	*COL1A1-PDGFB*
Epithelioid haemangioendothelioma	t(1:3)(p36.3:q25)	*WWTR1-CAMTA*
Ewing sarcoma	t(11;22)(q24;q12)	*EWSR1-FLI1*
Fibromyxoid sarcoma	t(7;16)(q33;p11) t(11:16)(p11;p11)	*FUS-CREB3L2* *FUS-CREB3L1*
Inflammatory myofibroblastic tumour	t(2:19)(p23;p13.1)	*TPM4-ALK*
Alveolar soft part sarcoma	t(X:17)(p11.2;q25)	*ASPSCR1-TFE3*
Synovial sarcoma	t(X;18)(p11;q11)	*SS18-SSX1/SSX2*

In myxoid liposarcoma, the FUS-DDIT3 fusion transcript is identifiable in more than 95% of cases and, although not fully understood, is widely regarded as the primary oncogenic event (Figure 4.1) (189). There has been relative success in targeting myxoid liposarcoma with the drug trabectedin. It binds to the DNA minor groove dissociating the aberrant FUS-CHOP transcription factor from promoters of its target genes and also induces lethal DNA strand breaks (190).

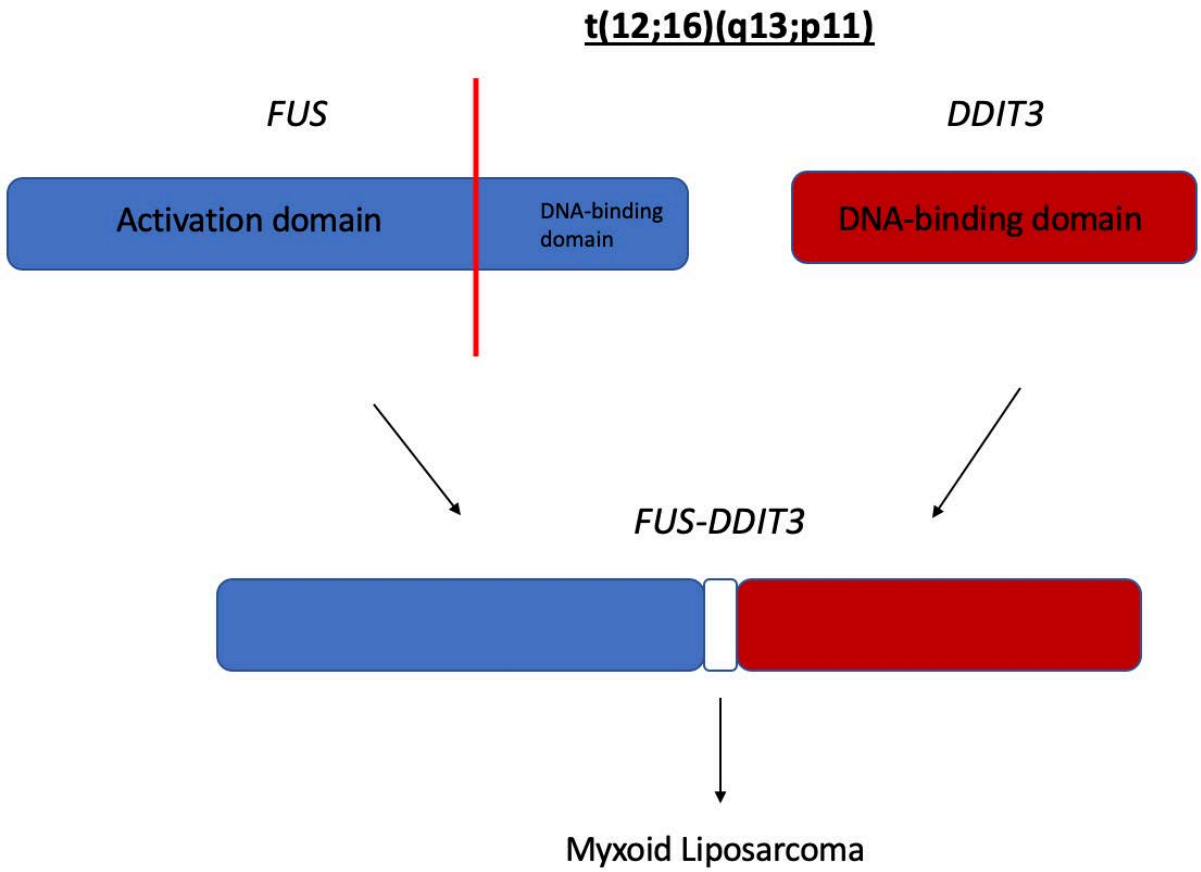

Figure 4.1: Diagrammatic representation of the *FUS-DDIT3* fusion gene. This consists of the promoter region (activation domain) of the *FUS* gene, binding to the complete coding region of *DDIT3*. Adapted from Atlas of Genetic Oncology (191).

SDHA is located on the short arm of chromosome 5 (192). It encodes the largest catalytic subunit of succinate dehydrogenase (complex II) – the smallest complex of the electron transport chain, located at the inner mitochondrial membrane (193). This complex converts succinate to fumarate through oxidation, a key step in the

tricarboxylic acid (TCA) cycle, and transfers electrons to ubiquinone in the electron transport chain.

Mutations in *SDHA* have been described in GISTs, hereditary paragangliomas and phaeochromocytomas (194-196). Burnichon et al. in paragangliomas described a germline mutation in *SDHA* (p.Arg589Trp), which caused loss of expression of *SDHA* and *SDHB*, and the loss of function of complex II (195). In GISTs, the most common gastrointestinal sarcoma, loss of *SDHA* expression predicts mutations in *SDHA* which tend to be nonsense mediated (197).

It is hypothesised that in SDH dysfunction or loss of function, accumulated succinate leaves the mitochondria and inhibits the actions of prolyl hydroxylase domain enzymes (198). This allows hypoxia-inducible factor 1-alpha (HIF-1-a) to escape degradation in the cytosol - abnormally stabilising it - leading to the induction of a hypoxic response, under normoxic conditions (pseudo-hypoxia) (193). HIF-1-a then migrates to the nucleus where it induces the expression of genes involved in angiogenesis and proliferation (199).

Loss of function caused by gene fusion has been observed in colorectal cancer. The *LACTB2-NC0A2* fusion gene resulting from intra-chromosomal rearrangements on chromosome 8 has been implicated in colorectal cancer. *NCOA2* in healthy tissues supresses Wnt/B-catenin signalling, but when fused to *LACTB2* this is lost, leading to uncontrolled Wnt signalling.

We hypothesised that *SDHA* fusions lead to a loss of function of the SDH complex. SDH loss of function, leads to an accumulation of the oncometabolite succinate

which is pro-tumourigenic. To further explore *SDHA* in liposarcoma; the transcriptome was assessed through an RNA fusion panel to look for evidence of gene fusions. Following this, evidence of SDHA expression was sought via Western blotting, to see if gene fusions were preventing the effective translation of the gene. The function of SDHA was then assessed through a colorimetric assay. Finally using four paired tumour and normal samples from the SHARPS study (Chapter 5), total RNA sequencing was performed to look for evidence of SDHA fusions.

4.2 Results

4.2.1 SDHA Fusion Panel

Of 32 samples (Table 4.2), there were 13 WDLS and 19 normal (of which 13 were adjacent fatty tissues, three testicular, two adrenal and one skin). There were 8 paired samples. These originated from FFPE scrolls, obtained from the HBRC and are an entirely separate cohort to Chapter 3.

Table 4.2: Histology for *SDHA* cohort

Patient Number	Sample ID	Tumour histology
1	1	WDLS
	2	Adjacent normal
2	3	WDLS
	4	Adjacent normal
3	5	WDLS
	6	Adjacent normal
4	7	WDLS
	8	Adjacent normal
5	9	WDLS
	10	Adjacent normal
6	11	WDLS*
	12	Testicular normal
7	13	WDLS*
	14	Testicular normal
8	15	WDLS*
	16	Skin normal
9	17	Adjacent normal
10	18	Adjacent normal
11	19	WDLS
12	20	Adjacent normal
13	21	Adjacent normal
14	22	Adjacent normal
15	23	Adjacent normal
16	24	WDLS
17	25	WDLS
18	26	WDLS
19	27	Adjacent normal
20	28	Adrenal normal
21	29	Adrenal normal
22	30	Testicular normal
23	31	WDLS*
24	32	Adjacent normal

*WD component of a tumour that contained a DD component

To avoid the reporting of false fusion discoveries, only significant ($p<0.05$) fusions were reported. Significant counts were determined by counts that align to the genome above the background level (i.e > x1). The presence of statistically significant SDHA fusions in samples 1-32 are reported in Table 4.3.

Table 4.3: Significant fusions present in *SDHA* (p-Value <0.05) by sample ID

Sample ID	Histology	*SDHA* Fusion Present (Y/N)
1	WDLS	N
2	Adjacent normal	Y
3	WDLS	Y
4	Adjacent normal	Y
5	WDLS	N
6	Adjacent normal	Y
7	WDLS	Y
8	Adjacent Normal	Y
9	WDLS	Y
10	Adjacent Normal	N
11	WDLS	Y
12	Testicular normal	Y
13	WDLS	N
14	Testicular normal	Y
15	WDLS	N
16	Skin normal	N
17	Adjacent normal	Y
18	Adjacent normal	Y
19	WDLS	N
20	Adjacent normal	Y
21	Adjacent normal	Y
22	Adjacent normal	Y
23	Adjacent normal	Y
24	WDLS	N
25	WDLS	Y
26	WDLS	N
27	Adjacent normal	Y
28	Adrenal normal	Y
29	Adrenal normal	N
30	Testicular normal	N
31	WDLS	N
32	Adjacent normal	Y

20/32 (63%) samples had statistically significant fusions. 15/19 (79%) normal samples and 5/13 (39%) WDLS had significant fusions. Normal samples without fusions were found in two testicular, one skin and one adjacent fat sample. Table 4.4 lists the significant fusion partners of *SDHA*.

Table 4.4: Significant *SDHA* fusion partners

Sample ID	Histology	Fusion Partner(s)
2	Adjacent normal	*SLC9A3,AHRR,EXOC3,CLPTM1L,LPCAT1*
3	WDLS	*PLEKHG4B*
4	Adjacent normal	*SLC9A3,EXOC3,CTD-2228K2.7*
6	Adjacent normal	*SLC9A3*
7	WDLS	*SLC9A3*
8	Adjacent normal	*CTD-2231H16.1,PLEKHG4B,LRRC14B*
9	WDLS	CTD-2231H16.1,SLC9A3,LRRC14B,TPPP,PDCD6
11	WDLS	*RP11-811I15.1,CTD-2231H16.1*
12	Adjacent normal	CTD-2231H16.1,PLEKHG4B
14	Testicular normal	*ZDHHC11*
17	Adjacent normal	*SLC9A3*
18	Adjacent normal	*SLC9A3*
20	Adjacent normal	*AHRR,CTD-2228K2.7, SLC9A3*
21	Adjacent normal	*SLC9A3,AHRR,CTD-2228K2.7,*
22	Adjacent normal	CTD-2231H16.1,PLEKHG4B
23	Adjacent normal	*SLC9A3,AHRR*
25	WDLS	*CTD-2228K2.7, SLC9A3*
27	Adjacent normal	*SLC9A3*
28	Adrenal normal	*AHRR, PDCD6*
32	Adjacent normal	*ZDHHC11,SLC9A3*

The significant fusion partners are all present on the short arm of chromosome 5. They lie within 1,398,282 bases of each other, between chr5:58,198 and chr5:1,456,480, as shown in Figure 4.2. Distance between fusion partners ranged from 62,602 to 439,581 bases.

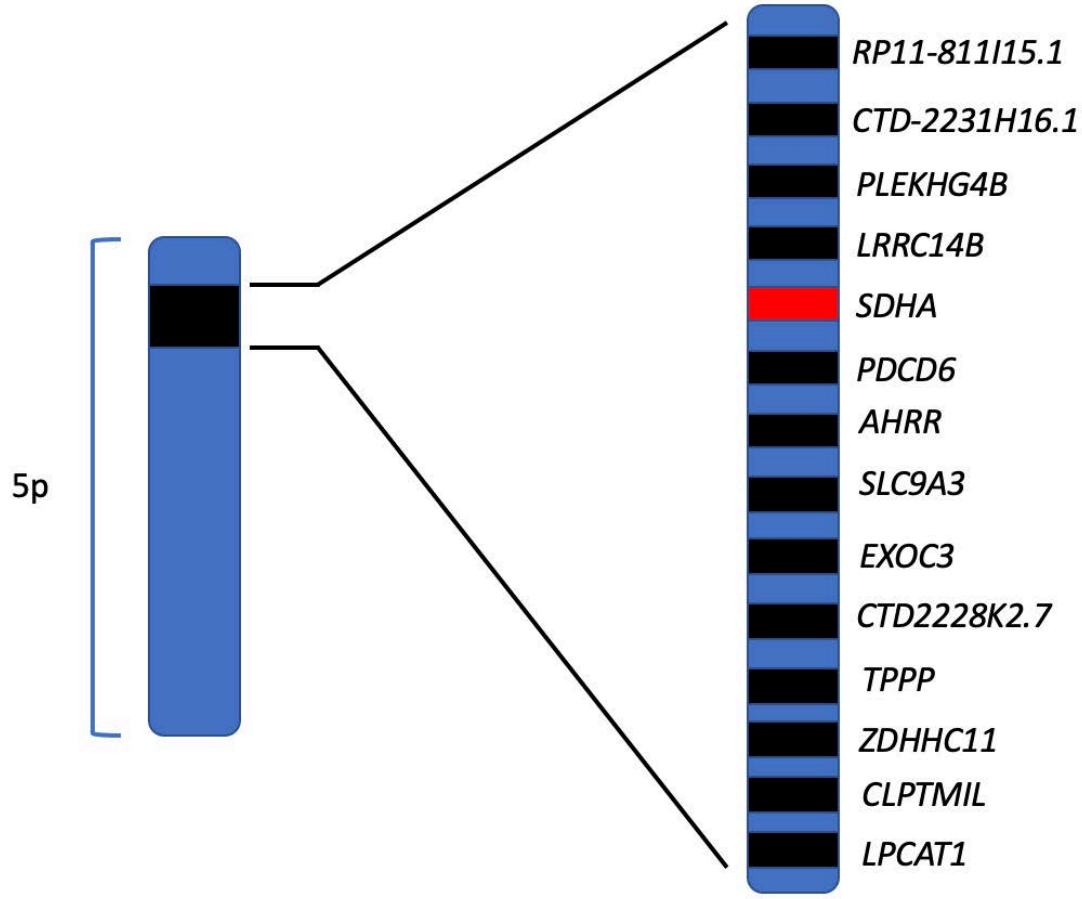

Figure 4.2: Map of *SDHA* fusion partners. All fusion partners were contained between chr5:58,198 – chr5:145,480. *SDHA* highlighted in red.

Fusion partners unique to tumours were *RP11-811I15.1, LRRC14B* and *TPPP*. Fusion partners unique to normal samples were *AHRR, EXOC3, CLPTM1L, LPCAT1, CTD-2228K2.7* and *ZDHHC11.* Shared fusion partners were *SLC9A3, PLEKHG4B, CD2231H16.1* and *PDCD6*. Table 4.5 provides a detailed analysis of significant fusions occurring within tumour samples.

Table 4.5: Significant *SDHA* fusions in tumour samples. All chromosomal co-ordinates are for Chromosome 5. Break Point Location 1 gives the functional region covered by Start 1 to End 1. Break Point Location 2 gives the functional region covered by Start 2 to End 2.

Sample ID	Start 1	End 1	Break point location	Start 2	End 2	Break point location	p-Value	Gene1	Gene2	Supporting reads
3	144,486	144,686	Intron 5/19	224,501	224,701	Exon 3/13-Intron 3/12	2.49E-02	*PLEKHG4B*	*SDHA*	3
3	155,017	155,217	Intron 8/19	254,460	254,660	Exon 12/13-Intron 12/12	2.40E-04	*PLEKHG4B*	*SDHA*	8
7	94,476	94,789	Long non-coding RNA	225,438	225,638	Exon 4/13-Intron 4/12	3.21E-02	*CTD-2231H16.1*	*SDHA*	4
9	92,336	92,744	Long non-coding RNA	218,363	218,563	Exon 1/13-Intron 1/12	1.79E-04	*CTD-2231H16.1*	*SDHA*	16
9	92,356	92,568	Long non-coding RNA	225,461	225,662	Exon 4/13-Intron 4/12	4.91E-02	*CTD-2231H16.1*	*SDHA*	3
9	155,788	155,988	Intron 9/19	218,390	218,590	Exon 1/13-Intron 1/12	7.14E-03	*PLEKHG4B*	*SDHA*	5
9	193,081	193,334	Intron 1/1	256,348	256,548	Exon 13/13-Exon 13/13	1.79E-04	*LRRC14B*	*SDHA*	16
11	58,386	58,586	Single exon processed pseudogene	235,207	235,420	Exon 7/13 - Intron 7/12	3.38E-02	*RP11-811I15.1*	*SDHA*	10
11	92,345	92,681	Long non-coding RNA	254,463	254,663	Exon 12/13-Intron 12/12	4.13E-02	*CTD-2231H16.1*	*SDHA*	9
3	251,049	251,249	Exon 11/13- Intron 10/12	481,884	482,239	Intron 8/16-Intron 7/16	3.60E-03	*SDHA*	*SLC9A3*	4
7	224,459	224,659	Exon 3/13 - Intron 3/12	481,884	482,239	Intron 8/16-Intron 7/16	3.21E-02	*SDHA*	*SLC9A3*	4
7	251,444	251,644	Exon 11/13 - Intron 11/12	481,882	482,236	Intron 8/16 - Intron 7/16	3.21E-02	*SDHA*	*SLC9A3*	4
7	254,391	254,591	Intron 11/12 - Intron 12/12	481,883	482,237	Intron 8/16 - Intron 7/16	3.21E-02	*SDHA*	*SLC9A3*	4
9	225,458	225,658	Exon 4/13-Intron 4/12	481,884	482,395	Intron 8/16-Intron 7/16	7.14E-03	*SDHA*	*SLC9A3*	5
9	228,249	228,449	Exon 6/13-Intron 4/12	353,544	353,744	Intron 4/13-Exon 3/4	2.13E-02	*SDHA*	*AHRR, PDCD6*	4

9	228,249	228,449	Exon 6/13-Intron 4/12	423,773	423,973	Intron 8/13-Exon 5/9	2.13E-02	*SDHA*	*AHRR*	4
9	233,590	233,790	Exon 8/13- Intron 6/12	481,817	482,171	Intron 8/16-Intron 7/16	2.13E-02	*SDHA*	*SLC9A3*	4
9	233,590	233,790	Exon 8/13- Intron 6/12	673,171	673,371	Intron 2/3-Intron 2/3	2.13E-02	*SDHA*	*TPPP*	4
25	251,049	251,249	Exon 10/13-Intron 10/12	473,147	473,365	Exon 17/17-Exon 17/17	1.77E-02	*SDHA*	*CTD-2228K2.7, SLC9A3*	4

4.2.2 Western Blots

The results of the RNA fusion panel were inconclusive, as not all tumour samples had fusions, and that a large number of normal samples also had evidence of fusions. For this reason, the protein expression of *SDHA* needed investigating in both tissue types. Protein lysates were made from three paired samples of WDLS and matched normal, and four WDLS only. These did not correspond with FFPE samples from the RNA fusion panel (see Table 4.6)

Table 4.6: Demographics and histology for Western blotting and colorimetric assay samples. T = Tumour, N = Normal. All samples underwent testing for MDM2 amplification.

Sample ID	Gender	Age	Histology	MDM2 Amplified
T1	M	64	90% WDLS/10% DDLS	Yes
N2				
T3	F	62	100% WDLS	Yes
N4				
T5	M	63	100% WDLS	Yes
N6				
T7	F	60	100% WDLS	No
N8				
T9	M	76	80% WDLS/20% DDLS	Yes
T10	F	61	100% WDLS	Yes
T11	F	74	100% WDLS	Yes
T12	F	73	100% WDLS	Yes
T13	F	72	100% WDLS	Yes

Western blots were performed, probing for Actin, SDHA and SDHB. Actin was chosen as a loading control protein as it shows steady expression profiles in adipose tissue (200). In pairs 2 and 3 SDHB was probed for, as when SDHA is lost the SDH complex is lost, resulting in no SDHB expression, thus acting as a validation tool (201).

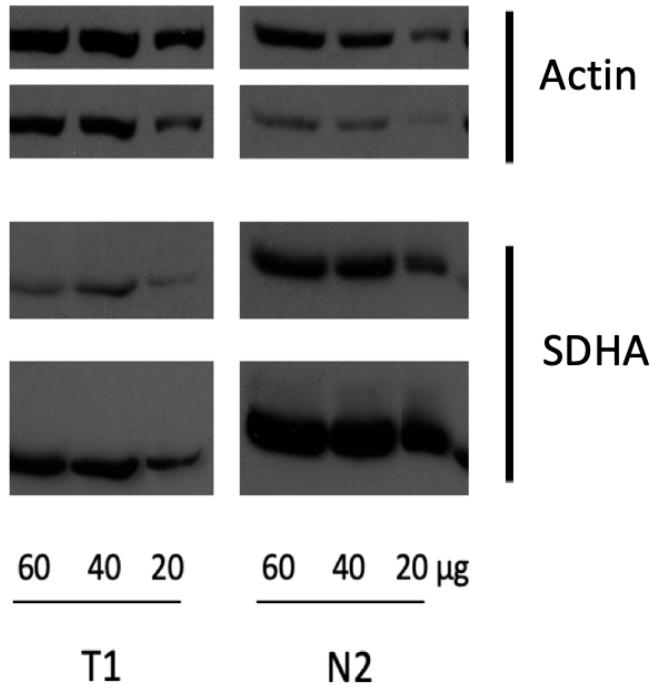

Figure 4.3: Western blots for SDHA and Actin in T1/N2. Three protein quantities were loaded (20,40,60 µg). As a pilot, two different exposure lengths were tested (30 s and 5 minutes). N2 shows subjectively higher protein concentrations for SDHA compared to T1.

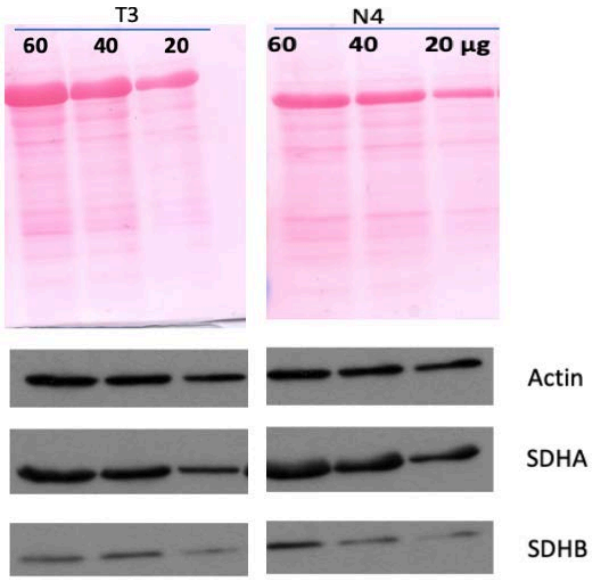

Figure 4.4: Western blots for Actin, SDHA and SDHB in one paired WDLS sample (T3/N4). Three protein quantities were loaded (20,40,60 µg). Ponceau stain included. There is comparable SDHA/B and Actin protein quantities between T3 and N4.

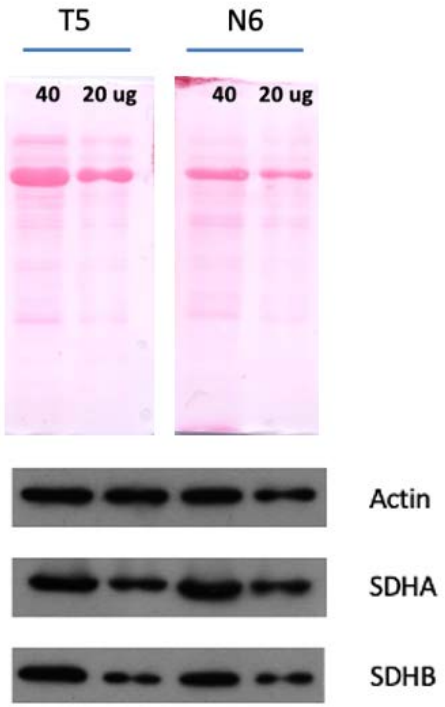

Figure 4.5: Western blots for Actin, SDHA and SDHB in one paired WDLS sample (5/6). Two protein quantities were loaded (20,40 µg). Ponceau stain included. There is comparable SDHA/B and Actin protein quantities between T5 and N6.

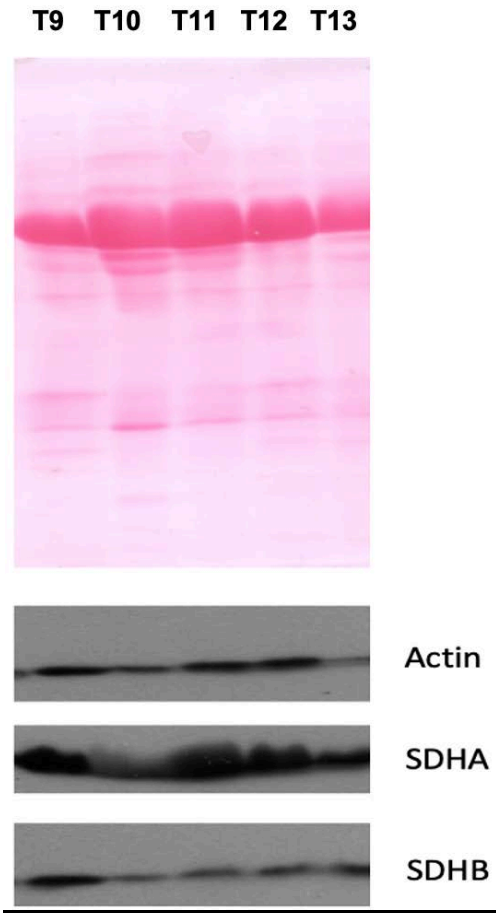

Figure 4.6: Western blots for Actin, SDHA and SDHB in 5 WDL only samples (7-11). 40 µg of protein was loaded per sample. Samples T9-13 show evidence of SDHA and SDHB expression.

Figure 4.3 compared a paired tumour and normal sample. It appears there is a higher level of SDHA expression in the normal sample compared to tumour. This is evident across three loading amounts, and in both exposures. Paired samples 3/4, 5/6 and 7-11 differ. In the pairs T3/N4 and T5/N6 there appears to be similar expression of SDHA between tumour and normal. In tumour samples 7-11 SDHA is consistently expressed. Importantly in samples 3-11 SDHB expression was retained in tumour samples.

For a more objective analysis, band densities were then quantified using ImageJ, as described in Chapter 2.

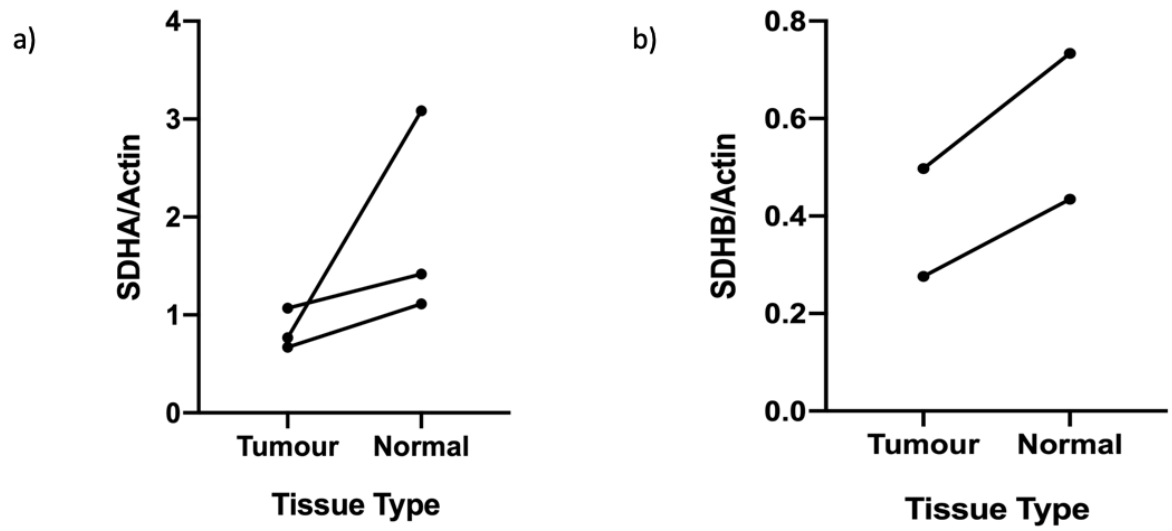

Figure 4.7: SDHA and SDHB expression normalised against Actin in tumour and matched normal (plotted on X-axis). Expression values were generated by band densities, measured in semi-quantitative fashion by ImageJ.

Figure 4.7(a) shows a consistent difference in band densities of SDHA between tumour and normal when normalised to Actin. Normal samples have a relative increase in SDHA expression compared to paired tumour. This is more marked in the first pair. However, when tumour and matched normal were compared with an unpaired t-test there was no significant difference between SDHA ($p=0.173$) or SDHB ($p=0.399$).

4.2.3 SDHA Colorimetric Assay

200 mg of four paired WDLS and normal were assessed for SDHA activity using a colorimetric assay. Pairs 1-3 correspond to T1-N6, as per Table 4.6. The fourth pair (T7+N8) is additional 100% WDLS and normal tissue.

The absorbance values at 600 nm were plotted against time (as per Figure 4.8). As the degree of downward slope defines the activity of SDHA, linear regression was performed using GraphPad (Prism, USA) to quantify slope values. The slope of

each sample with malonate, was subtracted from the paired sample without malonate. The mean of the duplicate samples was calculated and is show in Figure 4.9.

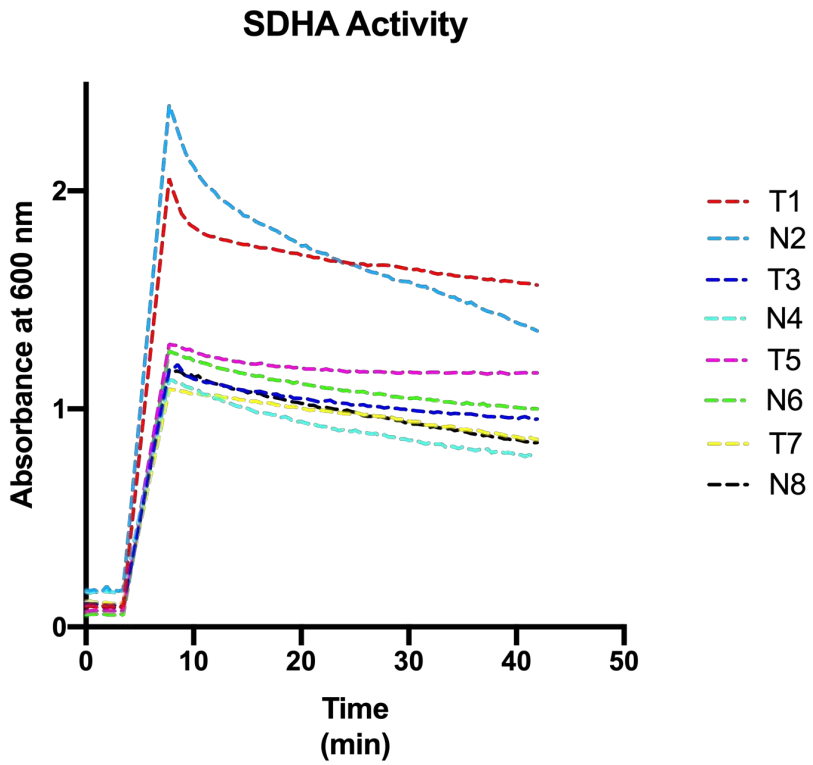

Figure 4.8: Line graph of *SDHA* activity in four paired sampes, measured through calorimetric assay. Absorbance at 600 nm is plotted against time. A more rapid decline indicates increased *SDHA* activity. The early rise in the graph is caused by the addition of DC-PIP.

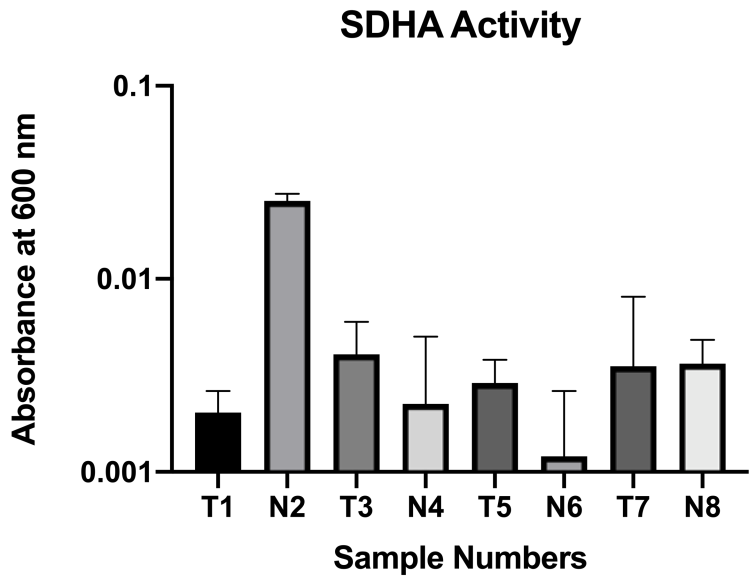

Figure 4.9: Bar chart of *SDHA* activity for four paired samples. Absorbance figures (y-axis) determined by linear regression using degree of slopes from Figure 4.7. Y-axis plotted using log_{10} values. Error bars are displayed.

4.2.4 RNA Sequencing

Total RNA sequencing using 150 base pair paired end sequencing was performed to look for evidence of *SDHA* fusions and relative expression of *SDHA* across four paired samples from the SHARPS study (Table 5.1). As per Table 5.1, sample 9 is an additional dedifferentiated tumour sample from Patient 4 (Sample 8 is well differentiated, Sample 9 is normal). Samples were loaded on a 300-cycle high output flow cell and ran on a NextSeq. For detailed methods, see Chapter 2.14. The run yielded 97.34 gigabase pairs, with a quality score (Q30) of 94.08%. The read depth was 50 M reads per sample. The density of clusters on the flow cell was 134-151 K/ mm^2.

A fusion calling pipeline (STAR-Fusion) with a minimum of 50,000 base pairs between ends was run using Partek Flow. The default settings for STAR-Fusion

were used throughout. The number of all significant gene fusions per sample are shown in Table 4.7.

Table 4.7: Significant gene fusions per sample ($p<0.05$).

Sample ID	Significant Fusions
1	77
2	277
3	56
4	11
5	23
6	20
7	6
8	19
9	130

None of the samples contained significant levels of *SDHA* gene fusions.

Expression patterns of *SDHA* across the samples was analysed. Log fold-change of *SDHA* expression in normal versus tumour was 3.58 which was significant ($p=0.016$). Expression values are given in Table 4.8, with corresponding bar chart in Figure 5.13.

Table 4.8: Normalised (TMM) *SDHA* expression values by sample

Sample ID	Expression value
1	0.0280884
2	0.398605
3	0.0777207
4	0.0884677
5	0.0574903
6	0.127852
7	0.0563664
8	0.140474
9	0.0443362

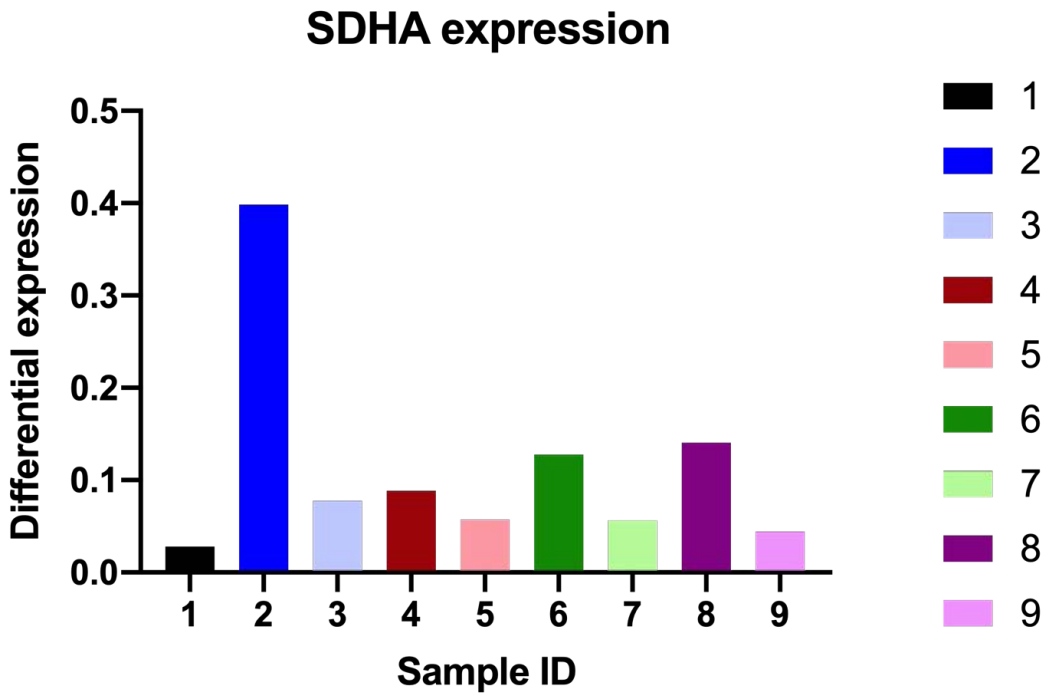

Figure 4.10: Bar chart of SDHA expression in the SHARPS study (Chapter 5) by sample ID. Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5 (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour). All paired normal samples have a higher relative expression than paired normal.

Across all paired samples, SDHA expression was lower in tumour samples compared to paired normal. This was also the case for Patient 4, where both WDLS and DDLS components of the tumour had lower expression values than paired normal. When expression values were compared with a Wilcoxon signed rank test the difference was significant (p=0.02). Other differentially expressed genes are discussed in Chapter 7.

4.3 Discussion

The *SDHA* fusion panel showed 20 statistically significant fusions in 32 samples, across both normal and tumour samples. Gene fusions can arise through various mechanisms. Two possible explanations for the observed fusions – where two genes are located on the same chromosomal strand – are deletions and

translocations. 50% of prostate cancer is caused by the *TMPRS2/ERG* fusion, which results from a 2.8 Mb deletion (202). In this data, when fusions occurred, they were between genes present on 5p, lying within 1.39 Mb of each other. Deletion as a mechanism is therefore a strong possibility. Another possibility is translocation, commonly seen in sarcomas, where a small genomic fragment breaks off and reattaches elsewhere in the genome. Myxoid liposarcoma is driven by a translocation between chromosome 12 and 22 (t(12;16)(q13;p11) resulting in FUS-CHOP (203).

Whether the resulting fusion genes have any biological impact requires further analysis. In myxoid liposarcoma the promoter of FUS binds to the complete coding region of DDIT3. The resulting protein acts as an aberrant transcription factor acting on a number of downstream target genes (204). Unique fusion partners in tumour samples were *RP11-811I15.1*, *LRRC14B* and *TPPP*. The only referenced article for *RP11-811I15.1* classes it as a processed pseudogene acting as a promoter (205). In its fusion to SDHA it could increase transcriptional activity, rather than supress it. The breakpoint in *TPPP* was intronic. Intronic breakpoints are more common as these are 35 x longer on average than exons, but rarely disrupt coding sequences (206). Therefore, this data does not consistently support a promoter/proto-oncogene model, such as observed in myxoid liposarcoma.

Another oncogenic mechanism that requires further discussion is that of oncogenic breakpoints. The majority of the breakpoints in *SDHA* were exonic in location (18/19 cases). Furthermore, when these breakpoints occurred, they were exonic in SDHA, whilst intronic in their respective fusion partner. Exonic loss could disrupt the gene's

open reading frame, leading to nonsense mediated decay and a truncated protein (207). As *SDHA* is a recognised tumour suppressor gene, a potential mechanism exists by which exonic breakpoints could be implicated in oncogenesis.

It is important to address the high number of normal samples that contained gene fusions (15/19). Gene fusions were once thought to be unique to cancer, but as sequencing has progressed numerous studies have shown high levels of fusion transcripts in normal tissues. In a recent fusion screen of The Cancer Genome Atlas, 192 gene fusions were identified in 364 normal tissue samples (208). The authors hypothesised these were a way of increasing proteomic diversity.

Interestingly, 14/15 adjacent tissue samples contained fusions, whereas only 1/4 non-adjacent samples (testicular/skin) did. A recent theory regarding fusion transcripts present in normal tissues could therefore be relevant. Oliver et al. assert that low levels of pathogenic fusions occur in select tissues and remain benign due to an unsuitable microenvironment or immune detection. However, subclonal events may be a precursor to malignancy when the surveillance and repair mechanisms are escaped (209). This is particularly relevant in liposarcoma where many believe the retroperitoneal fat harbours a genomic field change, as local recurrence is so common (35).

To assess whether exonic breakpoints could alter the final protein product, western blots were performed. *SDHA* expression was retained across the tumour samples. *SDHB* expression was also retained, which cannot be present without *SDHA*. On semi-quantitative analysis with densitometry, all pairs had a higher expression in

normal samples compared to tumours. However, the difference was not significant (p=0.173).

In theory, *SDHA* could be expressed but lack functionality. For this reason, an *SDHA* colorimetric assay was performed to test *SDHA* activity. Pair one showed a large difference as per the western blots, with increased *SDHA* activity in the normal. The remaining pairs (2+3) showed the converse with higher activity in tumours, with a similar activity level shown in pair four.

Burnichon et al. describe a missense mutation in *SDHA* (c.1765C>T;pArg589Trp) resulting in a loss of *SDHA* and *SDHB* on immunostaining. They further induced this mutation in a yeast model, which in functional studies showed no *SDHA* activity. It is likely the observed mutation leads to an error in the reading frame, and a non-functional protein. Although our data shows multiple exonic breakpoints, this may not have altered the final reading frame. However, the combination of exonic breakpoints, and intronic binding partner location, does increase the chances of protein function being affected. Unfortunately, our fusion panel used historical FFPE samples, whereas the western blots used prospective fresh tissue from different patients. For this reason, we cannot compare patients exonic breakpoints with final protein products.

The western blots were limited by a lack of paired samples for comparison due to their rarity. Secondly, *SDHA* is expressed in low levels in adipose tissue according to the Human Protein Atlas, which may have made it more difficult to detect subtle differences in expression patterns (210). Nevertheless, adipose tissue is highly

metabolically active and prefers the TCA cycle as its primary biochemical pathway, with adipocytes using 15% of circulating glucose (211). *SDHA* is therefore unlikely to be poorly expressed in adjacent fat, which then acts as a good comparison.

The fusion calling pipeline found no significant *SDHA* fusions in any of the SHARPS cohort. It did find evidence of other fusions ranging from 6-277 significant fusions per sample. It is feasible that fusions were present and not detected as our flow cell metrics showed under clustering, i.e the input library concentration was too low, a recognised issue when using the bioanalyser to perform library quantification. Secondly, RNA sequencing can suffer from poor sensitivity for detecting fusion genes expressed at a low level or diluted by non-malignant cells (212). It also struggles to detect chromosomal rearrangements that fuse a promoter upstream of a separate gene (213). Lastly, it is possible that only some WDLS contain fusions whereas these did not. The preliminary experiments and the fusion panel in chapter four found fusions at a rate of 4/8 (50%) and 5/13 (39%) respectively. This dataset simply may not have been large enough, compounded by the fact that Patient 2's final histology was equivocal for a diagnosis of WDLS.

An important consideration is the concept of intratumoural heterogeneity (ITH). ITH refers to the biological and genomic variations within a tumour lesion (214). Gerlinger et al. biopsied different areas of renal cell cancers and found over 2/3 somatic mutations were not detected in all the samples (215). Retroperitoneal liposarcomas are the largest tumours on record with a median size of 20 cms, so it is not unreasonable to suggest a high degree of ITH could exist. Future experiments should consider comparing the genomes of different areas of these tumours.

Although no statistically significant fusions in SDHA were found, there was evidence of impaired *SDHA* expression in tumours. As figure 4.10 shows, there was higher *SDHA* expression in normal tissues (p=0.01) with a log-fold change of 3.58. Patient 2's *SDHA* expression was similar (0.07 v 0.08, tumour v normal) although this was likely a benign retroperitoneal lipoma. The results then suggest a downregulation of SDHA, rather than a complete loss of expression. As fusions cannot explain this, alternative methods of impaired SDHA transcription need considered. Truncating and missense germline mutations in *SDHA* are responsible for 1/3 of all SDH-deficient GISTs, and also reported in paragangliomas and phaeochromoctyomas (195, 216). The epigenome could also be responsible, with a hypermethylated phenotype well recognised in this subset of tumours (217). Regardless of cause, whether this translates into metabolic dysregulation needs further investigation.

4.4 Conclusion

Based on the data presented in this chapter, we cannot say there is a loss of function in *SDHA.* Sequencing data also showed no evidence of *SDHA* fusions. Furthermore, there is evidence of both protein expression and *SDHA* activity in tumour samples. However, SDHA expression tended to be higher in the normal samples, and there was evidence of recurring exonic breakpoints in SDHA. To interrogate SDHA more confidently and its function in the TCA cycle, in vivo experiments were required in the form of metabolic tracing.

Chapter 5: In vivo assessment of WDLS using metabolic tracing

5.1 Introduction

Chapter four provided a genomic and proteomic assessment of *SDHA* fusions in WDLS. Even with the observation of recurrent exonic breakpoints in *SDHA*, it appeared expressed, albeit at lower levels, and functional in tumour samples. However, even though in vitro assessment of *SDHA* is useful, it may not accurately reflect the metabolic tumour microenvironment. For example, glutamine is a major contributor to the TCA cycle in cultured lung cancer cells but is a minority contributor for tumours in vivo *(218, 219)*. It is therefore important to assess in vivo *SDHA* functionality.

We hypothesised that if SDHA was non-functional in vivo, an increase in levels of succinate relative to fumarate would be observed (Figure 5.1). Levels of TCA intermediates, both relative and absolute, can be measured using Gas Chromatography-Mass Spectrometry (GC-MS). This is particularly relevant as there is now good evidence that TCA cycle intermediates also possess non-metabolic signalling functions (220).

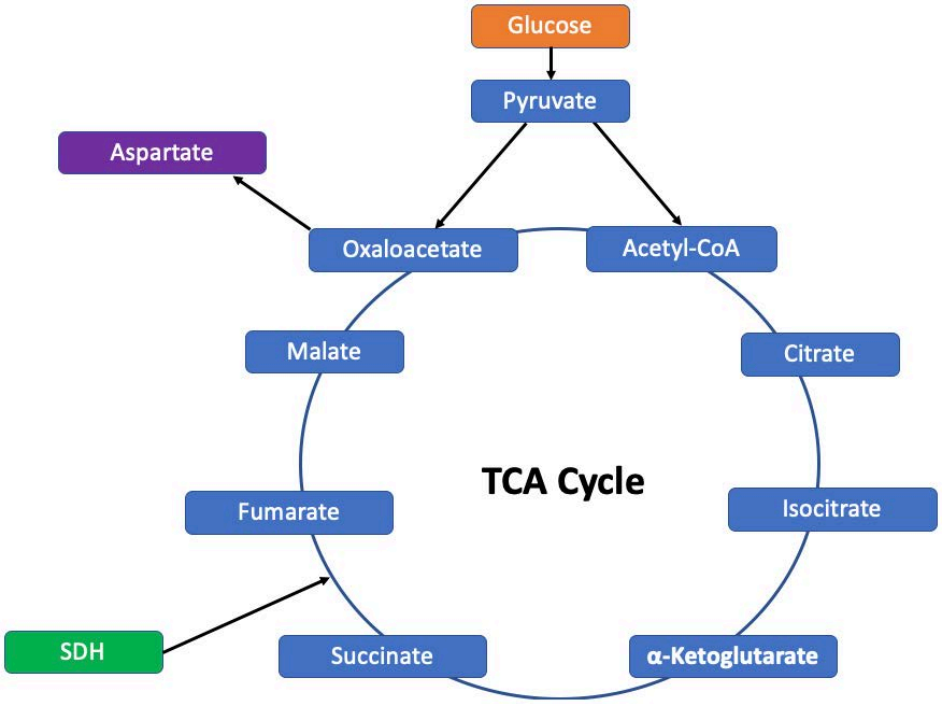

Figure 5.1: TCA cycle. Primary substrate glucose in orange, SDH converts succinate to fumarate in green. SDHA is one component of SDH. Key amino acid aspartate (purple) is made from oxaloacetate.

As stated in Chapter 4, succinate can inhibit prolyl hydroxylases and stabilise HIF-1a which has been shown to promote tumour progression. Therefore, quantitative measures of TCA intermediaries succinate and fumarate through GC-MS are required in the first instance.

However, while raw counts ('steady-state' measures) of metabolites are useful, they do not reflect the complete metabolic picture. High levels of metabolites (e.g., succinate) can be due to faster production or slower consumption, and differentiating between these two is crucial, especially when metabolite levels may be enhanced in a disease process, and could be inhibited therapeutically.

This problem can be solved by measuring metabolic flux analysis, which provides a measurement of pathway activity (221). One means of achieving this is by

administering a stable isotope-enriched tracer which is taken up by both healthy and tumour cells. The choice of tracer depends on the pathway of interest, and for the TCA cycle $^{13}C_6$ labelled glucose can be used. In this case, every carbon of glucose ($C_6H_{12}O_6$) is ^{13}C rather than ^{12}C – the most-common naturally abundant form – and differs by containing one more neutron, which acts as a 'label.'

When a labelled $^{13}C_6$ substrate is metabolised by cells, enzymatic reactions result in specific labelling patterns in downstream metabolites, specifically succinate, malate, citrate and alpha-ketoglutarate (222). These labelling patterns are detected using gas chromatography which vaporises and separates the analytes, feeding them into a mass spectrometer which measures the deflection of each analyte and the time taken to travel through. This combination creates peaks on a readout which can be mapped to spectral libraries, and thus metabolic flux is inferred (223).

Using $^{13}C_6$ metabolic flux analysis Faubert et al. demonstrated that lactate is a TCA cycle carbon source in non-small cell lung cancer whilst Maher et al performed similar experiments in patients with glioblastoma (224, 225). No such experiments have been carried out in sarcomas of any form. As liposarcomas grow to very large sizes, it is self-evident that they are able to guarantee a generous supply of metabolites to grow and divide, and are able to continue to do so without outstripping this supply or vasculature.

To further explore *SDHA* in liposarcoma, an in-depth assessment of tumour metabolism in vivo is required. To facilitate this, a $^{13}C_6$ metabolic tracer was administered to patients undergoing surgery to remove a WDLS. Methodology from

Faubert and Maher as previously stated was used as a guide. Analysis was performed using standard GC-MS techniques. Metabolic tracing results was compared with RNA sequencing results (Chapter 4.2.4) assessing for the *SDHA* fusion in vitro.

We hypothesised that the labelling patterns of TCA cycle intermediates in tumours showing transcriptomic evidence of *SDHA* fusions would be consistent with a non-functional SDH complex. There have been several studies that describe the metabolic consequences of a non-functional SDH complex, mostly using in vitro methodology. These find a build-up of intracellular succinate with a concomitant decrease in fumarate. Those studies that use in vivo methodology describe labelling patterns consistent with SDH dysfunction, which require further discussion.

In simple terms, when $^{13}C_6$ glucose is introduced, by the action of pyruvate dehydrogenase, labelled two carbon is incorporated into the TCA cycle via Acetyl-CoA and produces M+2 and M+4 in the first and second rounds of the TCA cycle, respectively (226) (See Figure 5.2). However, as per Tennant et al. showed, SDH knockout cells rely on high levels of pyruvate carboxylase dependent aspartate production (227). This occurs through the action of pyruvate carboxylase in reductive carboxylation. Pyruvate carboxylase can incorporate one unlabelled carbon in the uniformly labelled pyruvate and generate M+3 oxaloacetate which can go to the TCA cycle and generate M+3 labelled intermediaries, namely M+3 fumarate. Therefore, one labelling pattern consistent with SDH dysfunction, would be elevated levels of M+3 fumarate.

Combining this we hypothesise that SDH dysfunction will result in:

1) High intracellular succinate
2) Low intracellular fumarate
3) Enhanced presence of M+3 fumarate

5.2 Results

Patient demographics and sample histology is summarised in Table 5.1.

Table 5.1: Demographics and histology of SHARPS cohort

Patient ID	Sample ID	Sample Type	Gender	Age	Tumour Histology	MDM2 Status
1 1	1 2	Tumour Normal	F	72	Well differentiated liposarcoma	Amplified
2 2	3 4	Tumour Normal	M	54	Retroperitoneal Lipoma	Negative
3 3	5 6	Tumour Normal	F	59	Well differentiated liposarcoma	Not tested
4 4 4	7 8 9	Tumour Normal Tumour	M	48	Well differentiated liposarcoma Dedifferentiated liposarcoma	Amplified

Patients 1 and 3 had tumour taken with matched healthy fat from adjacent to the tumour. Patient 2's pre-operative imaging was equivocal for the nature of the tumour. It was resected as it had grown on surveillance. On sampling the tumour, macroscopically it appeared to be a large lipoma. As a control, subcutaneous fat

was taken from the subcutaneous layer which was separate from the tumour capsule. Patient 4's pre-operative biopsy was indicative of a WDLS, but on sampling found to have areas of dedifferentiation. For this reason, a sample of WDLS (Sample ID 7), healthy fat (Sample ID 8) and DDLS (Sample ID 9) was taken.

Pathological examination revealed Patients 1 and 3 to have WDLS. MDM2 amplification was present in Patient 1, but not carried out in Patient 3 as the diagnosis could be confirmed on microscopic features alone. Patient 2 had a retroperitoneal lipoma, a benign entity which was MDM2 negative. Patient 4 had a WDLS with areas of dedifferentiation, and was MDM2 amplified.

Polar metabolites were extracted from samples on two separate occasions to generate duplicates. Samples then underwent GC-MS testing with data analysis performed by MetaboliteDetector software. Averages from the duplicates were always reported. The labelling patterns of mass isotopomers are illustrated in Figure 5.2.

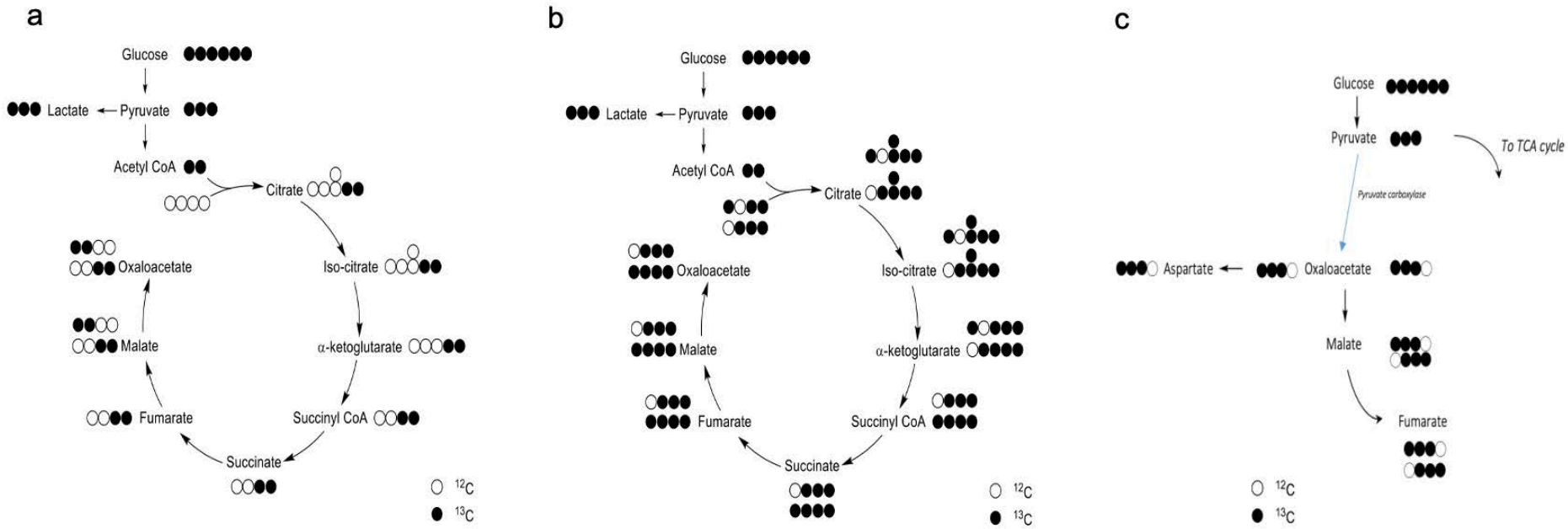

Figure 5.2: Labelling patterns of different mass isotopomers. Cycle starts with a glucose molecule fully labelled with C13. (a) 1st turn of the TCA cycle (M+2). (b) 2nd turn of the TCA cycle (M+3/4). (c) M+3 intermediaries produced through reductive carboxylation by the action of pyruvate carboxylase. Circles correspond to a carbon atom. Clear circles show naturally abundant carbon 12. Black circles show C_{13} tracer. As labelled isotopomers re-enter the TCA cycle, they become progressively more labelled (a-b), and distinct labelling patterns allow TCA labelling patterns to be distinguished from other pathways. Figure kindly created with help from Dr. Paul Walker, University of Birmingham.

5.2.1 TCA Cycle Abundance

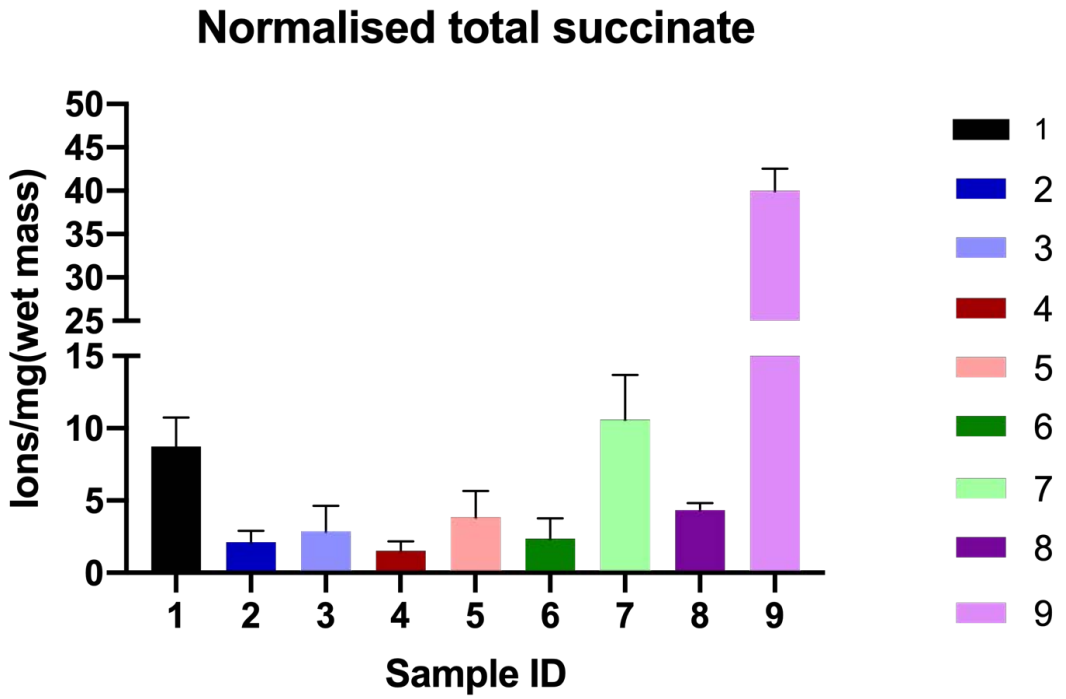

Figure 5.3: Normalised total succinate counts in ions/mg wet tissue mass. Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5 (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour). Y-axis is non-linear to allow for Sample 9.

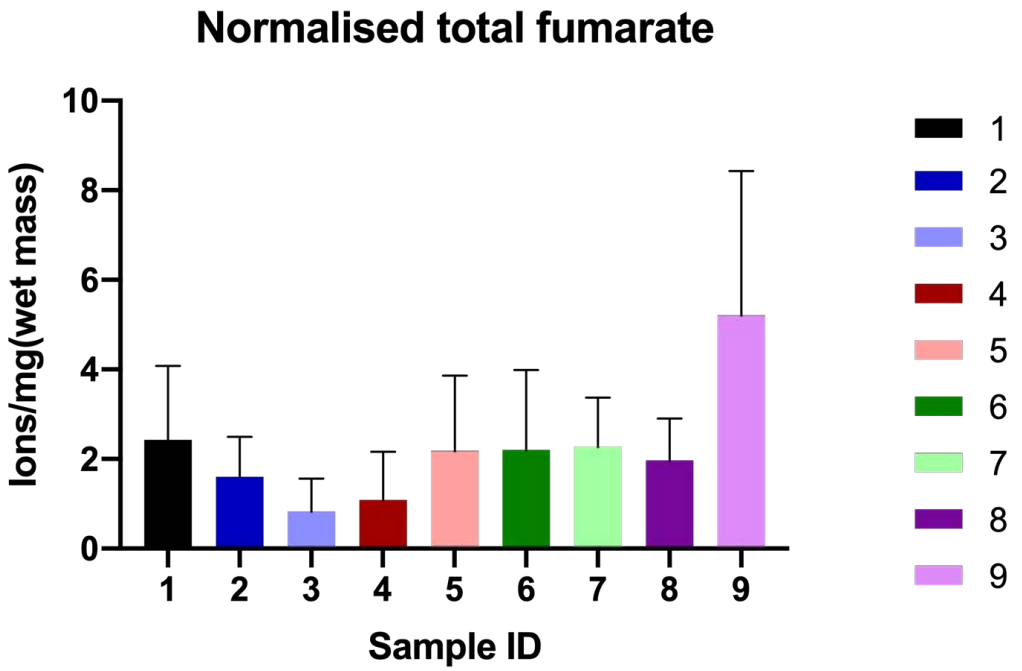

Figure 5.4: Normalised total fumarate counts in ions/mg wet tissue mass. Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5 (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour).

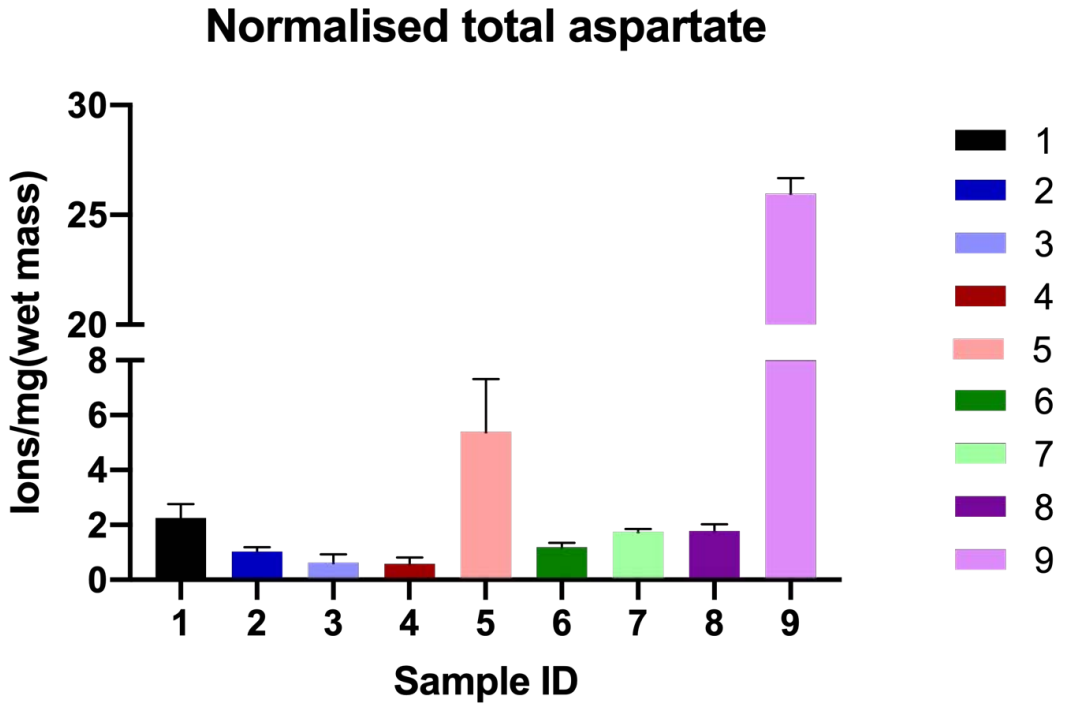

Figure 5.5: Normalised total aspartate counts in ions/mg wet tissue mass. Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5. (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour). Y-axis is non-linear to allow for Sample 9.

Metabolites were normalised to wet tissue mass to allow comparison between tumour samples. We found that succinate levels were consistently elevated in tumours compared to their paired normal samples (Figure 5.3). Interestingly, sample 9, which arose from a dedifferentiated area of tumour, had far higher levels of succinate.

Patients 1 + 4 had higher levels of fumarate in tumour compared with normal. This differed from patient 2, where a higher fumarate level was noted in the normal sample, and patient 3 where levels were similar (Figure 5.4). As with succinate Patients 1+ 3 had higher levels of aspartate in tumour compared to normal. Patients 2 + 4 had similar levels between tumour and normal (Figure 5.5). Although, as noted previously, the dedifferentiated component of Patient 4's tumour (Sample 9) showed very high levels of aspartate.

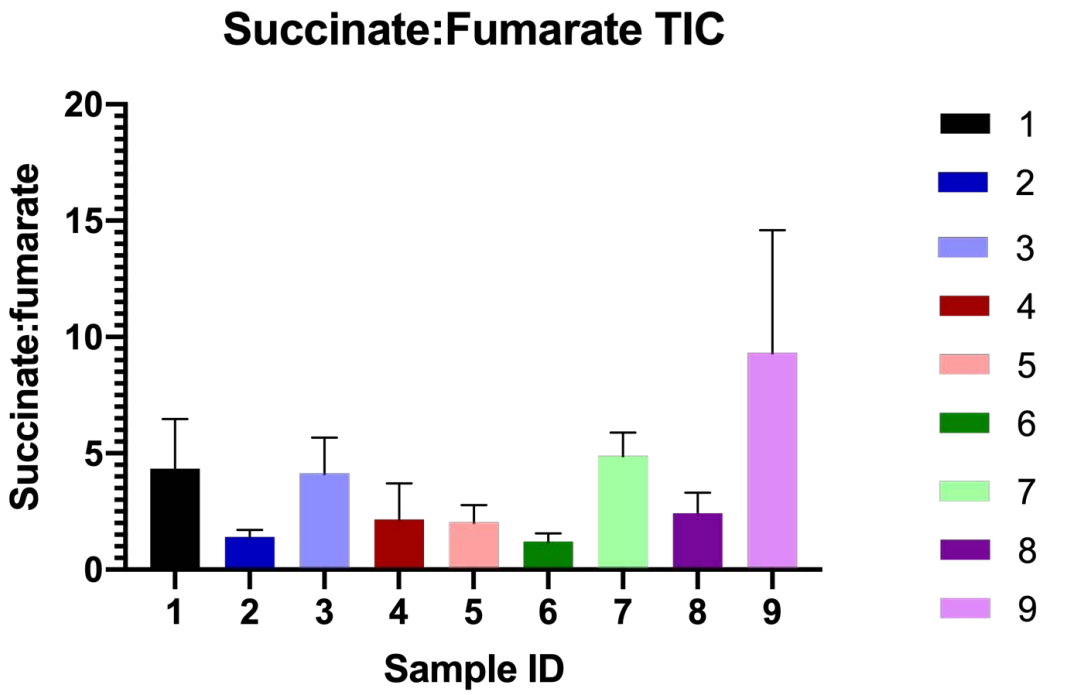

Figure 5.6: Ratios of Succinate to Fumarate Total Ion Count (TIC). Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5 (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour).

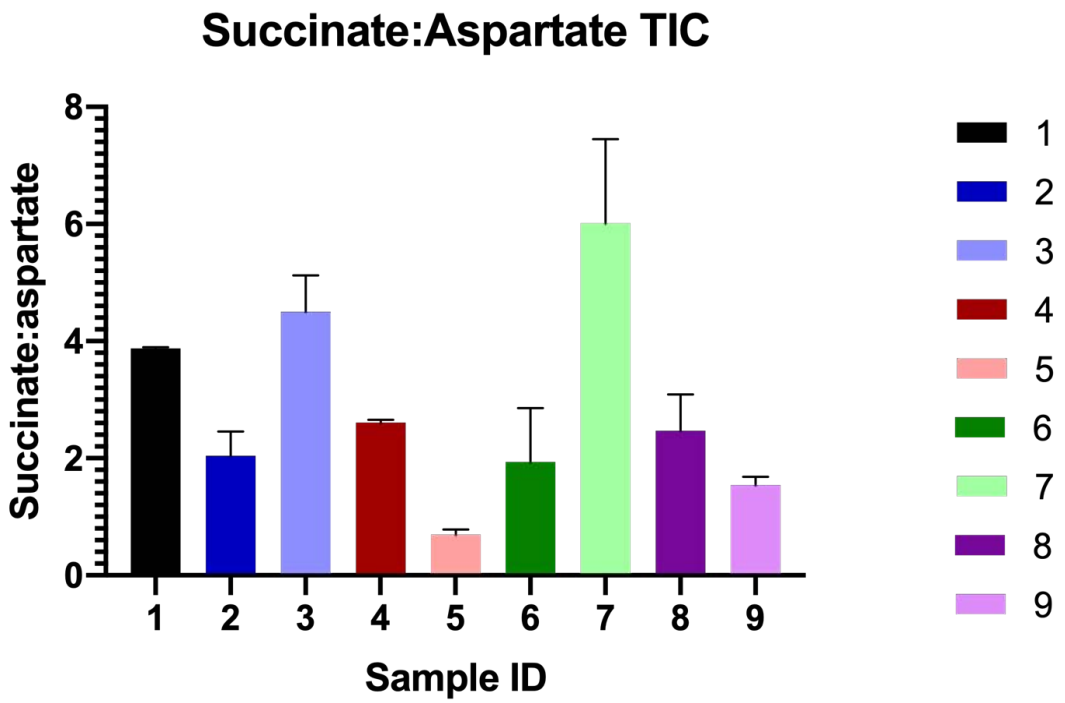

Figure 5.7: Ratios of Succinate to Aspartate Total Ion Count (TIC). Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5 (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour).

Increased succinate: fumarate ratios were observed in all tumours compared to normal. When compared with a Wilcoxon-signed rank test, this result did not reach significance (p= 0.064).

In patients 1,2 and 4 there appeared to be a trend towards increased succinate: aspartate ratio in tumours compared to normal, except patient 3 which is lower in tumour compared to normal. However, when compared with a Wilcoxon-signed rank test, there no significant difference was found (p=0.343). Sample 9, from a dedifferentiated part of the tumour has an elevated succinate: fumarate ratio, but a low succinate: aspartate ratio. This is driven by a very high level of aspartate in this tumour (25.98 ions/mg wet mass).

5.2.2 TCA Cycle Flux

As outlined in the introduction, the flux through the TCA cycle was monitored using $^{13}C_6$-glucose infusion given to patient prior to resection. The mass isotopomers – incorporation of ^{13}C atoms – into metabolites within the TCA cycle was examined, and is presented as a mass isotopomer distribution (MID), representing the fractional abundance of each isotopologue normalised to the sum of all isotopologues (228). TCA metabolites measured in plasma were analysed first, to show proof of ^{13}C enrichment in tissue. Then in order, glycolysis, TCA cycle and amino acids were examined.

5.2.2.1 Measurement of ^{13}C fractional enrichments in blood

Blood samples were taken at three different time points, to try to find the appropriate timepoint at which isotopic steady-state was reached. Mean time for tumour removal was 118 minutes, whilst mean time for tumour to be frozen when removed from the body cavity was 13 minutes (Table 5.2).

Table 5.2: Sampling times by patient. Sample 1 was taken prior to surgery. Sample 2 and 3 were taken at two separate time points from the beginning of surgery. Times also given for time taken to remove tumour surgically, and from tumour removal to snap frozen.

Patient	Sample 1	Sample 2	Sample 3	Tumour out	Tumour out to frozen
1	Prior to surgery	+ 1 hour	+ 2 hours	+ 2 hours 23 minutes	+ 9 minutes
2	Prior to surgery	+ 30 minutes	+ 1 hour 15 minutes	+ 1 hour 33 minutes	+ 12 minutes
3	Prior to surgery	+ 30 minutes	+ 2 hours 7 minutes	+ 1 hour 43 minutes	+ 10 minutes
4	Prior to surgery	+ 58 minutes	+ 1 hour 56 minutes	+ 2 hours 20 minutes	+ 25 minutes

M+3 labelling patterns in alanine, lactate and pyruvate were indicative of ^{13}C enrichment (see Figure 5.1 for labelling patterns).

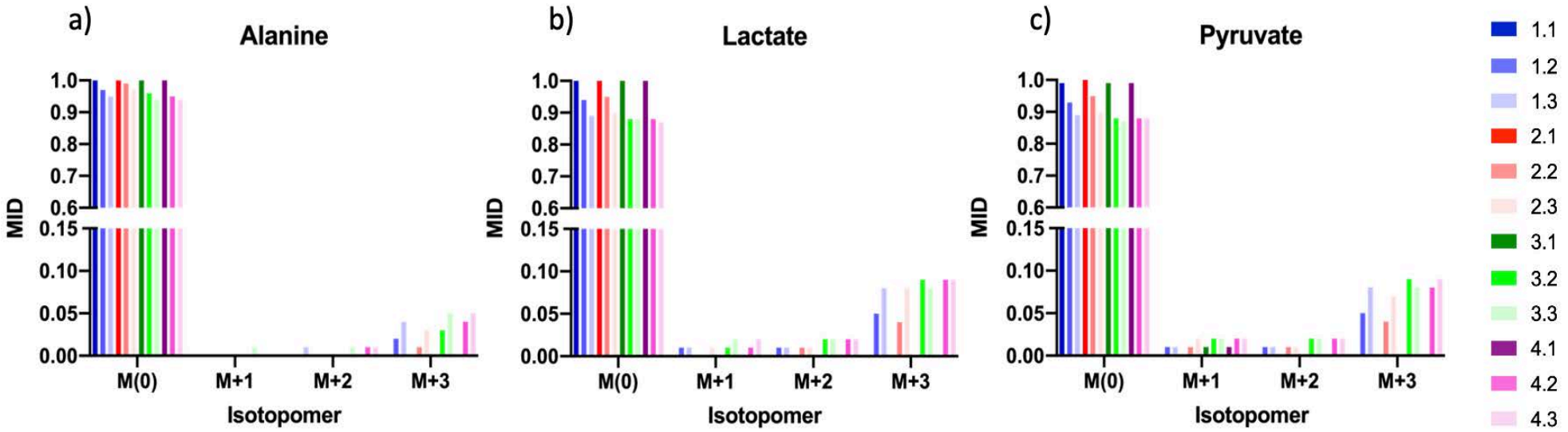

Figure 5.8: Blood plasma samples for Patients 1-4, showing mass isotopomer distribution (MID) of glycolysis metabolites (alanine, lactate and pyruvate) from $^{13}C_6$-glucose (%). Sample 1.1 corresponds to Patient 1, Sample 1. Sample 1 taken prior to surgery, sample 2/3 taken at time points afterwards (see Table 5.2). Non-linear y-axis to clearly delineate >M+1 isotopomer.

In Figure 5.8 each of the '.1' samples showed no evidence of M+3 isotopomer. This is expected as these samples were taken prior to surgery.

Samples '.2' and '.3' contained M+3 in generally increasing amounts. Patients 1 and 2 showed a continual increase in M+3 across all metabolites. Patient 4 showed increases in pyruvate and alanine over time, with lactate levelling off. Patient 3 showed pyruvate and lactate slightly dropping off at the third time point, whereas alanine increased.

Patients 1 and 2 may not have reached steady state within the tumour (where the proportion of label doesn't increase further) as isotopomer levels continued to increase. It is possible with one further time point we could prove steady state had been reached here. In patients 3 and 4, time points 2 and 3 are similar in lactate and pyruvate and may have reached steady state. A slight drop in patient 3's pyruvate and lactate could be explained by a longer interval (2 hours 7 minutes) to the final blood sample.

5.2.2.2 Measurement of ^{13}C fractional enrichments in tissue

Pyruvate and lactate were analysed to ascertain evidence of enrichment, but also to assess glycolysis. There was evidence of M+3 labelling patterns in pyruvate and lactate consistent with evidence of enrichment. During glycolysis glucose is ultimately broken down to pyruvate, which can feed into the TCA cycle or be reduced to lactate (229). Figure 5.8 shows M+3 lactate present in tumours, showing some degree of lactate metabolism, although less pronounced than their paired tumours.

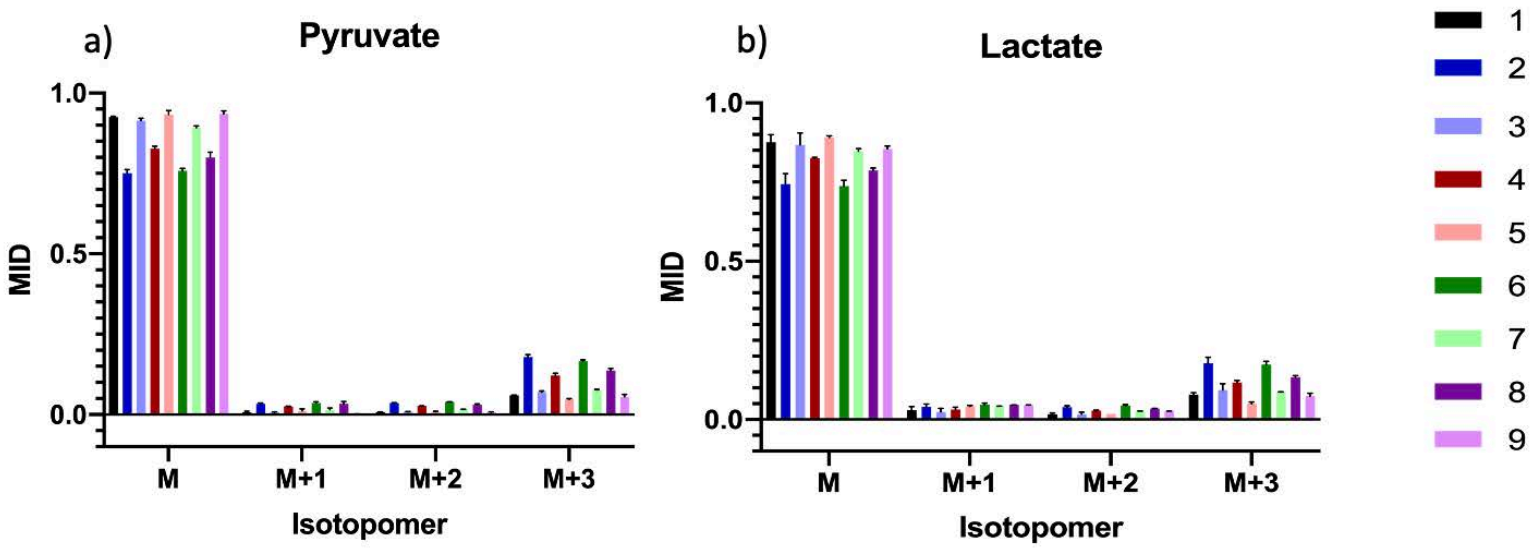

Figure 5.9: Mass isotopomer distribution (MID) of glycolysis metabolites from $^{13}C_6$-glucose (%) for Sample ID 1-9. Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5 (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour). MID is a % relative to natural abundance. Labelling in order of TCA cycle (a-f). M represents no label incorporation.

Of note is the differences between lactate incorporation in normal tissue (Fig 5.9 b) compared to in blood (Fig 5.8 b). There appeared to be higher labelling (MID) in normal tissue compared to the respective blood sample. When M+3 MID was compared between these by Wilcoxon signed rank test, the difference was significant (p=0.02). When compared between

tumours and blood this was not significant (p=0.88). High intracellular labelled lactate in normal tissues is therefore an area for discussion. Oxidative metabolism was then measured by examining TCA cycle intermediaries (Figure 5.10).

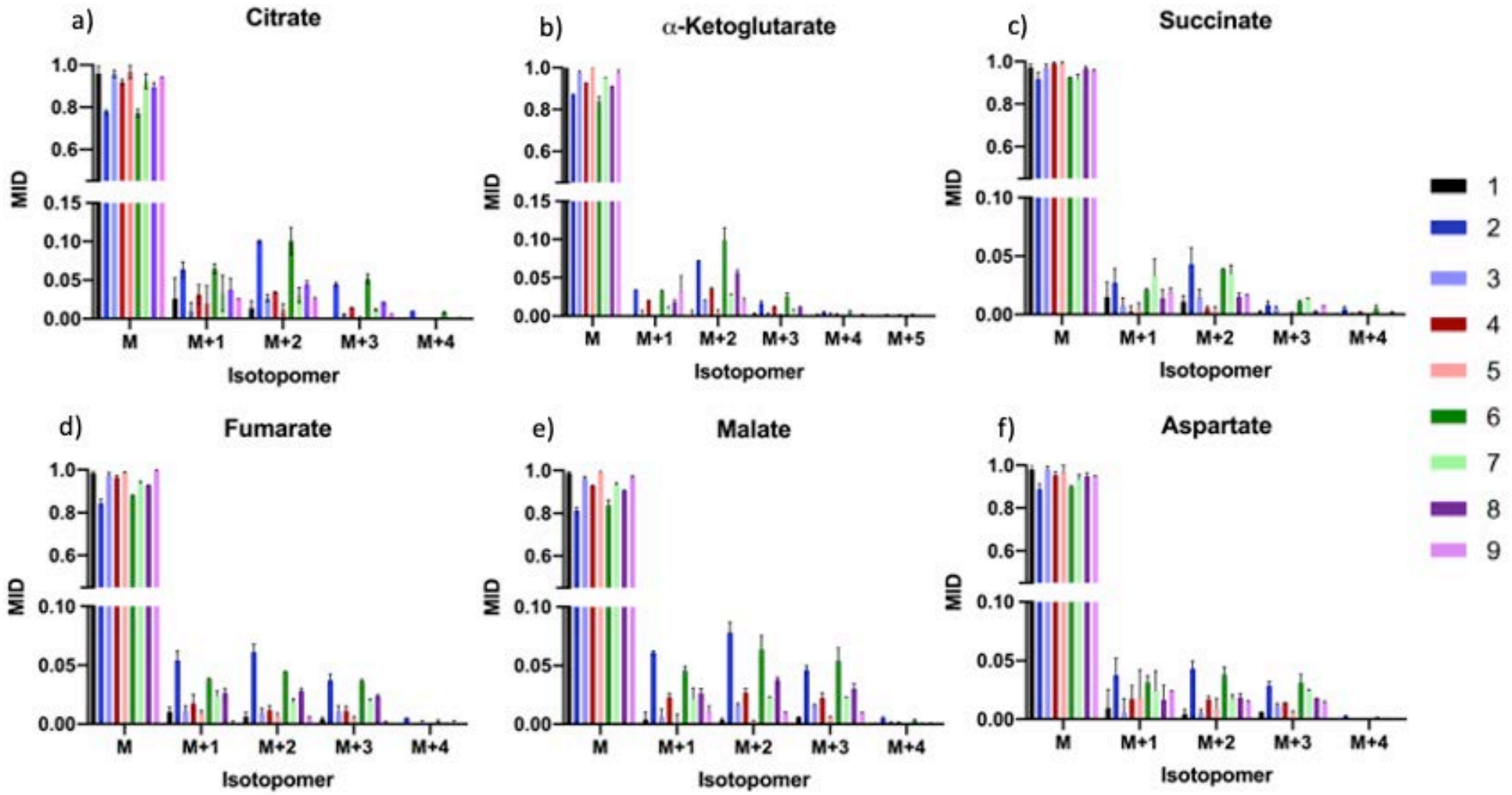

Figure 5.10: Mass isotopomer distribution (MID) of TCA cycle intermediates and aspartate from $^{13}C_6$-glucose (%) for samples 1-9. Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5 (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour). Non-linear y-axis to clearly delineate >M+1 isotopomer. Labelling in order of TCA cycle (a-f).

In general, we see low levels of label incorporation across the samples, with MID of no greater than 0.1 across all isotopomers M+1 or greater. This makes comparisons of samples more difficult, with an emphasis on describing any patterns that are observed. Nevertheless, in some samples up to 20% of metabolites were made of glucose carbons (see Sample 2, Figure 5.10), which shows successful sample uptake of the metabolic tracer.

All metabolites with M+2 were of interest as this indicates flux specifically through oxidative TCA cycle. All normal samples have a higher relative M+2 distribution than their paired tumours across each of the metabolites. The exception to this is succinate. In patients 2 + 4 there is higher M+2 in tumours compared to normal (0.014 v 0.004, 0.04 v 0.02, respectively). This pattern continues in M+3 samples. The next metabolite fumarate shows no significant differences between tumour and normal (0.01 v 0.01, 0.02 v 0.03). After this (malate/aspartate) the pattern of normal>tumour is seen again.

A more detailed examination of fumarate was then performed. M+3 labelling patterns in fumarate cannot occur by the conversion of succinate to fumarate, i.e through the action of SDH. Only M+2 labelling patterns are derived from this conversion. The M+3 isotopomers of fumarate are made by reductive carboxylation, through pyruvate carboxylase activity, and other mechanisms. Enhanced levels would be indicative of enhanced reductive fumarate synthesis, and help support a degree of loss of function of the SDH complex (227). The ratios of M+2/M+3 fumarate isotopomers are shown in Figure 5.11.

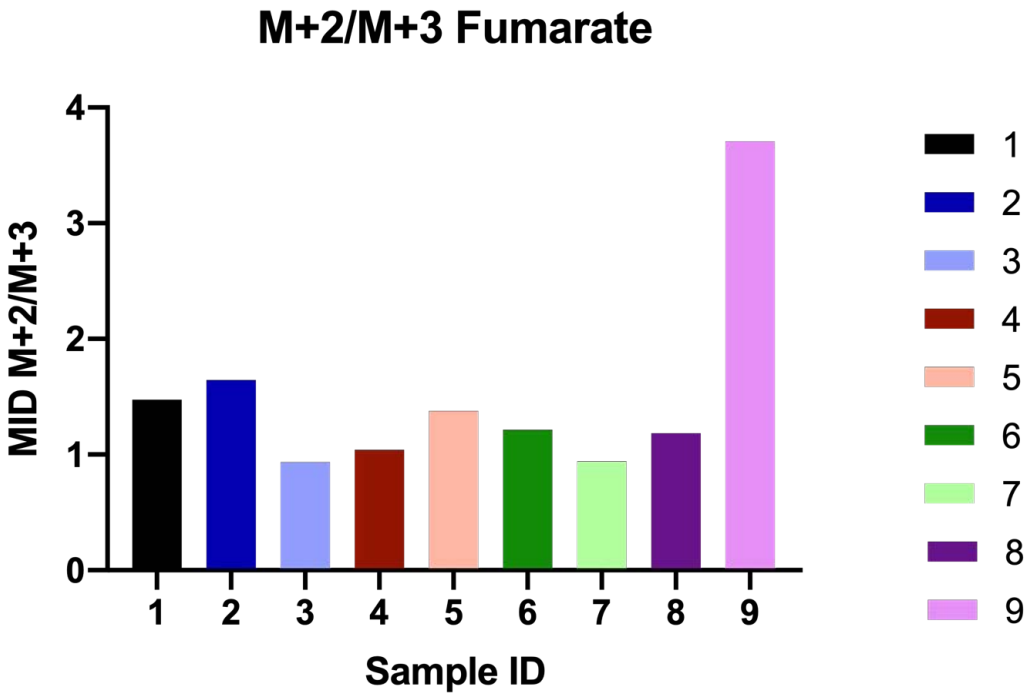

Figure 5.11: Ratio of M+2/M+3 fumarate isotopomers. Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5 (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour). M+2 isotopomers made through TCA cycle using SDH. An enhanced presence of the M+3 isotopomer would be indicative of increased reductive fumarate synthesis (via pyruate carboxylase), which SDH dysfunctional cells have been shown to favour (See Figure 5.2).

Pairs 1+2 show lower M+2/M+3 fumarate ratios. Pair 4 shows a lower ratio in the WD component, but not the DD component. Pair 3 has higher M+2/M+3 ratio in tumour. When these ratios were compared in tumour versus normal by the Wilcoxon signed rank test, it was not significant. Although the results are not significant, there is a pattern in the WD components of patients 1,2 and 3 of increased activity of reductive carboxylation (lower M+2/M+3 ratio in tumours), which could signal a dysfunctional SDH complex.

For amino acids serine and glycine there was negligible levels of label incorporation. In the case of glycine, no M+3 was recorded for any sample (Figure 5.12). As

observed in blood plasma, M+3 labelling patterns were observed in alanine across the samples indicating tracer uptake in the tissue sample.

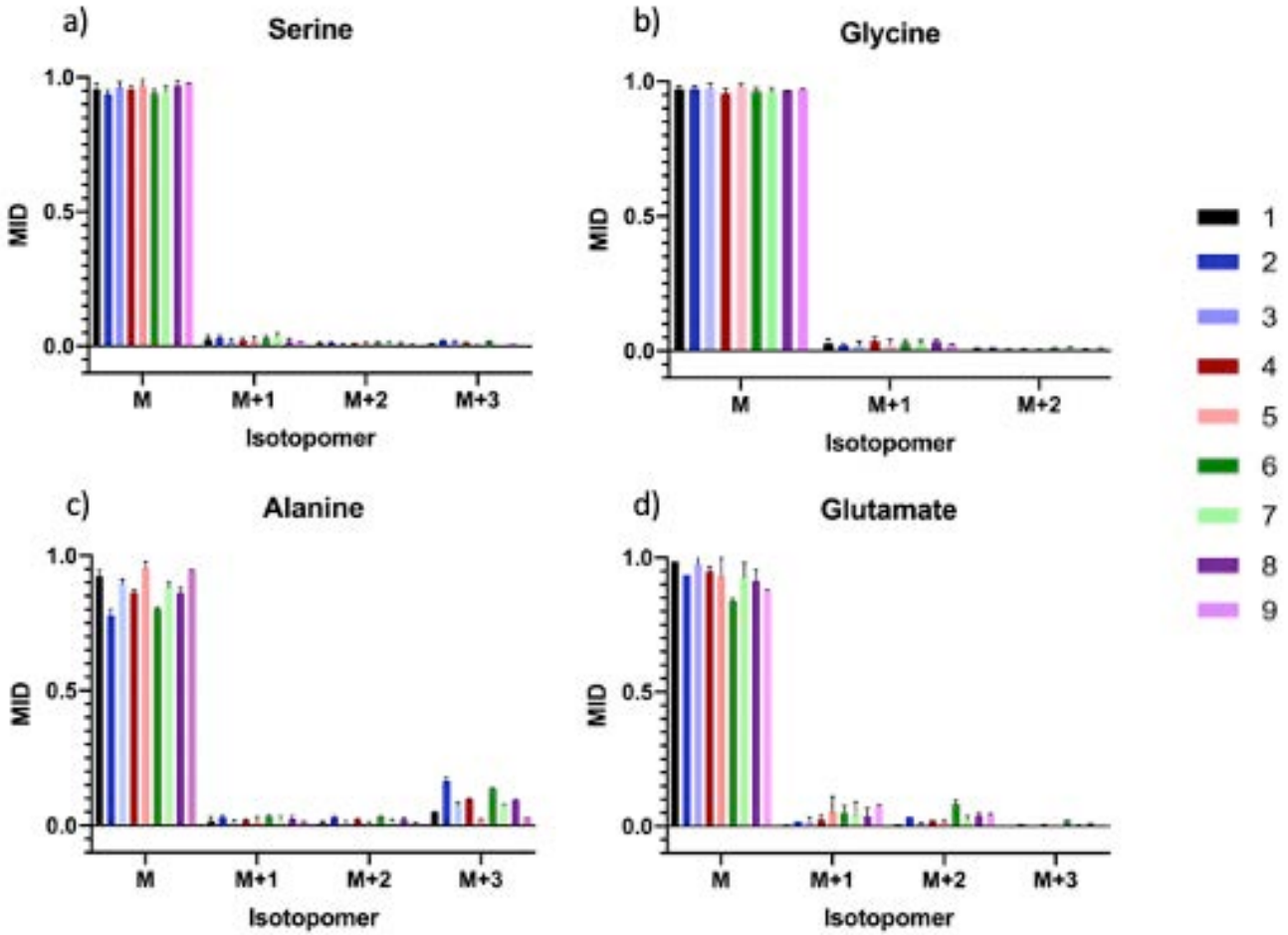

Figure 5.12: Mass isotopomer distribution (MID) of amino acids from $^{13}C_6$-glucose (%). Patient 1 – Sample ID 1 (Tumour) and 2 (Normal). Patient 2 – Sample ID 3 (Tumour) and 4 (Normal). Patient 3 – Sample ID 5 (Tumour) and 6 (Normal). Patient 4 – Sample ID 7 (WDLS Tumour), 8 (Normal) and 9 (DDLS Tumour). Glycine (b) contains negligible label incorporation. Alanine (c) displays M+3 labelling pattern, consistent with labelled glucose as feeding molecule.

5.3 Discussion

Chapter four data gave no convincing evidence of a loss of SDH function due to gene fusions, but a suggestion that *SDHA* expression was impaired. Metabolomic data interrogates the metabolic phenotype of WDLS showing increased TCA metabolite content within tumours, but lower levels of metabolic flux.

5.3.1 SDH function

We originally hypothesised that SDH deficient tumours would have elevated succinate: fumarate ratios which Figure 5.6 supports. Tumours consistently have elevated succinate: fumarate ratios, which trended towards significance (p=0.064). This was driven by elevated succinate levels in tumours, rather than lower fumarate levels. This supports our hypothesis, as succinate accumulation is the expected sequelae of SDH dysfunction. It is possible that lower levels of SDHA expression hinders the ability of the SDH complex to function at normal capacity but does not prevent it working completely.

However, measuring raw abundance does not allow us to assess flux. Changes in succinate could be indicative of increased activity of metabolite producing enzymes or decreased activity of metabolite consuming enzymes (228). We assessed flux through the TCA cycle with $^{13}C_6$ tracing. M+2 metabolites are generated through the TCA cycle and originate from a fully labelled glucose molecule. All TCA cycle intermediates had elevated M+2 in normal samples compared to tumour. Interestingly the exception was succinate. As Figure 5.10 shows, patients 2 and 4

(Samples 3,4,5,8,9) had elevated label incorporation in succinate in tumours compared to normal. Label incorporation continued to be reduced in tumours compared to normal in malate and aspartate.

In Patients 2 and 4 this pattern could be explained by a loss of SDH function, as elevated succinate is followed by reduced levels of fumarate, malate and aspartate. Unfortunately, the MID differences between them are very small, and not large enough to draw conclusions. Examining the presence of M+3 fumarate can further help in assessing SDH function. In patients 2 + 4, the ratios of M+2/M+3 were lower in tumour, but again not significantly so. With more samples, and a greater tracer uptake these results may have become significant.

Lastly, although RNA expression of *SDHA* appeared lower in tumours, there was no evidence of gene fusions. Although raw abundance levels looked promising, the metabolic flux data is equivocal, and does not fully support a loss of SDH function. Importantly, though, it does not refute it. Furthermore, a consistent pattern of a downregulated *SDHA* both in the transcriptome and metabolome is shown.

5.3.2 Metabolic phenotype of WDLS

The data produced here affords us an examination of the metabolic phenotype of WDLS. Although we cannot say through which pathway these metabolites were made, all tumours have elevated levels of succinate, fumarate and aspartate. These differences are most noticeable in succinate and aspartate, and particularly in sample 9 – a dedifferentiated part of the tumour. This contains nearly ten times (40.03 v 4.33

ions/mg) the level of succinate, which is relevant as this is a more clinically aggressive tumour.

The mechanisms for accumulation can be speculated, but the consequences of high succinate levels are well documented. So far, the genomic data suggests *SDHA* impairment, which in turn leads to succinate accumulation. Rasola et al. showed elevated *TRAP-1* expression in tumours impairs SDH, resulting in increased succinate (230). Others believe in obesity, adipocytes exposed to hypoxia release succinate, which may be relevant in this adipocytic cancer (231).

The resultant elevation in succinate levels is interesting for several reasons. The HIF-1-alpha pathway has already been addressed, but more recent data suggests succinate can also regulate the Jumonji-C-domain-containing histone demethylases and the ten eleven translocations (TET) family, both of which regulate histone and DNA methylation (232). Furthermore, succinate accumulation in tumours has been show to promote angiogenesis via upregulation of *VEGF*, in a HIF-1a independent manner (233). Clearly succinate has the capacity to act as an oncometabolite, through HIF-1a, epigenome remodelling and angiogenesis.

Similar patterns were observed in aspartate, with the exception of patient 2. Again, the dedifferentiated component (sample 9) showed significantly higher levels. Aspartate is a proteinogenic amino acid but is also essential for nucleotide biosynthesis in proliferating cells (234). For these reasons, cellular aspartate can

drive cancer proliferation (235). Why levels in tumours are high is therefore of interest.

Primarily, the elevated aspartate levels suggest that tumours have created a better aspartate supply for themselves, possibly through enhanced TCA cycle activity, which is also observed by the presence of m+2 isotopomers of several TCA cycle metabolites. Aspartate has poor cell permeability and is taken up by very few cells, except hepatocytes (235). Levels are highest in the dedifferentiated tumour, which is highest in grade. It is feasible that these tumours require aspartate to dedifferentiate, which helps drive this more aggressive phenotype. Garcia-Bermudez et al. showed aspartate levels are low in hypoxic conditions, but hypothesised that in the longer-term, chronic hypoxia leads to upregulation of gene expression of aspartate transporters and levels increase (236). Whether these tumours are chronically hypoxic requires separate work, and equally other mechanisms may explain these elevated levels.

Understanding alterations in tumour metabolism can lead to therapeutic targets. In paragangliomas, Tennant et al. described a loss of succinate dehydrogenase activity resulting in reduced metabolic plasticity and a dependence on pyruvate carboxylase to fulfil cellular metabolism (227). Efforts to drug pyruvate carboxylase or its precursors are ongoing. Birsoy et al. are examining the reliance on aspartate by tumour cells, in the hope of targeting these pathways in vulnerable tumour types (236). Any reliance on aspartate, particularly through its membrane transporters,

could be a relevant target in WD/DDLS. In theory this would mainly affect tumour cells and hepatocytes, with close monitoring of liver function easy to perform.

5.3.3 Tracer uptake in tumour versus normal tissues

Although tumours contained elevated raw abundance of these metabolites, this pattern was not reflected in $^{13}C_6$ label incorporation. Except succinate, label incorporation was higher in normal tissues for all glycolytic and TCA metabolites, as well as amino acids. As cellular glucose metabolism is one of the metabolic hallmarks of cancer, this requires further discussion (237).

A potential explanation for the higher label incorporation in normal tissues involves lactate. There is evidence of higher M+3 lactate in normal tissues than blood, which was significant ($p=0.02$). Tumour lactate compared to blood was comparable. Rabinowitz et al. in a landmark paper showed that lactate can be used as a primary source for the TCA cycle in healthy tissues (238). It is possible that normal tissues can take up lactate from the blood and metabolise it. This gives an additional TCA cycle fuel, hence creating a higher proportion of labelled TCA intermediates.

Paired normal tissue was taken from retroperitoneal fat which is a mixture of white and brown adipose tissue (239). Brown adipose tissue is the primary site for thermogenesis and uses a very high amount of glucose (240). When pathway analysis was performed in our dataset, thermogenesis was the most upregulated pathway. Furthermore, glucose uptake in BAT is stimulated by sympathetic stressors,

i.e surgery (240). It is conceivable that paired retroperitoneal fat, due to its functional role is able to import glucose at a higher rate than tumour tissue.

Glucose uptake by tumours has been harnessed for diagnostic purposes in Positron Emission Tomography (PET) scanning. 18-fluorodeoxyglucose (^{18}F-FDG), a radiolabelled glucose analogue, is taken up by proliferating cells and trapped, allowing the PET scanner to indicate areas with increased glucose uptake (241). It is the gold standard for diagnosis in several cancers, but not routinely used in RPLS. A recent study across several sarcoma subtypes using PET scanning, found significant correlation between both glucose metabolism, GLUT-1 expression and grade (242). However, this did not apply to WDLS, of which 0/4 expressed GLUT-1.
Furthermore, a recent study of RPLS recurrences, found a higher tracer uptake in DDLS v WDLS ($p=0.0035$). When applying PET metabolism to WDLS, we may infer a less aggressive, more indolent metabolic phenotype.

Limitations in experimental methodology may have meant isotopic steady state was never reached, and therefore the MID values could be falsely low. TCA metabolites take longer to reach steady state, sometimes up to several hours, whilst the mean time from infusion to tumour removal was just under two hours. One way around this would have been to start the glucose infusion earlier, over a longer period in the hope of steady state being reached by the time the tumour was removed. However, it was not practical to start the infusion earlier in this operative setting due to anaesthetic priorities. Importantly, these limitations would have applied to the normal samples as well limiting any bias.

Lack of label incorporation in the amino acids serine and glycine could indicate a failure to reach steady state in these metabolites. This is unsurprising considering there was only 10% labelling in the blood to start, and these two amino acids take multiple reactions to generate. Furthermore, cells are able to make their own serine and glycine, and would not rely solely on labelled glucose to do so (243).

5.4 Conclusion

The metabolic findings do suggest a degree of impairment of SDH function. This is shown through elevated levels of raw succinate in tumours, and an enhanced presence of M+3 fumarate in patients 1, 2 and 4. Unfortunately, these findings did not reach statistical significance, which may be aided by an increased study size, and ensuring steady state is reached. Critically, none of the tumour samples had *SDHA* fusions, which means we cannot ascribe the metabolic suggestion of SDH dysfunction to gene fusions within *SDHA*. New findings include elevated label incorporation across normal samples, which may come from the use of lactate as a primary carbon source. Elevated aspartate levels also offer a novel insight into this tumour and require further investigation.

Chapter 6: Genome Wide CRISPR Knockout Screen

6.1 Introduction

FOXD4L3 and *SDHA* could easily be passengers in a wider genomic rearrangement and have no direct role in tumorigenesis. A broader approach, using a genome-scale loss-of-function screen allows for the unbiased identification of multiple genes, essential for survival and proliferation.

Barrentina et al. performed knockdown with short hairpin RNAs (shRNAs) on 385 genes in three dedifferentiated liposarcoma cell lines and were able to identify 99 genes whose knockdown significantly decreased cell growth in at least one cell line (112). shRNA screens can, however, be limited by incomplete protein depletion and off-target effects. Genome-scale CRISPR-Cas9 knockout (GeCKO) libraries utilise RNA-guided endonuclease Cas9 from the microbial CRISPR (clustered regularly interspaced short palindromic repeat) immune system (149). This has proved more robust and specific, and GeCKO libraries have been used successfully to explore gene essentiality, drug and toxin resistance and the hypoxia response (244).

The library is delivered through lentivirus, and the cas9 endonuclease is guided to a genomic target by short RNAs. Here it is able to produce double strand breaks that are repaired through either homology-directed repair, or non-homologous end-joining (NHEJ). NHEJ is error prone, producing insertions or deletions, which in a coding

region results in loss-of-function mutations (245). These produce premature stop codons, and either decay of the transcript or a non-functional protein.

GeCKO libraries which target the 5' conserved coding exons of 19,050 human coding genes can be transduced in a cell line at a low multiplicity of infection, to ensure that each cell receives only one guide RNA (149). The cell line can be subjected to either a positive selection, e.g radiotherapy or drug, or negative selection where it is simply grown over a period of time, also called a drop-out screen. Negative selection screens allow the identification of essential genes, where the experimental and control conditions are analysed at two different time points (244). Once complete, next generation sequencing is used to compare the number of sgRNAs in the experimental condition relative to the control, thus identifying essential genes or cell proliferation and survival.

To date, no GeCKO screens in WDLS have been published, and this screen does not appear in the cancer dependency map project (Depmap) (246). A GeCKO screen represents a genuine opportunity to further explore the liposarcoma genome. It will identify essential pathways to all cancer cells, but through comparisons with other datasets and cross-referencing with total and single cell RNA sequencing in Chapter 7, it could identify liposarcoma specific targets.

6.2 Results

Preliminary experiments to calculate the multiplicity of infection (MOI) required for optimal lentiviral transfection were performed. The first of these was a cell viability assay to determine the number of cells per well for continuous growth over 72 hours. This is shown in Figure 6.1, with 5000 cells (red line) the most linear curve, indicating continuous growth.

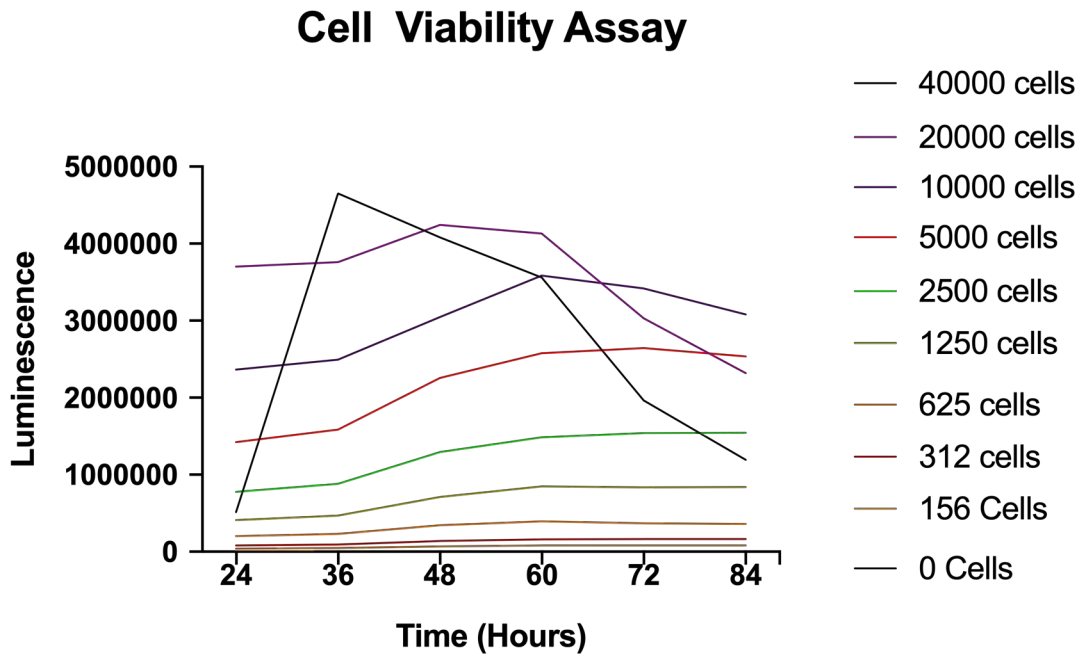

Figure 6.1: Cell viability assay plotting luminescence in different cell concentrations over time. Luminescene was generated by the EnSpire Multimode Plate Reader. 5000 cells (red line) shows the closest to a linear growth pattern over time, plateauing at 72 hours.

The second preliminary experiment aimed to determine the lowest concentration at which the selection antibiotic, puromycin, was toxic to the cells. A kill curve was performed which showed 0.6 µg/ mL as the lowest concentration required to generate no cellular luminescence activity (Figure 6.2).

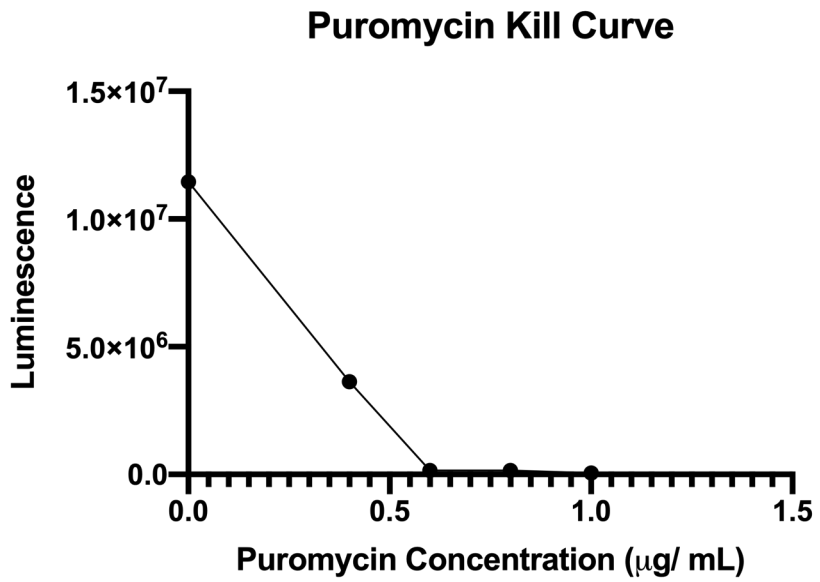

Figure 6.2: Cell titre glow luminescence at increasing puromycin concentrations (µg/ mL). 0.6 µg/ mL was found to be the lowest concentration needed for no cellular activity.

The last preliminary experiment calculated MOI by dividing the average luminescence of 2 wells treated with puromycin by the average luminescence in the two wells without puromycin. This was done at 6 virus condition volumes (Table 6.1).

Table 6.1: Multiplicity of infection at different lentiviral volumes

Virus condition (µL)	0	2.5	5	10	20	40
Antibiotic selection (absorbance)	58085	759885	1925000	3670000	4750000	2835000
No antibiotic selection (absorbance)	8635000	9610000	10050000	10400000	9650000	7280000
Average MOI	0.006727	0.079072	0.191542	0.352885	0.492228	0.389423

Using GraphPad, the lentiviral supernatant volume (log_{10}) was plotted against MOI (Figure 6.3)

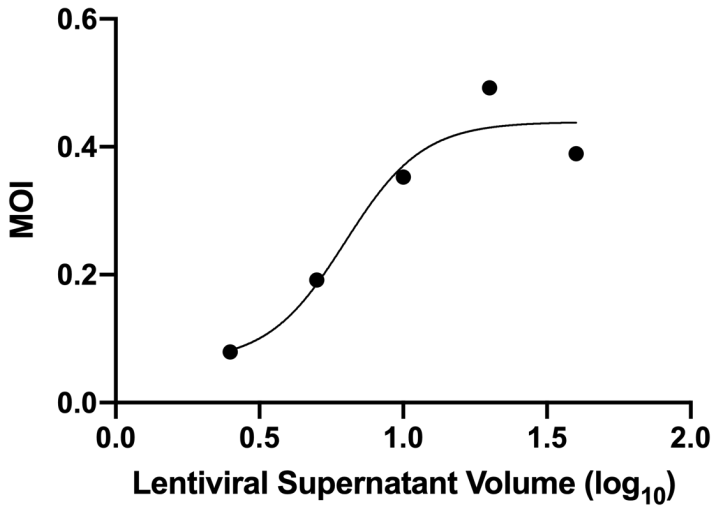

Figure 6.3: Sigmoidal curve generated in Graphpad Prism. Curve plots multiplicity of infection against lentiviral supernatant volume. Bullet points were average of lentiviral replicate concentrations.

This then underwent logarithmic back transformation to give 8.97 µL lentiviral supernatant.

The screen was then performed as per Chapter 2.12. Output DNA was sequenced and outputted in a fastq format. This was then kindly processed by Professor Andrew Beggs using Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout software (MAGeCKFlute, SourceForge, USA), an integrative analysis pipeline for GeCKO screens (247). A summary of the screen is highlighted in Figure 6.4 below.

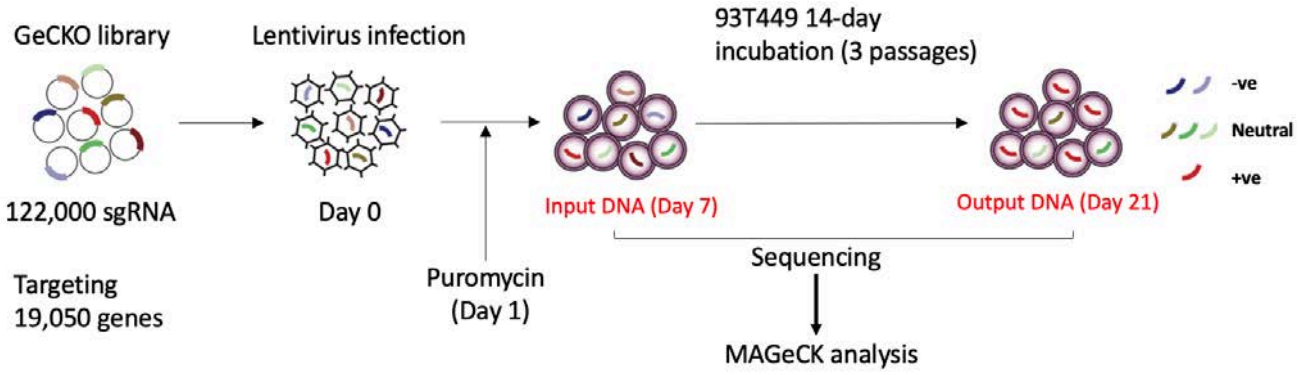

Figure 6.4: Schematic representation of GeCKO screen for WDLS cell line (93T449) adapted from Yamauchi et al. (248). The GeCKO library consists of 122,000 sgRNAs which target 19,050 genes. Lentiviral transfection marks Day 0. Cells are grown in puromycin media for 7 days before half are extracted for sequencing (Day 7). The remaining cells are grown for a further 14 days prior to sequencing. The MaGECKFlute pipeline identifies positively and negatively selected genes. Blue coloured genes drop out, and are negatively selected. Red coloured genes have accumulated, and are positively selected.

The first output from the MaGECKFLute pipeline is a comparison of day 14 versus day 0 log-fold-change (logFC) and false discovery rate (FDR). This generates a beta score which measures the degree of selection after target perturbation, similar to 'log fold change' in differential expression analysis (247). A positive beta score indicates a positive selection, and a negative score indicates a negative selection. A positive beta score indicates that gene knock out may promote cellular proliferation, whilst negative beta scores indicate that gene knock out may inhibit cellular proliferation or cause cell death.

Day 0 here applies to input DNA, and day 14 output DNA. We anticipated genes encoding basal cellular machineries to be enriched in negatively selected genes. Therefore, to identify potentially actionable targets, as per Yamauchi et al. we excluded histones, ribosomal proteins and polymerases as targeting them would affect healthy tissues. This left five genes of interest which are shown in Table 6.2, and then discussed briefly afterwards

Table 6.2: Most significant negatively selected genes in GeCKO screen. Protein function from OMIM database (249)

Gene Name	Protein Function	Beta Score	FDR
FDX1L	Mitochondrial iron-sulphur protein	-7.0994	0.002475
IARS2	Mitochondrial isoleucine-tRNA synthetase	-6.1992	0.04703
PCBP2	Multifunctional adapter protein that binds iron	-5.5982	0.051705
MAPKAP1	Subunit unit of mTOR complex 2	-2.4056	0.071131
BARD1	Forms heterodimer with BRCA1 co-ordinating DNA damage repair	-5.3227	0.071131

FDX1L and *PCBP2* are both involved in iron homeostasis. *FDX1L* encodes the mitochondrial ferredoxin 2 (Fdx2) protein involved in Fe-S cluster biogenesis, the mitochondrial respiratory chain and haem biosynthesis (250). *PCBP2* has been found to deliver iron to ferritin for storage, as well as other intracellular hydroxylases. It is also involved in RNA processing and translation (251). *FDX1L* is highly significant (FDR=0.002) whilst *PCBP2* is borderline (p=0.05), and their negative selection indicates iron may play an essential role in liposarcoma.

IARS2 encodes an aminoacyl-tRNA synthetase (ARS) which was significantly negatively selected (FDR=0.04). RNA modification is essential for both tumour and healthy cells, but their inclusion here is based on wider evidence that ARS as a class

can perform non-canonical functions, and are implicated in cancer development and progression (252).

Mitogen-activated protein kinase 1 (*MAPKAP*) is crucial for the assembly of mammalian target of rapamycin complex 2 (*mTORC2*), which in turn regulates cell growth and is highly sensitive to growth factors (253). *MAPKAP1* has also been shown as a direct binding partner of *PCBP2* in apoptosis pathways, which may explain their negative selection here *(254)*.

The last gene of interest here is BRCA1-associated ring domain 1 (*BARD1),* which has roles in cell cycle progression and cell viability (255). Although this negative selection may also be applicable to healthy cells, as *BARD1* is involved in DNA damage repair and is the subject of therapeutic interest in breast cancer, it deserves further discussion.

A limitation of a direct comparison between day 14 and day 0 is that it does not control for commonly selected genes. To prioritise the identification of interesting genes that play an essential role in liposarcoma, depmap data can be used as a control (246). Depmap is a project that performed hundreds of GeCKO screens in cancer cell lines, and by comparison excludes commonly selected essential genes. Figure 6.5 compares the beta score distribution of Depmap to Day 14.

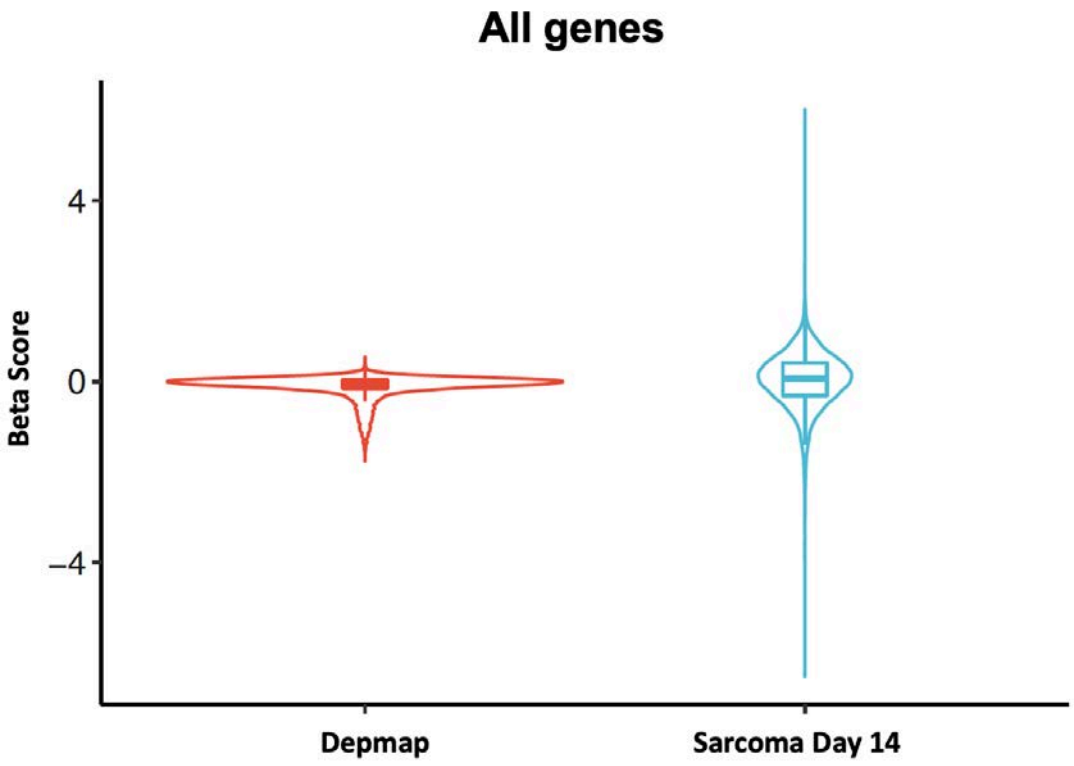

Figure 6.5: Violin plot of all genes distributed by beta score. Red plot represents DepMap which acts as a control. Light blue plot shows our final pool of sequenced cells. The tails represent those genes which differ from Depmap by beta score, and are outliers.

The narrow tails indicate a small number of significantly depleted and enriched genes that will be of interest. There are many ways to explore these genes using MaGECKFlute. Firstly, they are sorted based on differential beta score, which is calculated by subtracting the control beta score (from DepMap) from the treatment score (Day 14) (Figure 6.6).

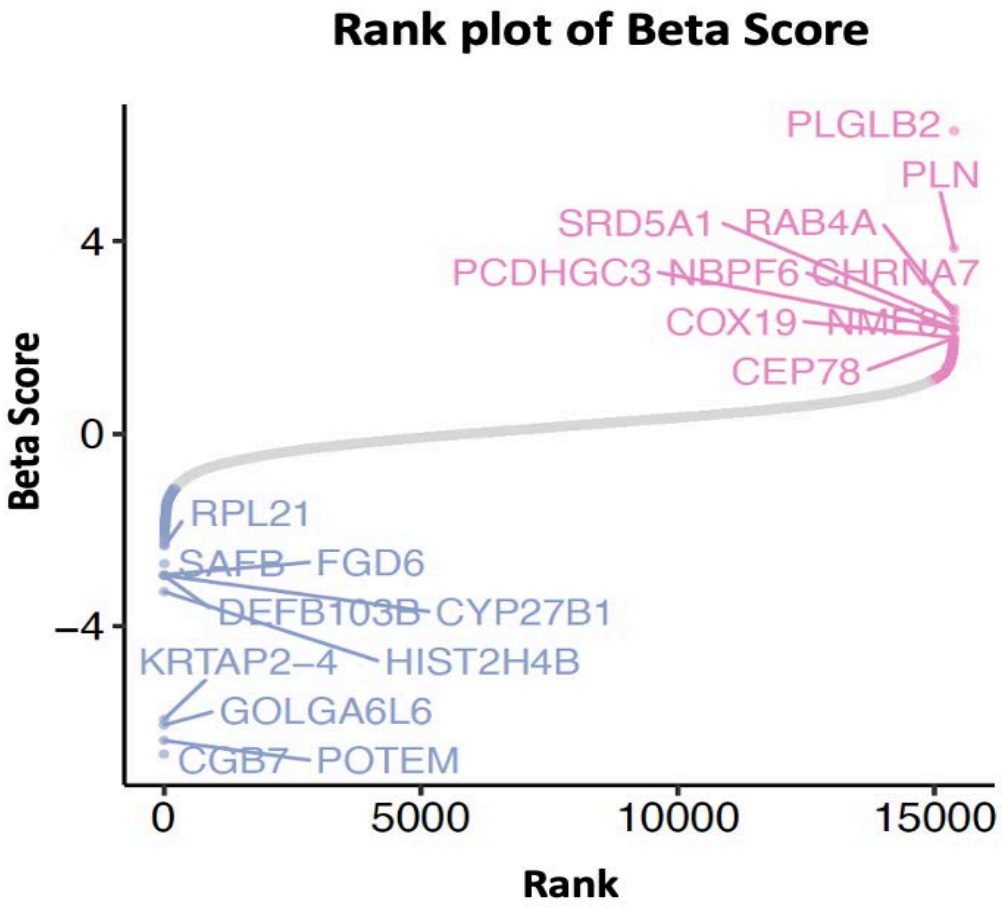

Figure 6.6: A rank plot of most selected genes by beta score. 10 most negatively selected (blue) and positively selected (red) genes highlighted. This is calculated by subtracting the control beta score (DepMap) from the treatment score (Day 14). Grey connecting line represents genes neither positively or negatively selected.

The lowest ranked genes (not confined to those in Figure 6.6) were then analysed using the Comprehensive Resource of Mammalian Protein Complexes (Complex) database (Figure 6.7) (256).

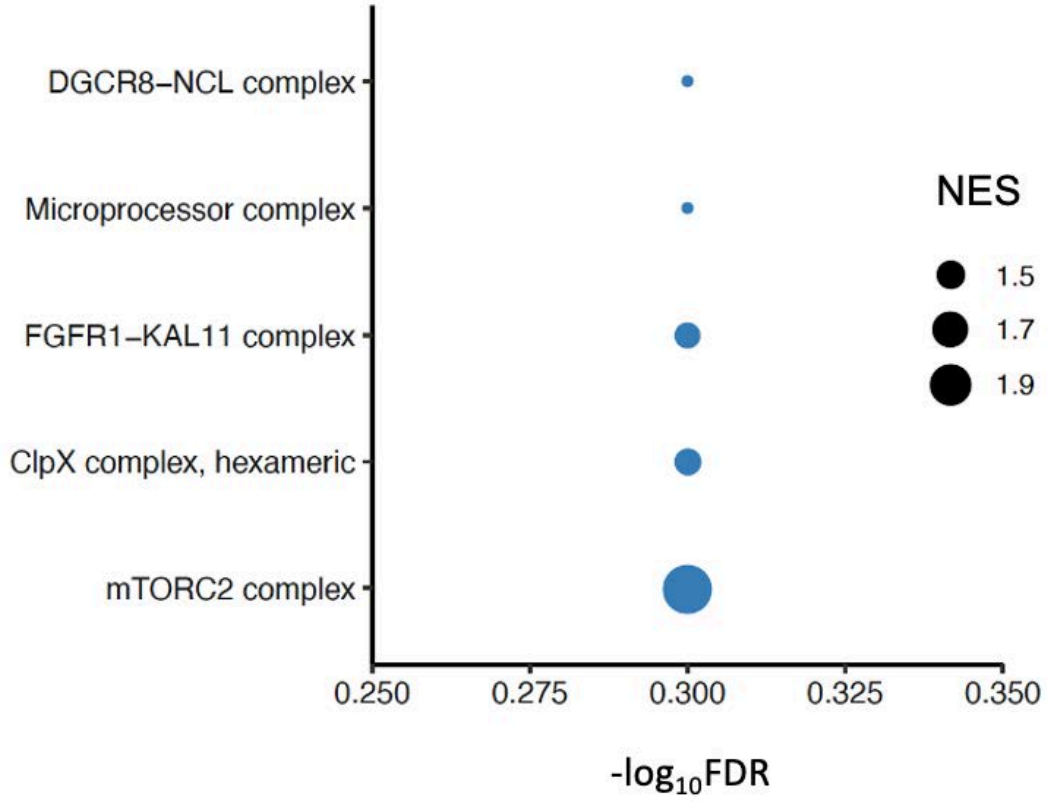

Figure 6.7: Complex enrichment of negatively selected genes. Complex identifies groups of genes associated with protein complexes. The size of each circle is proportionate to the numbers of genes enriched in the corresponding complex, and is given through a normalised enrichment score (NES). The mTORC2 complex is identified as negatively enriched.

The mTORC2 complex has been identified as negatively enriched using the protein complex database. It has a normalised enrichment score of 1.9, with a -logFDR of 0.3. This gives an FDR of 0.52, which is not significant. It is however, the second time this complex has been identified, with *MAPKAP1* discussed earlier.

The last output of MaGECKFlute is a scatter graph which classifies genes into four groups determined by differences in the beta scores between Day 14 and Depmap controls (Figure 6.8). By doing so those negatively selected by Depmap are filtered out.

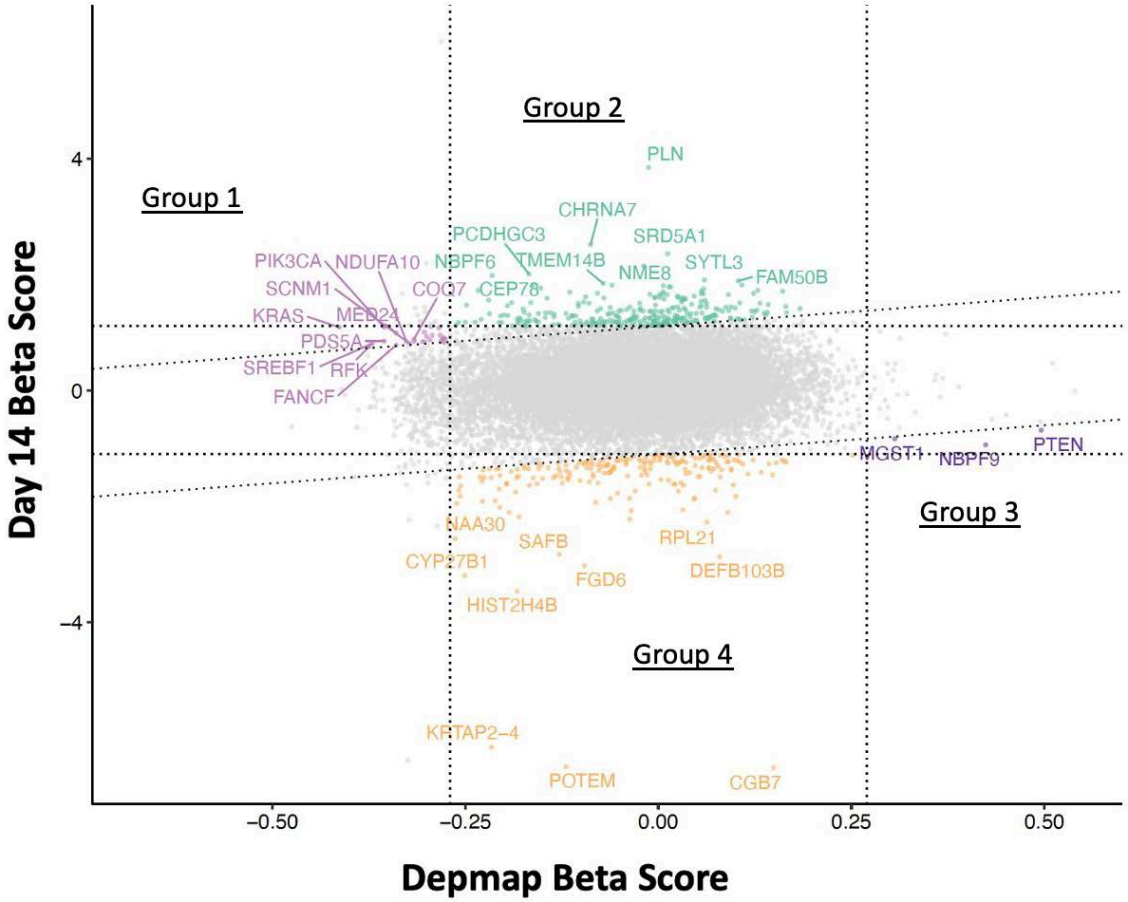

Figure 6.8: Scatter plot dividing genes in four different groups based on the largest absolute value of the differential beta score. Group 1 (Lilac) are strongly negatively selected in the control samples and are weakly positively selected at Day 14. Group 2 (Green) are strongly positively selected at Day 14 and weakly positively or negatively selected in the control samples. Group 3 (Dark purple) are strongly positively selected in the control and weakly negatively selected at Day 14. Group 4 (Orange) are strongly negatively selected at Day 14, and weakly positively or negatively selected in the control samples. The diagonal dashed line indicates the mean +/- 1 standard deviation of the differential beta score. The horizontal and vertical dashed lines indicate the mean plus or minus one stand deviation of treatment and control beta score, respectively(247).

The most relevant subgroups here are 2 and 4. Group 2 could contain tumour suppressors as gene knockout causes an enrichment in those cells. Group 4 contains genes that are essential for cell survival, as these are strongly negatively selected. A list of these genes are included in Appendix II.

10 genes in group 2 were entered into the Tumour Suppressor Gene Database which identified Protocadherin gamma-C3 (*PCDHGC3)* as a recognised tumour suppressor (257). It has been shown to supress Wnt and mTOR signalling pathways (258).

Group 4 was then examined, firstly by excluding ribosomal proteins, histones and polymerases as before. This left two genes of interest – Cytochrome P450 Family 27 Subfamily B Member 1 (*CYP27B1)* and N-alpha-acetyltransferase 30 (*NAA30).*

CYP27B1 is better known as 1-alpha-hydroxylase which converts calcifediol to calcitriol, the active form of vitamin D (259). When using MaGECKFlute to compare Day 14 to Day 0, *CYP27B1* had a beta score of -4.37 with an FDR of 0.16. Figure 6.5 shows a Depmap beta score of -0.25. Based on these findings, vitamin D metabolism could be another pathway of interest in liposarcoma. *NAA30* forms part of a catalytic subunit of NAT-C, which is involved in post-translational modification of proteins through N-terminal acetylation. Dysregulation of these modifications is a hallmark of many cancers (260).

For further exploration of these two groups, complex enrichment is shown in Figure 6.9.

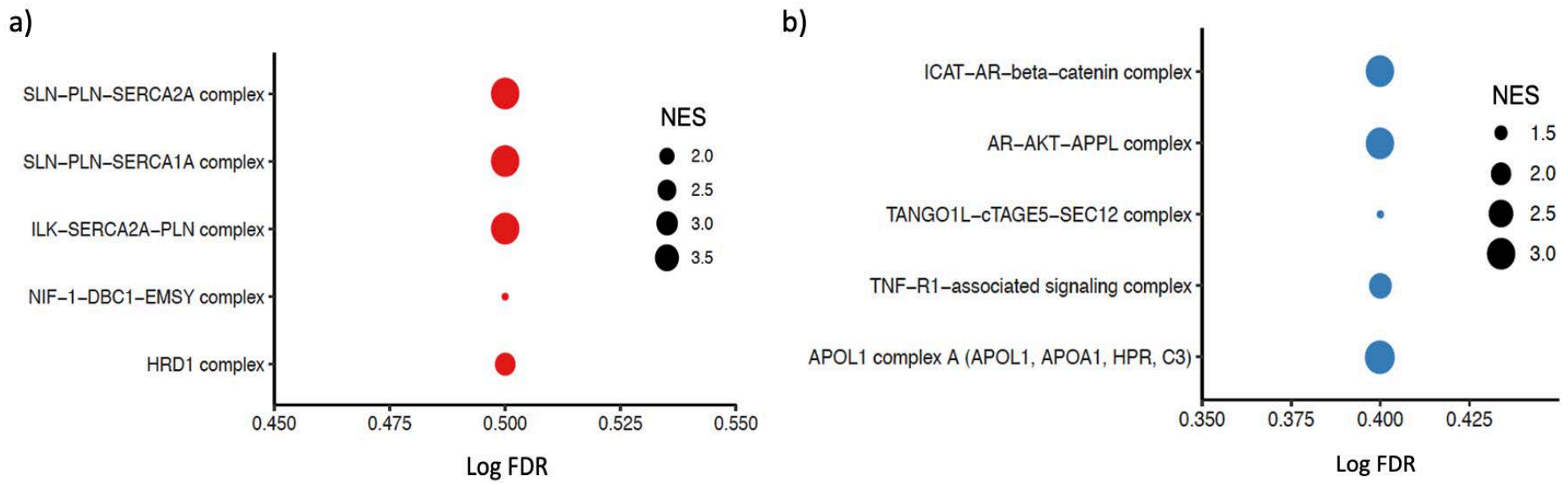

Figure 6.9: Complex enrichment of positively selected genes in Group 2 (a) and negatively selected genes in Group 4 (b). Complex identifies groups of genes associated with protein complexes. The size of each circle is proportionate to the numbers of genes enriched in the corresponding complex, and is given through a normalised enrichment score (NES).

Three of the most positively selected genes in Group 2 involve the Sarco/endoplasmic reticulum (SR/ER) Ca^{2+} ATPase pump (*SERCA*). Proteins of the BCL-2 family and p53 exert pro-apoptotic actions through *SERCA* (261). By knocking out *SERCA*, these pro-apoptotic actions are lost, and hence these cells are positively selected.

Figure 6.9(b) highlights the ICAT-AR-beta-catenin complex as being negatively selected. Wnt/ß-catenin signalling is implicated in many cancers, resulting in the increased transcription of many Wnt target genes (262). If this is negatively selected, combined with the positive selection of *PCDHGC3* (Figure 6.5) it is possible that Wnt signalling is used by liposarcoma cells to drive cell proliferation. Correlation with transcriptomic data will be required to see if there is more evidence for this.

6.3 Discussion

The data has identified several pathways and genes of interest. Two genes involved in iron homeostasis were present as significantly negatively selected (*FDX1L* and *PCPB2)*, whereas mTOR and Wnt signalling were implicated throughout. There is a fine balance between whether these genes and pathways are directly relevant to liposarcoma or simply applicable to all proliferating cells. This is especially important when considering any therapeutic implications, and therefore the relative merits of each requires further discussion in turn.

6.3.1 Iron homeostasis

FDX1L was the most highly negatively selected gene in our dataset when comparing Day 14 to Day 0. *FDX1L* encodes Fdx2, an iron-sulphur complex which acts as an electron transfer agent in a wide variety of redox reactions (263). Sheftel et al. performed Fdx2 knock down which led to increased cellular iron uptake (264).

PCPB2 is a documented iron chaperone, involved in iron storage. Human liver cells lacking *PCPB2* load less iron into ferritin (265).

Clearly both genes play a direct role in iron metabolism. Cancer cells are known to have higher metabolic iron demands as it is critical to proliferation and growth, and rely on a limited supply of free iron (266). Both knockdowns likely dysregulate this, but in potentially different ways. Fdx2 when deficient in healthy cells increases iron uptake, whereas in cancer it could be speculated iron pools are lower and iron uptake cannot increase, limiting cell growth (264). *PCBP2* knockout could inhibit growth by depletion or iron stores.

If reliance on iron is present, reducing availability of iron should be considered. New synthetic iron chelators such as Triapine have had some success in Phase 1 studies when combined with standard chemotherapy, reducing tumour burden with limited side-effects (267). More recently, higher throughput screening identified an iron chelating agent as a promising drug candidate for colon cancer (VLX600) (268).

It is important to note however, that neither *FDX1L* nor *PCPB2* appeared on the scatter graph due to filtering out performed by Depmap control. This is evidence that these genes are found to be essential in previous Depmap screens, i.e other cancers. Furthermore, Fdx2 is present in all tissues and is essential for Fe-S groups on helicases and DNA polymerases and directly targeting it may harm healthy tissues.

In summary this data suggests liposarcomas fall into a group of a number of cancers where iron homeostasis is involved in cell survival. In vitro cell proliferation studies growing 9T449 in iron-chelating agents would be straightforward and of interest.

6.3.2 mTOR

There are several links to the mTOR pathway presented in this data - both direct and indirect – which is relevant as abnormalities in mTOR signalling are associated with numerous cancers. *MAPKAP1* (also called Sin1) was negatively selected in this dataset, and forms part of mTORC2 along with *RICTOR* and *MLST8* (269). mTORC2 was selected in our complex analysis (Figure 6.7) and plays two main roles. Firstly, it regulates the actin cytoskeleton. Secondly, it phosphorylates the protein kinase Akt (270). Recently *MAPKAP1* has been identified as the most important component of mTORC2 required for Akt phosphorylation, which in turn leads to cellular proliferation and survival (253). mTORC2's role in metabolism was investigated by Li et al. by *MAPKAP1* knockout in 8 cancer cell lines. They found SREB1 levels declined, which indirectly supressed lipogenesis (271). *MAPKAP1* knock out in 93T449 could supress cell growth either through lack of Akt phosphorylation, adipocytic inhibition or both.

As observed with the previous genes, *MAPKAP1* did not appear in the scatter graph. This highlights that it was likely negatively selected in the Depmap screens, which is not surprising as mTORC2 is implicated in other cancers. The previously discussed *NAA30* did appear, which as part of the *NAT12/NAA30* complex has been proposed as a substrate of mTOR giving more evidence for this pathway in WDLS (272).

If we propose that liposarcomas are driven by Akt phosphorylation, targeting the MAPKAP1-mTORC2 interaction may yield clinical outcomes. Until recently, attempts to drug mTORC2 had off target effects, also inhibiting mTORC1. Benavides et al. used a small molecule inhibitor CID613064 to inhibit the mTORC2 phosphorylation targets AKT, NDRG1 and PKCα without effecting mTORC1 (273). There is some evidence in prostate cancer that mTORC2 can be targeted in tumour tissues to good effect, with no deleterious effects in healthy prostate cells. WDLS may benefit from mTOR signalling inhibition, and commercially available combined mTOR inhibitors would be a good place to start drugging cell lines in vitro, before more specific mTORC2 agents become available.

6.3.3 Wnt Signalling

There is some evidence from this data to suggest that Wnt signalling is utilised by liposarcoma cells. Complex analysis derived from Group 4 (Figure 6.9) shows the ICAT-AR-B-catenin complex is negatively selected. This complex is downstream of Wnt activation and part of the canonical pathway. When activated it causes an accumulation of ß-catenin in the cytoplasm where it translocates to the nucleus to act as a transcriptional activator (274). The complex analysis asserts that when this complex is prevented from functioning, there is resulting cell death. To further support this, *PCDHGC3* was positively selected. Dallosso et al. showed *PCDHG3* is pro-apoptotic in colorectal adenomas, and is able to repress Wnt/ß -catenin transcriptional activity (258). Moon and colleagues noted that Wnt signalling can promote or inhibit tumour growth in a cancer stage and type specific manner. The

combination of negatively selected pathway genes, and a positively selected pathway inhibitor suggest that Wnt signalling promotes growth, rather than inhibiting here.

In a similar manner to mTOR signalling, it is prudent to ask whether Wnt signalling can also be targeted. Wnt signalling is highly complex, and the ubiquitous nature of Wnt signalling leads to a cautious approach over the use of pathway agonists or antagonists (275).

There are several Wnt/ ß -catenin signalling inhibitors currently in clinical trials which manage to provide a therapeutic window against the tumour, whilst sparing the normal Wnt-dependent tissues such as the GI tract (276). PORCN (porcupine) inhibitors are the best example of these, which prevent Wnt ligands being released to the cell surface (277). There are many validatory steps to be taken before classing WDLS as a Wnt addicted cancer, but this pathway is certainly worthy of further attention.

6.3.4 Vitamin D

The biologically active form of Vitamin D is synthesised from an inactive prohormone by *CYP27B1* (278). This was strongly negatively selected in our dataset. It is surprising as Vitamin D has reportedly numerous anticancer properties; antiproliferation, proapoptotic and immunomodulatory and has been negatively correlated with several cancers including lung, prostate and colon (279). However, Clinckspoor et al. demonstrated increased expression of vitamin D receptor in papillary thyroid tissue versus normal tissue, whilst Urbschat et al. showed *CYP27B1*

was upregulated in renal cancer versus normal tissue (p=<0.001) (280, 281). An alternative explanation in this context is copy number variation. Non-essential genes that are present in an amplified region, often appear more negative than expected (282). *CYP27B1* is present on 12q14.1 which is often amplified in WDLS and may explain this inclusion here (283).

6.3.5 BARD1

BARD1 is the last gene of interest identified by this GeCKO screen. BARD1-BRCA1 complex localises to sites of DNA damage and is directly involved in repair. In mice *BARD1* knockout leads to defects in cell cycle progression and therefore could simply represent an essential gene for all cells (255). However, it is possible that liposarcoma cells are heavily reliant on BARD-1 for DNA repair, due to a lack of other DNA repair possibilities, and when knocked out results in loss of cell viability. This may be driven by the error-prone non-homologous end joining (NHEJ) route rather than the more efficient BRCA1 homology directed repair.

Pro-tumourigenic effects of *BARD1* have also been demonstrated through one of its two splice variants - BARD1ß. Excess BARD1ß has a pro-proliferative function overriding cell cycle check points resulting in a more invasive phenotype (284). This could explain its negative selection but would require further work demonstrating increased RNA and protein expression.

6.3.6 Limitations

GeCKO screens such as this are still in their infancy and require continuous refinement. Recent experience has led to certain groups encouraging higher levels of coverage for drop out screens such as this, in an attempt to increase the 'signal to noise' ratio (285). Kampfmann et al. suggest 500-1000x coverage, but yield that this may be prohibited by various factors (286). This screen had 250x coverage, which meant an increase in 'noise' and reduction in the number of significant sgRNA's identified. However, due to the size and fibroblastic nature of the 93T449 cells, a confluent T150 flask only contained 5 million cells. Preliminary calculations meant doubling the coverage to 500x would mean 41 x T150 flasks to start, with 82 required at passage 1. This would not have been economically viable or pragmatic, and therefore we have to accept the limitations of the screen.

The second major limitation is copy number variation. As stated with *CYP27B1,* genes contained from an amplified area of the genome, will be favoured in negative selection. Programmes do exist to control for this in GeCKO screens, but they are not validated, and currently not user-friendly. It is well documented that 12q.13-15 is amplified in WDLS. Importantly, the other genes and pathways discussed here, except for *CYP27B1,* do not lie in this amplified area.

6.4 Conclusion

This GeCKO screen has generated several new avenues of interest in WDLS. Reliance on iron pools may be a limiting factor in tumour growth, and simple chelation studies can answer this fairly reliably. mTOR and Wnt offer targetable signalling pathways, with which successes in other cancers could be piggybacked. Importantly, these pathways are corroborated both negatively and positively selected genes, at more than one point in each pathway. However, these results in isolation cannot implicate these pathways with certainty, and correlation with transcriptomic work will be required for further validation.

Chapter 7: Transcriptomic landscape of WDLS

7.1 Introduction

The GeCKO screen performed in Chapter 6 is a powerful tool in identifying essential genes. It is however limited by copy number variation and a lack of coverage. The advent of next generation sequencing has created orthogonal modalities to both validate pathways implicated by the GeCKO screen, and also generate new targets. Whilst the GeCKO screen assessed expression at a DNA level, transcriptomics allows the quantification of expression levels of RNA transcripts.

Transcriptomics has evolved from microarrays, through to total RNA sequencing and now single cell RNA sequencing. In this chapter we use total RNA sequencing to compare expression profiles in tumour versus normal, of patients included in the SHARPS study. Examining the transcriptome as such allows us to see which genes are active in WDLS.

A limitation of total RNA sequencing is tumour tissue is not entirely homogenous, and other cells from the microenvironment such as fibroblasts, red blood cells and immune infiltrates can contribute. When differential gene expression is performed these cells are sequenced which clouds the results. Single cell RNA sequencing (scRNA-seq) attempts to avoid this. Single cells are isolated from the tissue of interest within gel beads and then lysed. The resulting mRNA transcripts are given

unique barcodes, which then undergo reverse transcription to cDNA. This is amplified by PCR and then sequenced (287). At the time of writing, this had not been carried out in WDLS.

The final result is threefold. Firstly, we can more accurately assess expression data unique to tumour cells. Secondly, we can define any intra-tumoural heterogeneity. For example, the presence of an immune infiltrate in WDLS is not well characterised and eliciting this would deepen our understanding of the tumour microenvironment. Lastly, we can compare expression patterns with pathways identified in the GeCKO screen.

7.2 Results

7.2.1 Identifying cell populations from scRNA sequencing data

A well differentiated component of a dedifferentiated tumour underwent scRNA-seq. The final cell count after tumour dissociation was 6.33×10^5 with a target concentration of 7×10^5. The original alive cell count was 25% which after dead cell removal climbed to 32%. These figures were below the target yields, but enough of a cDNA library (198.4 ng) was generated for sequencing. The lower yields were due in part to the tumour sample being kept refrigerated overnight, which was unavoidable due to late running surgery.

scRNA-seq enables the annotation of cell-types based on transcription profiles. The Loupe Browser 4.1.0 (10xGenomics, USA) uses graph-based clustering as a form of

unsupervised clustering to identify these different cell populations. Clusters of cells (represented by nodes) are generated by the software, which are represented with a T-distributed stochastic neighbour embedding (t-SNE) plot.

t-SNE plots use log transformed normalised expression values when showing individual gene expression. This removes biases in sequencing coverage between cells, while allowing comparisons to be made on relative rather than absolute differences in expression (288). For differential expression analysis the software produces log fold changes, as a ratio of normalised mean gene counts in each cluster relative to all other clusters for comparison (289).

Figure 7.1 identifies three clusters. Clusters one and two both contain 84 cells, whilst cluster 3 contains 48 (Total = 216). Tables 7.1-7.3 detail each clusters most significant genes, with gene function data accessed via OMIM (249).

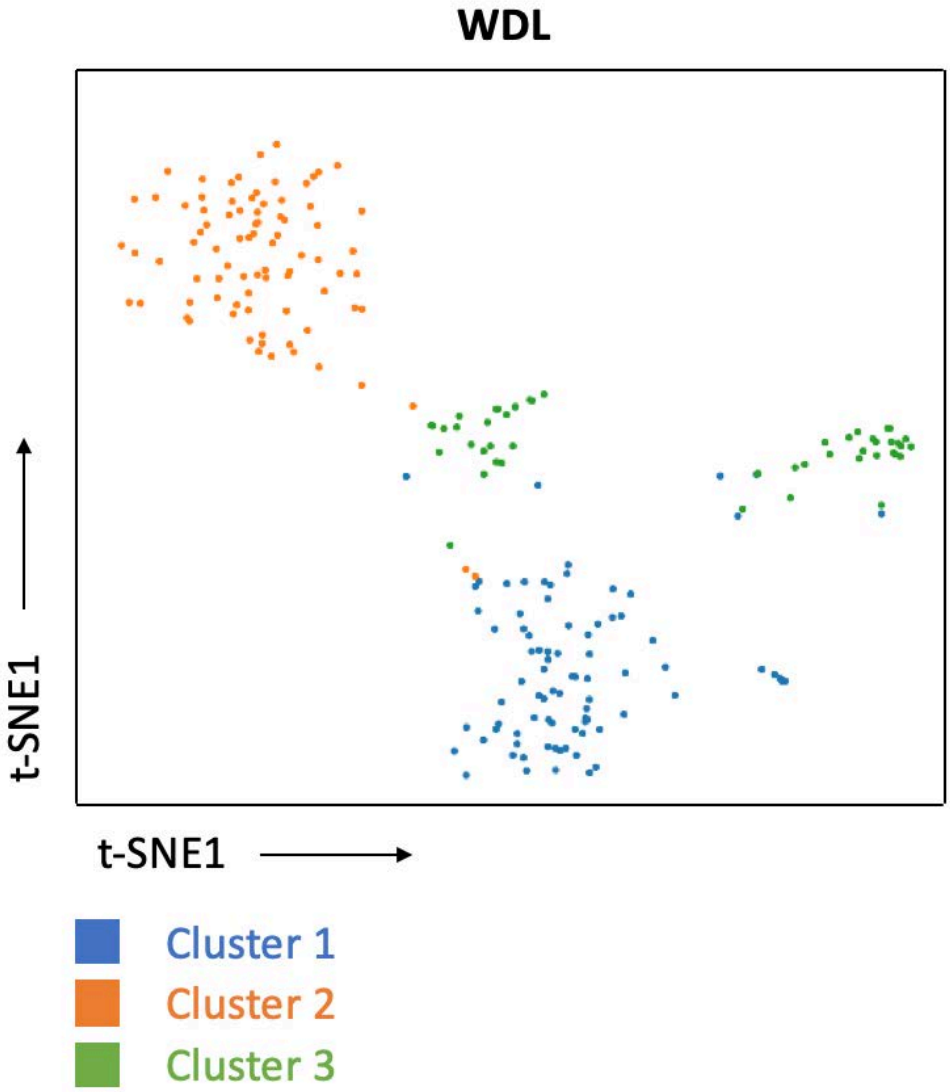

Figure 7.1: t-SNE plot of WDLS sample. Three clusters are identified with a total of 216 cells. Clusters one (blue) and two (orange) contain 84 cells each, whilst cluster 3 (green) contains 48.

7.2.2 Cluster One

Significant genes

A short discussion of each of the most highly expressed genes is given first, for further insight into the nature of the cluster.

Table 7.1: Most significant genes in Cluster One selected by log fold change when compared to Clusters 2 and 3.

Cluster	Gene Name	Gene function	LogFC	Cluster p-value
1	*SRGN*	T-cell secretory granule organisation	6.7	4.72E-51
1	*GPR183*	Adaptive immune response	9.492	7.28E-45
1	*PTPRC*	Activation of *MAPK* activity/ natural killer-cell differentiation	7.062	1.70E-40
1	*IL7R*	Encodes IL-7 receptor expressed on naive and memory T-cells	9.053	6.14E-40
1	*CXCR4*	Chemokine receptor involved in leukocyte chemotaxis	6.486	5.59E-38
1	*TGFB1*	Multifunctional cytokine with immunosuppressive function	5.048	1.16E-31
1	*HLA-B*	Encodes MHC 1 proteins	3.634	1.27E-30
1	*CYTIP*	Expressed weakly in natural killer and T-cells	7.2	3.07E-30
1	*HLA-A*	Encodes MHC 1 proteins	3.888	3.50E-30

7.2.2.1 SRGN

SRGN encodes serglycin, a haematopoetic proteoglycan expressed across the immune spectrum. It's function appears to differ between subtypes. In CD8+ T-cells it is directly involved in the storage of proteases, e.g Granzyme B (see Figure 7.2), whilst in macrophages it promotes the secretion of TNF-alpha (290). Interestingly, there is a growing body of evidence to suggest is upregulated in cancers, and can drive an aggressive phenotype, whilst conferring resistance to the complement system and drugs (291).

7.2.2.2 GPR183

GPR183 is a G-protein coupled receptor, present on the surface of B-cells, T-cells and monocytes. It helps determine the position of B-cells in lymph nodes and regulate T- cell migration through its ligand oxysterol (292).

7.2.2.3 IL7R

IL7R encodes the interleukin-7 receptor. The binding of IL7 to its receptor activates multiple signal cascades including the JAK-STAT and PI3-K pathways (293). The majority of IL7R expression in immune cells is on naïve CD8 and memory T cells. It's presence in cluster one may point towards a more immature T-cell infiltrate.

7.2.2.4 CXCR4

C-X-C chemokine receptor 4 is a G-Protein coupled transmembrane receptor, expressed on the cell surface of most leukocytes, but also adult stem cells such as fibrocytes (294). Data from the human protein atlas suggests these to be most highly expressed in the adaptive immune system (295). CXCR4 has been described as the most widely expressed chemokine receptor in cancer and is associated with invasion, angiogenesis and proliferation (296). Pooled analysis from 85 studies with 11,032 subjects (168 sarcoma patients) found overexpression of CXCR4 to be significantly associated with poorer PFS (HR 2.04; 95% CI, 1.72-2.42) and OS (HR=1.94; 95% CI, 1.71-2.20) (297). Subgroup analysis was not significant for sarcoma (HR=5.14; 95% CI 0.64-41.50), likely due to small sample size given the wide confidence interval.

7.2.2.5 HLA-A and HLA-B

HLA-A and B appear on all nucleated cells in the body, including platelets but not red blood cells. Their increased significance here is likely due to the absence of expression in cluster two.

7.2.2.6 Fine Clustering

Table 7.1 gives us a list of genes expressed across immune cells but does not contain any specific markers of cell type. Further interrogation of the Loupe Browser identified potential subpopulations within this data. Due to the limited number of

cells, it is hard to diversify subpopulations, as this requires a larger dataset. A broader approach is taken, to identify the main types of cells present.

Figure 7.2 highlights *CD2* and *CCR7* - the dominant immune cell markers. *CD2* (log fold change 7.99, p-Value <0.001) is a recognised pan T-Cell marker and present on most mature T-cells (298). *CCR7* (log fold change 7.58, p-Value <0.001) is expressed on the vast majority of T-cells, but is greatly enriched in CD4+ cells (299). There are many different subtypes of CD4 cells, but with limited cells confidently identifying these is not possible. FOXP3, which is indicative of T-regulatory cells (T-reg) do not appear to be present in large numbers here, however other markers of T-regs such as *CD69, IL2RA, CTLA4* and *IKZF2* are (300). More importantly, Transforming growth factor beta 1 (*TGF-ß1)* is highly overexpressed in this cluster which has immunosuppressive effects and is released by T-regs.

GZMB encodes granzyme B released by CD8+ cytotoxic T-cells, which shows some increased expression here in, but in only a limited number of cells. There does appear to be a small number of B-cells, as identified by increased expression of *MS4A1*.

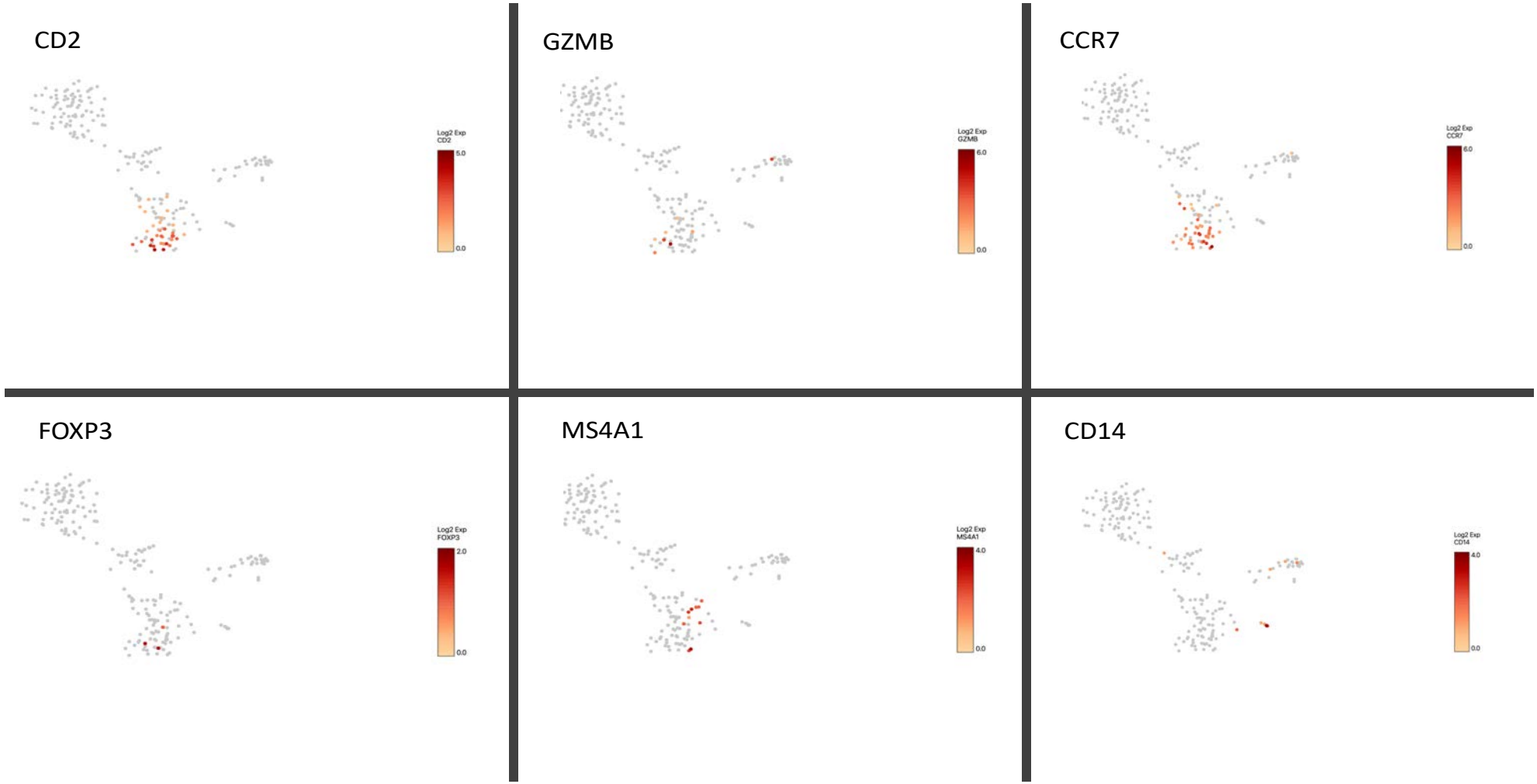

Figure 7.2: t-SNE plots by immune cell marker, across all clusters. Colour bar represents log normalised expression. CD2 – Pan T-Cell marker. GZMB – CD8 Cytotoxic T-Cell. CCR7 – CD4 T-Cell. FOXP3 – T-helper cell. MS4A1 – B-cell. CD14 – Monocytes. Cluster one has highest immune marker activity.

7.2.3 Cluster Two

Table 7.2: Most highly significant genes in Cluster 2 selected by log fold change when compared to Clusters 1 and 3.

Cluster	Gene Name	Gene function	LogFC	Cluster p-Value
2	*HBA2*	Subunit 2 of the alpha globin chain of haemoglobin	11.904476	3.11E-233
2	*HBB*	Encodes the beta globin chain of haemoglobin	11.4742984	2.69E-232
2	*HBA1*	Subunit 1 of the alpha globin chain of haemoglobin	11.9509163	2.04E-196
2	*SLC25A37*	Mitoferrin 1 - mediates iron uptake in developing erythroid cells	5.18550316	1.45E-37
2	*SLC25A39*	Mitochondrial carrier protein	3.45266545	2.46E-13
2	*GYPC*	Maintains erythrocyte shape	2.52014904	7.94E-09

The three most significant genes expressed in cluster two were *HBA1*, *HBA2 and HBB ($p<0.001$).* These encode subunits 1 and 2 of the alpha globin chain of haemoglobin, and the beta globin chain of haemoglobin respectively (301). Other highly significant genes were *SLC25A37*, *SLC25A39* and *GYPC.* The SLC genes are members of the solute carrier group, and play a role in mitochondrial iron transport in developing erythroid cells (302), whilst GYPC is an integral membrane glycoprotein regulating the stability of red blood cells (303).

7.2.4 Cluster Three

Within cluster three are a population of cells that neither appear to be immune or haematological in origin. This cluster has the potential to be malignant or stromal in nature, or a combination. Firstly, to interrogate this cluster, the most highly expressed genes are discussed first. This is followed by the expression pattern of genes known to be overexpressed in liposarcoma, and genes involved in adipogenesis.

Table 7.3: Most highly significant genes in Cluster 3 selected by log fold change when compared to Clusters 1 and 2. Genes denoted with * were also significant in the GeCKO analysis (see 6.5*). Genes denoted ** were included for a discussion point as critical genes implicated in and cancers and liposarcoma respectively.

Cluster	Gene Name	Gene function	LogFC	Cluster p-value
3	*FBLN1*	Glycoprotein incorporated into extracellular matrix and blood vessel walls.	9.12324495	4.00E-32
3	*FGF7**	Potent growth factor secreted by mesenchymal cells with broad mitogenic and cell survival activity	8.74657389	1.20E-21
3	*PLA2G2A*	Phospholipase A2 – enzyme that cleaves fatty acid.	8.70615962	8.96E-12
3	*C3*	Complement component 3 - part of innate immunity, activates both complement pathways	8.60867178	1.61E-16
3	*DCN*	Cellular/pericellular matrix proteoglycan + connective tissue component - binds to type 1 collagen	8.58657801	1.37E-36
3	*SLIT3**	Widely expressed in mammalian tissues. Responsible for kidney and diaphragm genesis.	6.38968596	8.96E-14
3	*THY1**	CD-90 - Cell/cell-matrix interactions. Marker for mesenchymal stem cells.	7.32887705	1.46E-19
3	*MYC***	Encodes nuclear phosphoprotein involved in cell cycle and apoptosis, recognised oncogene	3.00056742	4.29E-15
3	*JUN***	Encodes c-JUN which is required for cell cycle progression/ dimer in AT-1	1.6508642	4.22E-05

7.2.4.1 FBLN1

Fibulin 1 (*FBLN1*) is a member of a family of glycoproteins involved in the organisation of the extracellular matrix (304). Elevated expression levels are described in breast and ovarian cancers (305, 306). Pupa et al. suggested a link between high levels of *FBLN1* and chemoresistance. Through small interfering RNA-mediated suppression of *FBLN1*, they demonstrated a tenfold sensitivity to doxorubicin. This may have relevance, as liposarcomas are highly chemoresistant.

7.2.4.2 FGF7

Fibroblast growth factor 7 (FGF7) binds to Fibroblast growth factor 2 (FGFR-2), which regulates numerous cellular functions and processes (307).

Huang et al. investigated the role of the *FGF7/FGFR2* axis in gastric cancer, and it's interaction with thromobospandin 1 (*THBS1). THBS1* is a multifunctional extracellular matrix (ECM) protein implicated in tumour progression which was significantly overexpressed (log fold change 2.86, $p<0.001$) in cluster three alongside *FGF7* (308).

In search of targets, they pre-treated cells with FGF7, and then gave PI3K, MEK, MAPK and JNK inhibitors. Only the PI3K inhibitor supressed THBS1 activity, and a follow up with an mTOR inhibitor was performed which confirmed inhibition of THBS1. They concluded that FGF7 upregulated THBS1 through the PI3K/Akt/mTOR pathway in vitro and represented a therapeutic target.

7.2.4.3 THY-1

THY-1 or *CD90* is a cell surface glycoprotein involved in cell-cell and cell-matrix interactions (309). It is highly expressed in nervous and olfactory tissues as well as skin. More recently Kumar et al. proposed *THY-1* as a stem cell marker in mesenchymal stem cells (310). In certain cancers, *THY-1* seems to promote cancer growth, whilst in others it acts as a tumour suppressor.

The mechanisms underpinning CD-90 driven tumourigenesis or suppression are poorly understood and rarely described. Koumas et al. described activation of the MAPK pathway as one mechanism, whilst Chen et al. in HCC showed ectopic CD90 expression driving mTOR phosphorylation (311, 312). The beta score for THY-1 from Chapter 6's GeCKO experiment was -2.38. This negative selection indicates that THY-1 is important for cell growth and development and less likely to act as a tumour suppressor.

7.2.4.4 SLIT3

SLIT3 is a member of the Slit family of extracellular matrix proteins. It is discussed briefly here as it was significantly expressed in cluster 3 and also positively selected in the GeCKO screen (Beta score = 2.1). Very little is known about the mechanism of action of SLIT3. Fagerberg et al. performed analysis of human tissue expression across 27 tissue types; the highest mean RPKM for SLIT3 was in adipose tissue (22.998+/- 0.976) (313). Svensson et al. also demonstrated mRNA expression of SLIT3 in all adipose tissues, whilst Paul et al. observed it's secretion by 'pericytic' mesenchymal stem cell (MSC), which promoted angiogenesis (314, 315).

It is unclear why SLIT3 could be overexpressed in cluster 3, whilst being positively selected in the GeCKO screen. The latter implies that preventing the gene functioning, allows the cells to proliferate more rapidly. Of note, is the association with adipose tissues and mesenchymal stem cells, giving further insight into tumour origins and microenvironment.

7.2.4.5 MYC

A gene that cannot be overlooked in the context of malignancy is MYC. In this dataset it was significantly expressed in cluster 3 (log fold change 3.0, $p<0.001$). In addition *MYC* was negatively selected in the GeCKO screen (Beta score=-0.26) as well as its binding partner *NOP16 (*Beta score -6.16, FDR=0.07) (316). MYC is a recognised oncogene which shows increased expression in 50% of cancers (317). This is in part, due to it lying at the epicentre of many proliferative signal transduction pathways. Of more relevance is MYC's ability to promote rearrangements and amplifications at certain genomic loci inducing large karyotypic abnormalities, something which liposarcoma closely exhibits (318).

7.2.4.6 Liposarcoma Markers

Throughout the literature, the consistently amplified genes in WDLS/DDLS are *MDM2*, *CDK4*, *HMGA2*, *TPSAN31*, *JUN* and *FRS2*. *MDM2*, *CDK4*, *HMGA2* and *JUN* have also importantly shown increased expression levels of mRNA in liposarcomas (97, 119, 319, 320).

MDM2 (log fold change 2.96, p<0.001), *CDK4* (log fold change 4.77, p<0.001) and *JUN* (log fold change 1.65, p<0.001) are significantly expressed in cluster three compared to one and two. *TSPAN31*, *HMGA2* and *FRS2* all had a positive log fold change but did not reach significance. The comparison between these is illustrated through a violin plot, heatmap and combined t-SNE plots in Figure 7.3-7.4.

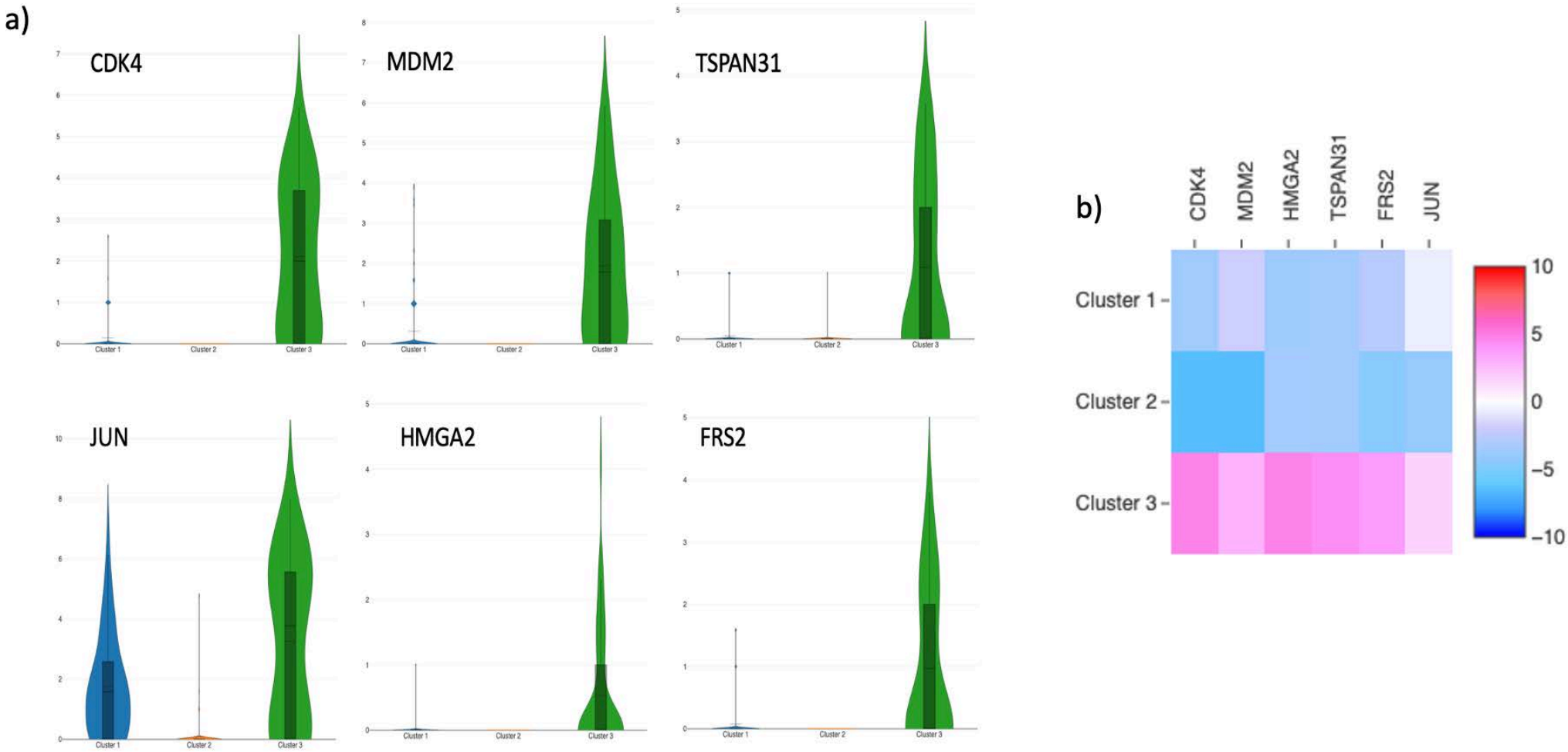

Figure 7.3: a) Violin plots for *CDK4*, *MDM2*, *TSPAN31*, *JUN*, *HMGA2* and *FRS2*. X-axis split by Cluster 1-3. Y-axis represent log fold change. Error bars indicate the IQR around the median. All genes consistently overexpressed in Cluster 3. b) Heatmap of key liposarcoma genes separated by cluster. Side bar indicates log fold change values.

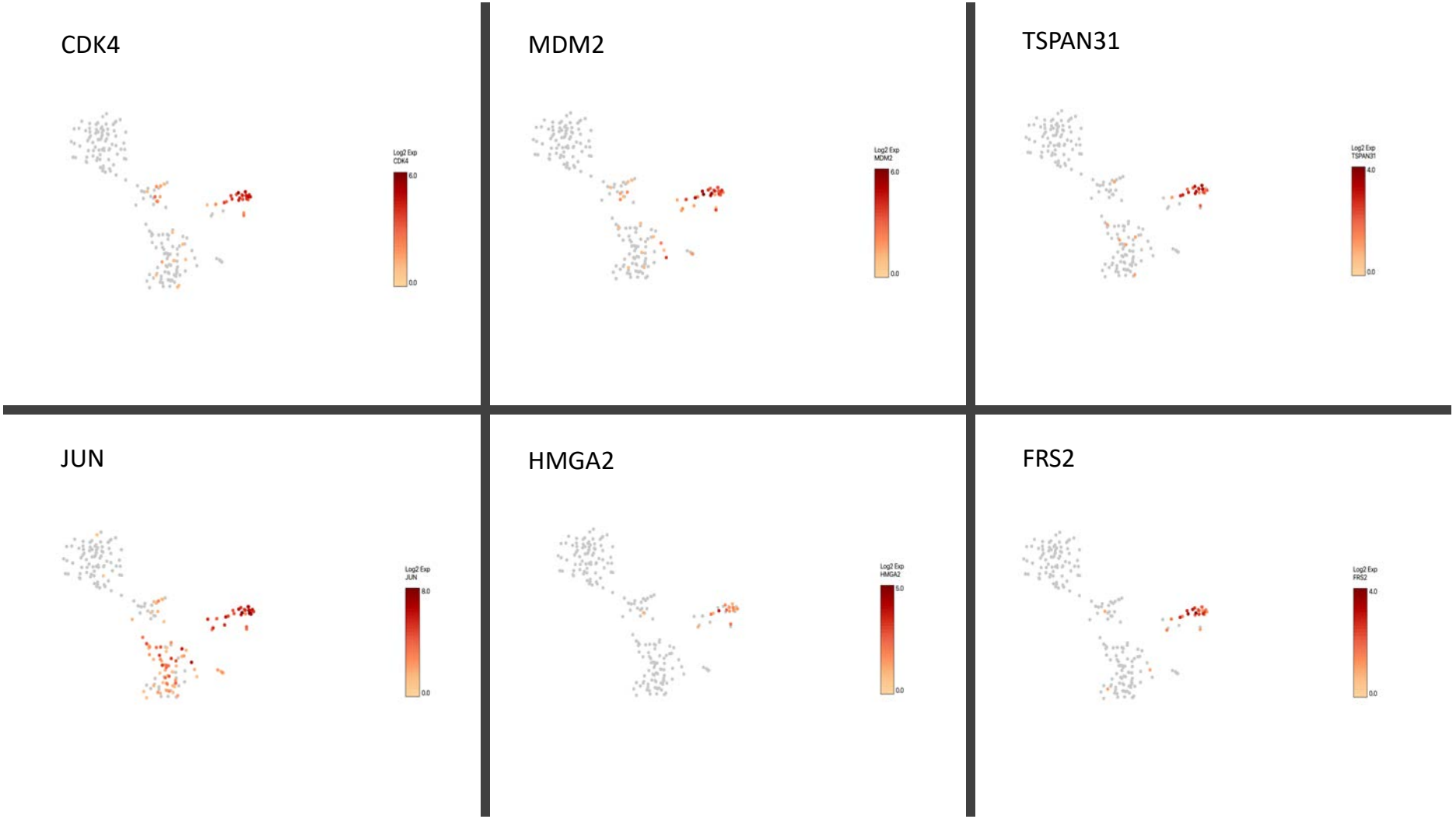

Figure 7.4: t-SNE plots per recognised liposarcoma gene. Colour bar represents log normalised expression. Cluster three contains highest expression values across all genes. There is notable *JUN* activity in Cluster one also.

7.2.4.7 Adipocytic Markers

If cluster three are indeed liposarcoma cells, there must be some supporting gene expression for an adipocytic microenvironment. WDLS microscopically appears as occasional atypical stromal cells against a mature adipocytic lipomatous backdrop (7). These adipocytes must be able to harness adipogenesis effectively, to generate such large lipid-laden cells. Differentiation of the adipocytic lineage is regulated by a complex network of transcription factors and genes, of which *PPAR-γ* and *C-EBP* proteins encoded by *CEBPA, B* and *D* are the most important (321).

In cluster three using the Loupe Browser 4.0, *PPAR*- γ has only increased expression levels in cluster 3, but this is not significant. *CEBP-D* showed significantly increased expression (log fold change 4.22, $p<0.001$), whilst *CEBP-B* was more highly expressed in Cluster 1 but not significantly so ($p=0.69$). *CEBP-D* has been proposed as one of the first transcription factors induced during induction of adipogenesis and adipocyte differentiation (322).

Other positive regulators of adipocyte differentiation include the Kruppel like factors and insulin-like growth factor binding proteins (323). In cluster three *KLF4* was significantly overexpressed (log fold change 3.93, $p<0.001$). KLF4 has been identified as an early marker of adipogenic differentiation and acts by binding directly to the promoter of CEBP-B (324). When KLF4 is knocked down in an adipocytic cell line (3T3-L1), it produced significantly less lipid than in control (325).

In a similar manner, Insulin-like growth factor 1 (IGF1) is a well-established inducer of pre-adipocyte differentiation the bioavailability of which is governed by seven IGF-

binding proteins (326). Interestingly, IGF 4-7 were all significantly expressed in cluster three, and the most significantly expressed was *IGF-6* (log fold change 6.68, $p<0.001$).

7.2.5 RNA Sequencing

The primary aim of RNA sequencing as per Chapter 5 was to elucidate evidence of SDHA gene fusions. However, differential expression analysis was also performed using a Partek Flow pipeline comparing normal tissues to tumour samples. This was undertaken to identify target genes, and to compare with both the GeCKO screen and ScRNA-seq results. Normalisation for the library size variation between samples was performed using weighted trimmed mean of log expression ratios (TMM). In total 22 genes were more significantly expressed in tumour compared to normal, of which the top ten are shown in Table 7.4.

Table 7.4: Most significantly expressed genes in tumour versus normal by log fold change. Gene function also summarised (249).

Gene symbol	Gene function	LogFC (Normal vs. Tumour)	P-value
DLK1	Pre-adipocyte transmembrane protein	-5.584962501	4.58E-02
MDM2	Ubiquitin protein ligase negatively regulates p53	-5.472487771	5.68E-03
SCUBE3	TGF-β ligand	-5.247927513	1.60E-02
FBN2	Encodes extracellular matrix protein Fibrillin-2	-5.141596278	2.57E-02
MKI67	Nuclear protein, facilitates cell cycle	-3.916476644	4.44E-02
CTNNB1	Encodes beta-catenin, which drives transcription	-3.807354922	3.39E-02
HMGA2	Encodes architectural transcription factor	-3.632268215	1.98E-02
H19	Long non-coding RNA	-3.240314329	4.63E-02
CPZ	Encodes carboxypeptidase enzyme	-3.181102551	4.37E-02
IGF2BP2	Encodes insulin-like growth factor binding protein	-3.066950244	3.85E-02

As expected *MDM2* and *HMGA2* were significantly overexpressed in tumour compared to normal. *HMGA2* has also been shown to directly regulate *IGF2BP2* which may explain its inclusion here. Li et al. described the *HMGA2-IGF2BP2-NRAS* signalling pathway as a critical oncogenic driver in lung cancer (327). *MKI67*,

still routinely reported by pathologists in soft tissue sarcoma, is a marker of cellular proliferation that tends to correlate with tumour grade.

CTNNB1 better known as ß-catenin, as described previously is an intracellular signal transducer in the Wnt pathway. Interestingly, the N-terminal domain of *CPZ* is a binding site for Wnt. The increased expression of both *CTNNB1* (log fold change -3.81, p=0.033) and *CPZ* (log fold change -3.18, p=0.043) adds to supporting data from the GeCKO screen implicating Wnt signalling in WDLS.

DLK1 is the most overexpressed gene in tumours for this dataset (log fold change - 5.58, p=0.045). This gene is normally only expressed in pre-adipocytes inhibiting differentiation into adipocytes (328). *H19* (log fold change -3.24, p=0.046) contains several binding sites for *C/EBP* transcription factors, which are central to adipogenesis. Between them these genes suggest a loss of regulation in adipocytic differentiation.

7.3 Discussion

7.3.1 Cluster One - The immune infiltrate

Gene expression patterns lead us to believe that cluster one is comprised of immune cells. Due to the number of cells sequenced it is difficult to achieve fine clustering. Nonetheless it is likely that the majority of these cells represent tumour infiltrating lymphocytes, which have been seen in multiple other solid tumours, although more samples are needed to make this assertion (329). It is also possible that a low number of sequenced tumour cells biases this cluster, overstating its importance.

Regardless of this, we see a CD4+ T-cell predominant infiltrate indicated through the high *CCR7* expression. CD4 T-cells are highly versatile cells that can differentiate into many subtypes. There are two main subtypes of CD4+ve T-cells: T-regs and T-helpers. There is evidence here to suggest T-reg infiltration, as they release TGF-ß, which appeared in the most highly expressed genes. There was also high levels of CD69 expression which T-reg precursors express also induces TGF-ß production (330). T-regs immunosuppressive action is predominantly through TGF- ß which has been shown to supress interleukins 1 and 2, and also prevent the formation of cytotoxic T-cells.

There are few studies investigating the immune landscape of liposarcoma. The first, by Tseng et al. in 2015, carried out flow cytometry and immunohistochemistry on 8 retroperitoneal liposarcomas (5WDLS/3DDLS) (331). They describe a substantial population of CD4 cells and a CD4/CD8 ratio of 4.2. These findings reflect ours, which is encouraging. However, they differ with their finding of high *PD-1* and low CD137 expression, whilst we showed low levels of PD-1 expression with some CD137 expression. The lack of PD-1 expression suggests that immune checkpoint inhibitors will have limited success here. However, an immune infiltrate rich in *TGF-ß1* expression, combined with a population of T-regs may benefit from future work looking at selectively supressing the T-reg population (332). Furthermore, recent work has shown blocking *TGF- ß1* signalling can enhance the efficacy of immune checkpoint blockade (333).

7.3.2 Cluster Two – Red cell component

It is highly likely that cluster two are red blood cells, owing to the fact that globin genes appeared most highly expressed, as well as *GYPC* which is essential to developing erythrocytes.

The exact cell type is harder to delineate. Erythroblasts undergo consecutive maturation steps, ending with enucleation to form a reticulocyte, which although contain no nucleus, contain high levels of mRNA transcripts (334). During maturation of a reticulocyte, globin mRNA's account for 95% of total cellular mRNA, which is then consumed during the lifespan of an erythrocyte (335). Kerkela et al. performed scRNA-seq on red blood cells. They detected mRNA transcripts in 10% of the cells, which were assumed to be reticulocytes at different stages of maturation (334).

The other potential source of the sequenced mRNA transcripts are platelets. Like erythrocytes, platelets are enucleated but also contain mRNA transcripts, likely derived from the megakaryocyte stage (334). Their presence here is less likely, as they do not contain haemoglobin. Additionally, two key platelet markers: P-selectin and Glycoprotein 9, were not significantly expressed in our data (336, 337).

Why red blood cells are here is a separate question. At the time the tumour was processed, it was dissected to a central portion and washed thoroughly in PBS so macroscopically had no contaminating blood cells. There was an optional red cell

lysis stage protocol which was not undertaken as we did not want to stress the cells further. It is hard to know whether a red blood cell lysis stage would have lowered the final cell count, and compromised the quality of the data, but as methodology improves in this relatively new technique, this component will likely diminish.

The vasculature of liposarcomas has not been previously described or studied, but as these are sizeable tumours, the central core must have an abundant metabolic supply network. It is possible, that these sequenced cells comprise part of the normal tumour microenvironment or are from microvasculature either in adipose or fibrous connective tissue. Mechanisms for blood cells present intratumourally include vasculogenic mimicry – where tumours cells generate vascular channels independently of endothelial blood vessels (338). Another is a hypoxic tumour microenvironment, driving HIF1-alpha mediated angiogenesis (339).

7.3.3 Cluster Three – Tumour and Stromal Component

The overexpression of MDM2 and CDK4 in cluster three support the notion liposarcoma cells are at least represented in this population. Somaiah et al. performed exome sequencing on 17 patients with retroperitoneal WDLS and DDLS (85). *MDM2* and *CDK4* amplification were the only two overlapping gene amplifications universally identified in all the samples.

Snyder et al. showed JUN expression in the majority of DDLS and their WD components, but only in the minority of pure WDLS. This is shown similarly here, as this sample is the WD component of a dedifferentiated tumour. As well as *JUN* being overexpressed in Cluster 3, it was also negatively selected in the GeCKO screen

(Beta score = -0.98). JUN's expression in cluster 2, which appear to be predominantly T-cells is not surprising. JUN is a member of the AP-1 transcription family, which plays role in T cell development and proliferation, as well as T-cell differentiation (340). More recently, Lynn et al. showed it's overexpression to prevent the induction of T-cell exhaustion programs (341).

As Table 7.1 shows, a number of ECM and mesenchymal stem cell genes are expressed. It is therefore important to consider the possibility these are a different population of cells all together, e.g fibroblasts. We accessed the Human Protein Atlas to look at other cells and tissues with high levels of *MDM2/CDK4* expression (342, 343). This found expression to highest in lung tissue, and then lymphoid tissue, which is rarely resected in this type of surgery, but most importantly low in fibroblast cell lines.

It is possible that these genes are poorly expressed at an mRNA level in immune and haematological cells creating false significance in cluster three. Using the human protein atlas again, with adipose tissue as a comparison, MDM2 expression was compared with immune cells. Values in granulocytes were comparable (19.3 NX v 16.6NX). This was albeit lower, but present in T-cells (16.6 NX v 4.4 NX). CDK4 expression levels appear very similar throughout both adipose tissue and immune cells. Therefore, our data combined with existing expression data, makes the previously discussed assertions unlikely, and that cluster three represents a genuine pool of liposarcoma cells.

Nevertheless, they display markers of mesenchymal stem cells, fibroblasts, adipocytes and liposarcoma cells. A critical, yet unanswered question in sarcoma is the exact cell of origin in each of the histological subtypes. It is now assumed that they originate from the malignant transformation of multipotent mesenchymal stem-cells, which are the precursor to mesodermal tissues, including bone, skeletal muscle and adipose tissue (344). What is not known is whether there is a distinct cell of origin for every subtype, or a multipotent cell of origin responsible for the development of all subtypes.

In favour of the latter are sarcomas with simple chromosomal defects which share a molecular pathogenesis and origin, e.g clear cell sarcoma and angiomatoid fibrous histiocytoma - completely different histologically - share both *EWSR1-CREB1* and the *EWSR1-ATF1* fusion genes (345). To counter this, and favour the 'distinct cell of origin theory', Matushansky et al. used gene clustering and distance correlation analysis to assign time points during adipogenesis that correlated to each of the four major liposarcoma subtypes (346). They believe an initial genetic change occurs in a differentiating MSC (along a mesenchymal lineage), which then arrests at that point of mesenchymal differentiation. With further genetic changes, differentiation is further impaired giving rise to dedifferentiated tissue, on the background of well differentiated tissue.

Although potentially explaining the different types of liposarcoma, Matushanksy's work conflicts with some key points. It is very rare to get MLS and PLS in the retroperitoneum, which if the theory of an adipocyte, undergoing differentiation arrest at different time points, should then occur in that anatomical location more

frequently. Secondly, Beird et al. showed shared somatic mutations, but more gene fusions in DDLS, suggesting a common ancestral clone. Genomically, WD/DDLS, PLS and MLS are all quite different which would also suggest they do not develop from one another. It is more likely that sarcoma is originates from a primitive MSC than a differentiated one, although more work is needed on this.

Assuming a mesenchymal stem-cell origin of liposarcoma cells may explain the significant expression in cluster three of *FGF7,* a fibroblast growth factor. *FGF7* was also negatively selected in the GeCKO screen and is therefore a gene of interest. *FGF7* preferentially binds to *FGFR2,* which is known to recruit the PI3k/Akt/mTOR pathways. Both FGFR and mTOR pathways are druggable, and Huang et al. have demonstrated reduced *FGF7* driven cell proliferation via mTOR inhibition (347). Combining both an mTOR inhibitor with a FGFR inhibitor could be a plausible follow-up experiment to this, based on the results presented thus far

7.3.4 RNA Sequencing

The most overexpressed gene in tumour tissue compared to normal in our SHARPS dataset was *DLK1. DLK1* is a pre-adipocyte transmembrane protein, which inhibits pre-adipocyte differentiation into adipocytes through the MAPK/ERK pathway (328). Wnt signalling has been proposed as initiating *DLK1* expression, as the Wnt/B-catenin pathway transcription factor – *LEF* – binds directly to the *DLK1* promoter (348). Skubitz et al. found *DLK1* more than eight times overexpressed in liposarcoma compared to paired normal tissue (349). The long asserted explanation for the development of liposarcoma is an adipocytic progenitor arrested in

development. Increased overexpression of *DLK1* could be responsible for this truncated adipogenesis, and therefore Wnt signalling is of interest again.

Activated Wnt signalling causes an accumulation of ß-catenin in the cell nucleus. *CTNNB1* which encodes ß-catenin was highly expressed in tumour compared to normal, as well as *CPZ* which is a Wnt signalling target. The GeCKO screen also showed evidence of the ICAT-AR-ß-catenin complex being negatively selected, implying this is needed for survival. Analysing the single cell expression data *CTNNB1* is weakly overexpressed in cluster three, which was significant (log fold change 1.21, p=0.01). We therefore have three separate experiments with data to implicate active Wnt signalling. High levels of ß-catenin can either occur through a loss-of-function mutation, resulting in an inability to degrade it. Or, it may be due to faults in other components of the pathway (350). ß-catenin mutations have not been documented in WDLS, which implies other parts of the pathway could be driving it, which is important when attempting any therapeutic intervention. Using this data, the *DLK1/LEF/B-Catenin* axis would be the target, although no drugs currently exist to block the B-Catenin/LEF interaction. Therefore, the more general inhibition of Wnt ligands using porcupine inhibitors would be the next obvious choice.

Unsurprisingly *MDM2* and *HMGA2* were significantly overexpressed compared to normal tissue. *MDM2* amplification in WDLS is historically shown through *FISH*. As sequencing becomes more economically viable, routine tumour sequencing may be all that is needed to prove *MDM2* amplification, which has the potential to offer more clinically relevant information. Although *HMGA2* is well described in WDLS, it is generally as an amplification partner of *MDM2* on chromosome 12 and its direct

contribution is not understood. The overexpression of *IGF2BP2* is noteworthy in this dataset, as this is a direct binding partner of *HMGA2*, and is implicated in embryonic rhabdomyosarcoma through the *HMGA2-IGF2BP2-NRAS* pathway (327). *IGF2BP2* was also negatively selected in the GeCKO screen (Beta score = -0.93). Overexpression in pancreatic cancers drives the *PI3K/Akt* pathway through *NRAS* which may represent an additional pathway of interest.

The number of significantly expressed genes in this dataset was lower than anticipated. This is possibly a result of underclustering on the flow cell, with a low concentration of the input library. It is also possible that expression is not significantly different between WDLS and neighbouring fat, supported by the very similar macroscopic appearance to them, or that with more paired samples there would be a greater number of significant genes. We can, however, be reassured by the appearance of MDM2 and HMGA2 within the results.

7.4 Conclusion

Cluster three shows ECM and mesenchymal growth factors predominating in tumour cells, in keeping with the microenvironment. One of these is *FGF7,* which is of interest as evidence suggests that *FGF7* drives mTOR/PI3k/AKT signalling. The overexpression of *MAPKAP1* combined with GeCKO complex data should not be ignored, and the pathway considered in WDLS. In addition, data presented in this chapter has emboldened the idea of targeting Wnt signalling. Beta-catenin was significantly expressed in tumour tissues in both total RNA sequencing and scRNA-seq.

In isolation, the scRNA-seq results suggest an predominantly T-cell immune cell infiltrate, which is intriguing. *PD-1*, an inhibitory receptor expressed on T-cells, did however, show low levels of expression in the immune cell cluster, as well as limited expression of its ligand *PD-L1* in the tumour cell pool (351). The low expression here suggests new PD1 inhibitors, which have had success elsewhere, would not be merited.

Although with checkpoint immunotherapies at the forefront of modern personalised cancer treatments, it is important that the immune microenvironment is further interrogated. Higher grade tumours such as DDLS may illicit a stronger immune response, a higher tumour mutational burden and therefore benefit from immunomodulation.

There are recognised limitations of this work, and this discussion generates more questions, than provides answers. A greater number of sequenced cells in scRNA-seq would have made the results more robust, but limitations working with adipose tissue, the tumour being left overnight and a lack of previous work in sarcomas with scRNA-seq made this a challenging protocol. Nevertheless, bringing together transcriptomic data in an attempt to link with a GeCKO screen although difficult has the potential to add meaningful data to the literature.

Chapter 8 – General Discussion

8. 1 Role of FOXD4L3 in WDLS

Preliminary work carried out before I started had identified what was originally thought to be a somatic mutation in *FOXD4L3* (chr9: 68,303, 273A>T). It was predicted to produce a stop gain, preventing the transcription of the forkhead box. This functional consequence combined with the 100% rate in preliminary experiments meant a further examination of the gene was required.

As *FOXD4L3* is contained within an area of ancestral gene duplication, with neighbouring paralogs consideration was required for the best sequencing modality. We chose nanopore sequencing due to the long-range amplicons generated, in-house expertise and reduced costs over whole genome sequencing. I found (chr9: 68,303, 273A>T) in 17/19 samples. This was seen in 10/11 tumours, and 7/8 normal samples including one blood sample. Patients were either heterozygous (50% adenine: 50% thymine) or had a 75% adenine: 25% thymine pattern. These results were difficult to interpret. Key findings were a 75:25 split in a blood sample indicating a germline origin, and a similar pattern in the cell line which can be seen as a ‘pure’ representation of WDLS, unbiased by stromal cells. These findings argued against a somatic mutation. At the time these results were generated the publication of the gnomAD database then indicated this was in fact a SNP (rs7034645) with a population frequency of around 40%.

A unifying theory to link the findings was copy number variation in the germline. In light of this new information, it appeared we had identified rs70346465 to be present in 11/12 (92%) patients. We then wanted to compare this rate with a healthy population using an orthogonal sequencing modality. I carefully designed a nesting primer that sat across an area of *FOXD4L3* that differed from its paralog *FOXD4L6*, allowing Sanger sequencing to be performed.

Comparing 4 cases with 46 healthy controls found rs7034645 present in 4/4 patients (100%) and 29/46 (63%) healthy controls giving an odds ratio of 5.34. This did not reach significance, but a wide confidence interval of 0.27-105.22 indicated it may do so with a larger sample size.

In summary, I found evidence that rs7034645 is a high-risk germline variant in WDLS. Patterns of zygosity at the SNP position could also indicate germline copy number variation. Whether it produces phenotypic effects is not answered here but further work is proposed. It may not contribute to the phenotype or do so only in combination with other SNP's. As WDLS is so rare, it is not a screening candidate, but this work further improves our biological understanding of this morbid cancer.

8.2 SDHA in WDLS and TCA cycle dysregulation

The same samples that yielded *FOXD4L3* as a gene of interest also identified recurrent fusions within *SDHA* in 5/8 samples. Sarcomas are often found to be driven by gene fusions, whilst SDH deficiency is the hallmark of certain cancers. We hypothesised that *SDHA* fusions would result in a loss of SDH function, truncating the TCA cycle with oncogenic consequences as previously described. To explore

this fully, I performed both in vitro and in vivo experiments assessing the transcriptome, proteome and metabolome of patients with WDLS.

An *SDHA* fusion panel found evidence of fusions in 5/13 tumour, and 15/19 paired normal samples. Fusions within normal tissue has been described before, although why this was at a higher level than in tumours is difficult to know. All five tumours had an exonic breakpoint in *SDHA* which raised the question of whether protein function would be affected. For this reason, evidence of protein expression was sought through Western blotting. *SDHA* expression was retained across tumours, although on semi-quantitative densitometry this was lower than in normal tissue.

At this point we had conflicting evidence for evidence of *SDHA* involvement in WDLS. In favour was a lower expression of *SDHA* and exonic breakpoints. To make an assessment of TCA cycle in vivo we made an ethics application to give patients a metabolic tracer intra-operatively. To our knowledge this had never been carried out in any sarcoma subtype before. This would give a comprehensive assessment of SDH regulation, but also the wider metabolic phenotype. We felt that if there was no loss of *SDHA* function in vivo, we would still benefit from metabolomic data.

I recruited four patients to the SHARPS study, giving them intra-operative C_{13} glucose and performing RNA sequencing and GCMS analysis on paired tumour: normal samples. The aim was to correlate transcriptional evidence of fusions with metabolomic evidence of SDH dysregulation. RNA sequencing found no evidence of significant gene fusions within *SDHA* but did show elevated expression in normal

compared to tumour (p=0.01). At this point, alternative mechanisms of impaired *SDHA* expression have to be considered. In GIST and phaeochromocytoma, these include truncating germline mutations and altered epigenetics.

GC-MS analysis found increased succinate: fumarate raw abundance across all tumour samples, but this did not translate fully when assessing label incorporation. This could have been caused by normal tissues ability to take up circulating lactate and utilise it as a carbon source. Additionally, 2 pairs had elevated succinate label incorporation compared to normal. These pairs may have had fusions which weren't detected by RNA sequencing.

To summarise, we showed impaired expression of SDHA in tumour compared to normal in both proteomic assessment and RNA sequencing experiments, with some evidence in tumour metabolomics. However, we found no evidence of gene fusions causing a loss of SDHA function in WDLS. Lastly, we demonstrated the presence of elevated aspartate and succinate levels which require further exploration.

8.3 Identifying new targets in WDLS

In general, the genomic landscape of WDLS is poorly understood. *MDM2* and *CDK4* amplification is accepted as a diagnostic marker, but there is no consensus that this is what either initiates oncogenesis or drives tumour growth. Efforts to drug these have had mixed success. *MDM2* suppression can cause adverse haematological effects, and new reports show tumour resistance mechanisms through *P53* mutations (126). We aimed to try and identify new genes or pathways responsible for WDLS by taking advantage of recent advances in CRISPR and scRNA-seq.

I performed a GeCKO dropout screen on a single WDLS cell line (93T449) to identify genes critical to WDLS survival. Two of the most negatively selected genes were involved in iron metabolism. *FDX1L* encodes an iron-sulphur complex whilst *PCPB2* an iron chaperone. These genes were also negatively selected in Depmap screens, indicating the role of iron in other cancers. However, the role of iron in liposarcoma is not established, and could yield therapeutic hope with a number of iron chelating agents available, some with anti-neoplastic activity. Further experiments are proposed in this chapter.

The screen also implicated both mTOR and Wnt signalling in WDLS. *MAPKAP*1, a key component of the mTORC2 complex was negatively selected along with *NAA30* – an mTOR substrate. The ICAT-AR-B-catenin complex was similarly negatively selected. Of interest, was the positive selection of the Wnt-inhibitor *PCDHGC3,* which was not selected by Depmap, suggesting a reliance on beta-catenin driven transcription. Evidence was sought of these pathways' involvement in both sc-RNA - sequencing and total RNA sequencing.

scRNA-seq afforded us the possibility of interrogating gene expression in the above pathways, but also assessing the tumour microenvironment. There was no published data at the time of scRNA-seq in WDLS. A WDLS component of a DDLS tumour was sequenced. t-SNE plots identified three subpopulations which on closer inspection appeared to be red blood cells, immune cells and tumour cells. The immune landscape of WDLS is very poorly understood. I found immune cells had a

high CD4:CD8 ratio, with a low PD-1 expression, which suggests a lack of immune surveillance.

The pool of tumour cells contained elevated expression of *MDM2*, *CDK4* and *HMGA2*, as well as various adipocytic markers. *FGF7* appeared significantly expressed in this population, which directly binds *FRS2*. FGFR inhibitors exist, as does evidence of *FRS2* recruitment of mTOR signalling. *CTNNB1* (beta-catenin) appeared also appeared overexpressed further implicating Wnt signalling.

The final dataset was obtained through differential RNA expression analysis of the SHARPS patients, originally undertaken to look for evidence of SDHA fusions. I was primarily interested in genes significantly expressed in tumour compared to normal. The most highly expressed compared to normal tissue was *DLK1*, a direct binding partner of *LEF* the Wnt signalling transcription factor. Furthermore, beta-catenin was again significantly overexpressed.

Three separate experiments have implicated both the Wnt and mTOR signalling pathways. These are worthy of further attention, largely thanks to the significant amount of interest and therapeutic work that has already been carried out in other cancers. In addition, FGFR inhibitors and iron chelating agents should not be discounted. The hope is this data now informs a more selected approach to a WDLS drug screen.

8.4 Limitations

The specific limitations of the methodology have been discussed in the relevant chapters. In general, we were hampered by both the type and rarity of the tumour. I found that in practice, pure WDLS are removed on average three times a year, as tumours that appeared to be WDLS both radiologically and on biopsy had dedifferentiated by the time of surgery. In the SHARPS study, Patient 2 had benign disease in the form of a retroperitoneal lipoma, whilst Patient 4 had a dedifferentiated tumour, which was unavoidable and gave a more heterogenous dataset.

I found working with adipose tissue fraught with difficulty. The quality of RNA obtained from FFPE for the SDHA fusion panel was low, even with purification steps and repeated extractions. Fat globules tend to clog up the filter chambers of spin columns and impair nucleic acid elution. Furthermore, protein extraction from adipocytic tumours faces several pitfalls. A choice was made not to use detergent as this can drop the final protein concentration. Instead, three high speed centrifugations were performed, and the non-lipid lower phase removed. The final bands were not distorted due to lipid content, but this may have lowered the final protein amount.

The SHARPS study required coordination with the anaesthetist, surgeon, theatre staff and pathologist. As surgery could be unpredictable, I could not guarantee that the infusion start time and follow on blood samples were taken at the same time in each patient. Great consideration was given to a method to immediately snap freeze tumour specimens once out of the patient. Unfortunately, hospital policy prevents liquid nitrogen in the operating room, a method adopted by Berardinis et al (224).

The quicker the cessation of metabolic activity, the closer the representation of the tumour metabolome. However, personal communication with Dr. Berardinis was reassuring. Their group compared C13 labelling ratios in needle biopsies (non-ischaemic) to those obtained after clamping the pulmonary vessels (ischaemic) and found little difference. Our specimens still frozen on average within 15 minutes and meaningful data was extracted.

The underclustering in our flow cell sequencing the SHARPS specimens was disappointing. This is a common issue when using a bioanalyser to quantify the final library. We still achieved a large number of reads and I am uncertain whether this would have made a significant difference as other gene fusions were detected. We are also limited drawing conclusions from only one sample in the scRNA-seq analysis. Furthermore, we sequenced a substantial number of red blood cells which dilutes the quality of the final comparisons. A specific red blood cell removal stage was not performed, in order to avoid stressing the cells which were low in number. This is in turn was caused by the specimen being left in the fridge overnight due to overrunning surgery. This was unavoidable, and at every stage I had to take a pragmatic approach to obtain the highest quality data from our samples.

8.5 Future Work

8.5.1 FOXD4L3

Nanopore sequencing revealed heterozygous patterns in the WDLS cohort, with a 50/50 or 75/25 ratio. Due to prohibitive costs nanopore sequencing was not performed in the healthy cohort. As large-scale nanopore sequencing has rapidly improved in the last two years, largely due to the global pandemic, a number of the healthy samples could be nanopore sequenced as part of a larger run using the PromethION (Oxford Nanopore, UK). This would illicit whether the copy number variation seen in healthy controls, or if it is unique to WDLS.

The importance of rs7034645 depends on whether it produces phenotypic consequences. Correlating SNPs and gene expression can be examined through expression quantitative trait loci (eQTL) analysis. eQTL databases are growing and these could be examined for the presence of this SNP. Whether rs7034645 is present in other sarcoma subtypes is of great interest. In 2019 Kanwal et al. found specific mutations in FOXD4L3 showed higher frequency in familial lung cancer patients compared with the healthy population (352). Gaining access to the 100K genomes database through the Genomics England Clinical Interpretation Partnership (GeCIP) would be the first step to cross-referencing the SNP with other cancer subtypes, particularly sarcomas.

8.5.2 SDHA

Metabolic tracing identified significantly high levels of aspartate in tumours, and more so in the dedifferentiated tumour. A simplistic explanation is that higher grade

tumours are more metabolically active and require more aspartate which is critical to cancer cells. However, aspartate has become an area of interest of late, with groups looking to exploit tumoural dependence on it (236). DDLS has a higher recurrence, metastasis and mortality rate. If efforts progress in other tumour types to thwart aspartate uptake in cancer cells, the treatment of DDLS could benefit.

Raw abundance of aspartate could be measured using GC-MS. DDLS tend to contain large areas of WDLS, with patchy areas of DDLS. A comparison should be made between the WDLS/DDLS components and paired normal tissue. If elevated aspartate levels are significantly higher in DDLS, then further work should be undertaken to see how these are generated. This could be achieved by infusing an aspartate tracer rather than a glucose tracer, using a similar protocol to the SHARPS study. The MID of labelled aspartate in tumours compared to normal would be the result of interest. It is possible to see this with a glucose tracer, but levels are far lower due to the prior completion of the TCA cycle. If levels are significantly higher there is the potential to use combine this with PET-scanning, to identify hotspots of aspartate uptake. This is ambitious, but often DDLS recurrences are ambiguous and require serial imaging to confirm. If DDLS is exquisitely sensitive to aspartate this may be provide a diagnostic answer.

The concept of intratumoural heterogeneity (ITH) was raised as a possible explanation for why no SDHA fusions were found in our tumour samples. To test for ITH, multiple specimens could be resected from across the tumour and whole genome sequencing compared. If tumours are polyclonal, it may explain why these tumours are so chemoresistant.

8.5.3 Identifying new targets in WDLS

Three main areas of interest were generated by this work: iron metabolism, mTOR and Wnt signalling.

Whether WDLS growth can be inhibited by iron chelation is an enticing prospect. Deferasirox, an orally administered iron chelator is routinely given in the clinical setting for transfusion related iron overload, with minimal side effects. It has been shown in human hepatoma lines to inhibit DNA replication and decrease cellular viability (353). Importantly, a much higher concentration of deferasirox is required to induce cytotoxicity in hepatocyte cultures compared to hepatoma cells (354).

Small molecule porcupine inhibitors may provide a breakthrough if WDLS has a druggable addiction to Wnt ligands. Several porcupine inhibitors are currently in phase 1 clinical trials, and have performed well in murine cancer models (276). Selective mTORC2 inhibition is now possible, with limited mTORC1 off-target effects. Importantly, our lab already has some expertise in drugging both these pathways in colorectal cancer lines. Interestingly, Zhong el al. found porcupine inhibition synergised with Pi3K/mTOR inhibition in Wnt driven pancreatic cancer, which cause enhanced suppressive effects (355).

The 93T339 cell line could be cultured in media +/- an iron chelating agent/mTORC2 inhibitor or PORCN inhibitor, and cell viability tested at different time points. This

would be a simple and inexpensive way to elucidate the reliance of WDLS on all of these pathways, and also test any therapeutic possibilities.

8.6 Conclusion

The work presented in this thesis addresses two genes in great detail, and then more broadly speculates at the involvement of several other genes and their respective pathways. Rs7034645, a SNP in *FOXD4L3*, is present at a higher frequency in WDLS. It represents a high-risk germline variant, of which none have been described to date. This data gives no evidence to suggest *SDHA* fusions cause a dysregulation of the TCA cycle in vitro or in vivo. The metabolic phenotype of WDLS is, however, altered in comparison to normal tissue and produces significantly higher amounts of aspartate. Iron metabolism could be critical to survival and proliferation, as has been shown in many previous cancers but not in WDLS. mTOR and Wnt signalling pathways were implicated in both screening and sequencing experiments and are worthy of more attention. This thesis offers new avenues of interest in WDLS – other than *MDM2* and *CDK4* – to make meaningful progress in both understanding tumour biology and targeting it.

List of References

1. J HL, G MR. Sarcomas of Soft Tissue. 5th ed. Abeloff's Clinical Oncology: Elsevier; 2013. p. 1753-91.
2. Ajithkumar T, Hatcher H. Multidisciplinary Management of Sarcomas–Where Are We Now? Clinical Oncology. 2017;29(8):467-70.
3. UK S. 'Its time to take sarcoma seriously.' News: Sarcoma UK; 2018 [Available from: https://sarcoma.org.uk/news/2018/06/its-time-take-sarcoma-seriously-new-data-suggests-sarcoma-not-rare-we-thought.
4. Cancer Research UK. Soft tissue sarcoma incidence statistics [Internet]. 2018 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/soft-tissue-sarcoma.
5. Loeb DM, Thornton K, Shokek O. Pediatric soft tissue sarcomas. Surgical Clinics of North America. 2008;88(3):615-27.
6. Lucas DR, Nascimento AG, Sanjay BK, Rock MG. Well-differentiated liposarcoma: the Mayo Clinic experience with 58 cases. American journal of clinical pathology. 1994;102(5):677-83.
7. Goldblum JR, Weiss SW, Folpe AL. Enzinger and Weiss's Soft Tissue Tumors E-Book: Elsevier Health Sciences; 2013.
8. Conyers R, Young S, Thomas DM. Liposarcoma: molecular genetics and therapeutics. Sarcoma. 2010;2011.
9. Goldblum JRMDFFF, Folpe ALMD, Weiss SWMD. Liposarcoma. In: Goldblum JRMDFFF, Folpe ALMD, Weiss SWMD, editors. Enzinger and Weiss's Soft Tissue Tumors2014. p. 484-523.
10. Strauss DC, Hayes AJ, Thomas JM. Retroperitoneal tumours: review of management. The Annals of The Royal College of Surgeons of England. 2011;93(4):275-80.
11. Grimer R, Judson I, Peake D, Seddon B. Guidelines for the management of soft tissue sarcomas. Sarcoma. 2010;2010.
12. Thway K, editor Well-differentiated liposarcoma and dedifferentiated liposarcoma: an updated review. Seminars in diagnostic pathology; 2019: Elsevier.
13. Gronchi A, Strauss DC, Miceli R, Bonvalot S, Swallow CJ, Hohenberger P, et al. Variability in patterns of recurrence after resection of primary retroperitoneal sarcoma (RPS). Annals of surgery. 2016;263(5):1002-9.
14. Henricks WH, Chu YC, Goldblum JR, Weiss SW. Dedifferentiated liposarcoma: a clinicopathological analysis of 155 cases with a proposal for an expanded definition of dedifferentiation. The American journal of surgical pathology. 1997;21(3):271-81.
15. Ishikawa K, Tohira H, Mizushima Y, Matsuoka T, Mizobata Y, Yokota J. Traumatic retroperitoneal hematoma spreads through the interfascial planes. Journal of Trauma and Acute Care Surgery. 2005;59(3):595-608.
16. Setsu N, Miyake M, Wakai S, Nakatani F, Kobayashi E, Chuman H, et al. Primary retroperitoneal myxoid liposarcomas. The American journal of surgical pathology. 2016;40(9):1286.

17. Hornick JL, Bosenberg MW, Mentzel T, McMenamin ME, Oliveira AM, Fletcher CD. Pleomorphic liposarcoma: clinicopathologic analysis of 57 cases. Am J Surg Pathol. 2004;28(10):1257-67.
18. Olimpiadi Y, Song S, Hu JS, Matcuk GR, Chopra S, Eisenberg BL, et al. Contemporary Management of Retroperitoneal Soft Tissue Sarcomas. Current Oncology Reports. 2015;17(8):39.
19. Excellence. NIfHaC. Improving Outcomes for People with Sarcoma: The Manual. NICE; 2006.
20. Lahat G, Madewell JE, Anaya DA, Qiao W, Tuvin D, Benjamin RS, et al. Computed tomography scan-driven selection of treatment for retroperitoneal liposarcoma histologic subtypes. Cancer: Interdisciplinary International Journal of the American Cancer Society. 2009;115(5):1081-90.
21. Crago AMMDP, Dickson MAMD. Liposarcoma. Surgical Oncology Clinics of North America. 2016;25(4):761-73.
22. Tseng WW, Wang SC, Eichler CM, Warren RS, Nakakura EK. Complete and safe resection of challenging retroperitoneal tumors: anticipation of multi-organ and major vascular resection and use of adjunct procedures. World journal of surgical oncology. 2011;9(1):143.
23. Group T-ARW. Management of primary retroperitoneal sarcoma (RPS) in the adult: a consensus approach from the Trans-Atlantic RPS Working Group. Annals of surgical oncology. 2015;22(1):256-63.
24. Clark MA, Fisher C, Judson I, Thomas JM. Soft-tissue sarcomas in adults. New England journal of medicine. 2005;353(7):701-11.
25. Van Houdt W, Schrijver A, Cohen-Hallaleh R, Memos N, Fotiadis N, Smith M, et al. Needle tract seeding following core biopsies in retroperitoneal sarcoma. European Journal of Surgical Oncology (EJSO). 2017;43(9):1740-5.
26. Rouse RV. Atypical Lipomatous Tumour Stanford Medicine*2007* [Available from: http://surgpathcriteria.stanford.edu/softfat/atypical_lipomatous_tumor/.
27. Lee ATJ, Thway K, Huang PH, Jones RL. Clinical and molecular spectrum of liposarcoma. Journal of Clinical Oncology. 2018;36(2):151.
28. Weaver J, Rao P, Goldblum JR, Joyce MJ, Turner SL, Lazar AJ, et al. Can MDM2 analytical tests performed on core needle biopsy be relied upon to diagnose well-differentiated liposarcoma? Modern Pathology. 2010;23(10):1301.
29. Thway K, Wang J, Swansbury J, Min T, Fisher C. Fluorescence in situ hybridization for MDM2 amplification as a routine ancillary diagnostic tool for suspected well-differentiated and dedifferentiated liposarcomas: experience at a tertiary center. Sarcoma. 2015;2015.
30. Alameddine M, Zheng I, Yusufali A, Mackrides N, Ciancio G. Adult renal sarcoma: A rare case of recurrence 13 years after initial resection. Clinical Case Reports. 2019;7(1):47.
31. Guillou L, Coindre J-M, Bonichon F, Nguyen B, Terrier P, Collin F, et al. Comparative study of the National Cancer Institute and French Federation of Cancer Centers Sarcoma Group grading systems in a population of 410 adult patients with soft tissue sarcoma. Journal of clinical oncology. 1997;15(1):350-62.
32. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. Annals of surgical oncology. 2010;17(6):1471-4.
33. Coindre J-M. Grading of soft tissue sarcomas: review and update. Archives of pathology & laboratory medicine. 2006;130(10):1448-53.

34. Gronchi A, Miceli R, Shurell E, Eilber FC, Eilber FR, Anaya DA, et al. Outcome prediction in primary resected retroperitoneal soft tissue sarcoma: histology-specific overall survival and disease-free survival nomograms built on major sarcoma center data sets. Journal of Clinical Oncology. 2013;31(13):1649-55.
35. Strauss DC. Patterns of recurrence in retroperitoneal liposarcomas: reflecting surgical approach or tumor biology? Annals of surgical oncology. 2014;21(7):2113-6.
36. Gutierrez JC, Perez EA, Moffat FL, Livingstone AS, Franceschi D, Koniaris LG. Should soft tissue sarcomas be treated at high-volume centers?: An analysis of 4205 patients. Annals of surgery. 2007;245(6):952.
37. Sandrucci S, Ponzetti A, Gianotti C, Mussa B, Lista P, Grignani G, et al. Different quality of treatment in retroperitoneal sarcomas (RPS) according to hospital-case volume and surgeon-case volume: a retrospective regional analysis in Italy. Clinical sarcoma research. 2018;8(1):3.
38. Bonvalot S, Rivoire M, Castaing M, Stoeckle E, Le Cesne A, Blay JY, et al. Primary retroperitoneal sarcomas: a multivariate analysis of surgical factors associated with local control. Journal of clinical oncology. 2009;27(1):31-7.
39. Gronchi A, Lo Vullo S, Fiore M, Mussi C, Stacchiotti S, Collini P, et al. Aggressive surgical policies in a retrospectively reviewed single-institution case series of retroperitoneal soft tissue sarcoma patients. Journal of Clinical Oncology. 2009;27(1):24-30.
40. Gronchi A, Miceli R, Colombo C, Stacchiotti S, Collini P, Mariani L, et al. Frontline extended surgery is associated with improved survival in retroperitoneal low-to intermediate-grade soft tissue sarcomas. Annals of oncology. 2011;23(4):1067-73.
41. Santos CER, Correia MM, Thuler LCS, Rosa BR, Accetta A, de Almeida Dias J, et al. Compartment surgery in treatment strategies for retroperitoneal sarcomas: a single-center experience. World journal of surgery. 2010;34(11):2773-81.
42. Lewis JJ, Leung D, Woodruff JM, Brennan MF. Retroperitoneal soft-tissue sarcoma: analysis of 500 patients treated and followed at a single institution. Annals of surgery. 1998;228(3):355.
43. Yang JC, Chang AE, Baker AR, Sindelar WF, Danforth DN, Topalian SL, et al. Randomized prospective study of the benefit of adjuvant radiation therapy in the treatment of soft tissue sarcomas of the extremity. Journal of clinical oncology. 1998;16(1):197-203.
44. Ballo MT, Zagars GK, Pollock RE, Benjamin RS, Feig BW, Cormier JN, et al. Retroperitoneal soft tissue sarcoma: an analysis of radiation and surgical treatment. International Journal of Radiation Oncology* Biology* Physics. 2007;67(1):158-63.
45. Davis AM, O'Sullivan B, Turcotte R, Bell R, Catton C, Chabot P, et al. Late radiation morbidity following randomization to preoperative versus postoperative radiotherapy in extremity soft tissue sarcoma. Radiotherapy and oncology. 2005;75(1):48-53.
46. Bonvalot S, Gronchi A, Le Péchoux C, Swallow CJ, Strauss D, Meeus P, et al. Preoperative radiotherapy plus surgery versus surgery alone for patients with primary retroperitoneal sarcoma (EORTC-62092: STRASS): a multicentre, open-label, randomised, phase 3 trial. The Lancet Oncology. 2020;21(10):1366-77.
47. Pasquali S, Gronchi A. Neoadjuvant chemotherapy in soft tissue sarcomas: latest evidence and clinical implications. Therapeutic advances in medical oncology. 2017;9(6):415-29.
48. Seymour MT, Morton D, Investigators IFT. FOxTROT: an international randomised controlled trial in 1052 patients (pts) evaluating neoadjuvant chemotherapy (NAC) for colon cancer. American Society of Clinical Oncology; 2019.

49. Wolfe AR, Miller ED, Abushahin LI, Cloyd J, Manilchuck A, Dillhoff M, et al. Neoadjuvant FOLFIRINOX versus adjuvant gemcitabine in pancreatic cancer. American Society of Clinical Oncology; 2019.
50. Saponara M, Stacchiotti S, Casali PG, Gronchi A. (Neo) adjuvant treatment in localised soft tissue sarcoma: the unsolved affair. European Journal of Cancer. 2017;70:1-11.
51. Antman K, Crowley J, Balcerzak SP, Rivkin SE, Weiss GR, Elias A, et al. An intergroup phase III randomized study of doxorubicin and dacarbazine with or without ifosfamide and mesna in advanced soft tissue and bone sarcomas. Journal of clinical oncology. 1993;11(7):1276-85.
52. Miura J, Charlson J, Gamblin T, Eastwood D, Banerjee A, Johnston F, et al. Impact of chemotherapy on survival in surgically resected retroperitoneal sarcoma. European Journal of Surgical Oncology (EJSO). 2015;41(10):1386-92.
53. Almond LM, Gronchi A, Strauss D, Jafri M, Ford S, Desai A. Neoadjuvant and adjuvant strategies in retroperitoneal sarcoma. Eur J Surg Oncol. 2018;44(5):571-9.
54. EORTC. A phase III randomized study of preoperative radiotherapy plus surgery versus surgery alone for patients with Retroperitoneal sarcomas (RPS) - STRASS.
55. Le Cesne A, Ouali M, Leahy MG, Santoro A, Hoekstra H, Hohenberger P, et al. Doxorubicin-based adjuvant chemotherapy in soft tissue sarcoma: pooled analysis of two STBSG-EORTC phase III clinical trials. Annals of Oncology. 2014;25(12):2425-32.
56. Datta J, Ecker BL, Neuwirth MG, Geha RC, Fraker DL, Roses RE, et al. Contemporary reappraisal of the efficacy of adjuvant chemotherapy in resected retroperitoneal sarcoma: evidence from a nationwide clinical oncology database and review of the literature. Surgical oncology. 2017;26(2):117-24.
57. Italiano A, Toulmonde M, Cioffi A, Penel N, Isambert N, Bompas E, et al. Advanced well-differentiated/dedifferentiated liposarcomas: role of chemotherapy and survival. Annals of Oncology. 2012;23(6):1601-7.
58. Sanfilippo R, Bertulli R, Marrari A, Fumagalli E, Pilotti S, Morosi C, et al. High-dose continuous-infusion ifosfamide in advanced well-differentiated/dedifferentiated liposarcoma. Clinical sarcoma research. 2014;4(1):16.
59. Casali P, Abecassis N, Bauer S, Biagini R, Bielack S, Bonvalot S, et al. Soft tissue and visceral sarcomas: ESMO–EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2018.
60. Schöffski P, Ray-Coquard IL, Cioffi A, Bui NB, Bauer S, Hartmann JT, et al. Activity of eribulin mesylate in patients with soft-tissue sarcoma: a phase 2 study in four independent histological subtypes. The lancet oncology. 2011;12(11):1045-52.
61. Demetri GD, Von Mehren M, Jones RL, Hensley ML, Schuetze SM, Staddon A, et al. 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. 2016;34(8):786.
62. Grosso F, Jones RL, Demetri GD, Judson IR, Blay J-Y, Le Cesne A, et al. Efficacy of trabectedin (ecteinascidin-743) in advanced pretreated myxoid liposarcomas: a retrospective study. The lancet oncology. 2007;8(7):595-602.
63. Zaidi MY, Canter R, Cardona K. Post-operative surveillance in retroperitoneal soft tissue sarcoma: The importance of tumor histology in guiding strategy. Journal of surgical oncology. 2018;117(1):99-104.

64. Ikoma N, Roland CL, Torres KE, Chiang Y-J, Wang W-L, Somaiah N, et al. Salvage surgery for recurrent retroperitoneal well-differentiated liposarcoma: early reoperation may not provide benefit. Annals of surgical oncology. 2018:1-8.
65. Gronchi A, Miceli R, Allard MA, Callegaro D, Le Péchoux C, Fiore M, et al. Personalizing the approach to retroperitoneal soft tissue sarcoma: histology-specific patterns of failure and postrelapse outcome after primary extended resection. Annals of surgical oncology. 2015;22(5):1447-54.
66. Anaya DA, Lahat G, Liu J, Xing Y, Cormier JN, Pisters PW, et al. Multifocality in retroperitoneal sarcoma: a prognostic factor critical to surgical decision-making. Annals of surgery. 2009;249(1):137-42.
67. Park JO, Qin L-X, Prete FP, Antonescu C, Brennan MF, Singer S. Predicting outcome by growth rate of locally recurrent retroperitoneal liposarcoma:"the one centimeter per month rule". Annals of surgery. 2009;250(6):977.
68. Jain S, Zhang DY, Xu R, Pincus MR, Lee P. Molecular Genetic Pathology of Solid Tumors. In: McPherson RAMDM, Pincus MRMDP, editors. Henry's Clinical Diagnosis and Management by Laboratory Methods2017. p. 1492-519.e7.
69. Grünewald TG, Cidre-Aranaz F, Surdez D, Tomazou EM, de Álava E, Kovar H, et al. Ewing sarcoma. Nature reviews Disease primers. 2018;4(1):1-22.
70. de Alava E, Gerald WL. Molecular biology of the Ewing's sarcoma/primitive neuroectodermal tumor family. Journal of Clinical Oncology. 2000;18(1):204-.
71. Tyler R, Wanigasooriya K, Taniere P, Almond M, Ford S, Desai A, et al. A review of retroperitoneal liposarcoma genomics. Cancer Treatment Reviews. 2020:102013.
72. Garsed DW, Marshall OJ, Corbin VD, Hsu A, Di Stefano L, Schröder J, et al. The architecture and evolution of cancer neochromosomes. Cancer cell. 2014;26(5):653-67.
73. Cortés-Ciriano I, Lee JJ-K, Xi R, Jain D, Jung YL, Yang L, et al. Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nature genetics. 2020;52(3):331-41.
74. Marino-Enriquez A, Fletcher CD, Dal Cin P, Hornick JL. Dedifferentiated liposarcoma with "homologous" lipoblastic (pleomorphic liposarcoma-like) differentiation: clinicopathologic and molecular analysis of a series suggesting revised diagnostic criteria. Am J Surg Pathol. 2010;34(8):1122-31.
75. Taylor BS, Barretina J, Maki RG, Antonescu CR, Singer S, Ladanyi M. Advances in sarcoma genomics and new therapeutic targets. Nature Reviews Cancer. 2011;11(8):541.
76. Szymanska J, Virolainen M, Tarkkanen M, Wiklund T, Asko-Seljavaara S, Tukiainen E, et al. Overrepresentation of 1q21–23 and 12q13–21 in lipoma-like liposarcomas but not in benign lipomas: a comparative genomic hybridization study. Cancer genetics and cytogenetics. 1997;99(1):14-8.
77. Coindre J-M, Pédeutour F, Aurias A. Well-differentiated and dedifferentiated liposarcomas. Virchows Archiv. 2010;456(2):167-79.
78. Taylor BS, DeCarolis PL, Angeles CV, Brenet F, Schultz N, Antonescu CR, et al. Frequent alterations and epigenetic silencing of differentiation pathway genes in structurally rearranged liposarcomas. Cancer discovery. 2011:CD-11-0181.
79. Oliner JD, Saiki AY, Caenepeel S. The role of MDM2 amplification and overexpression in tumorigenesis. Cold Spring Harbor perspectives in medicine. 2016:a026336.
80. Mendrysa SM, McElwee MK, Michalowski J, O'Leary KA, Young KM, Perry ME. mdm2 Is critical for inhibition of p53 during lymphopoiesis and the response to ionizing irradiation. Molecular and cellular biology. 2003;23(2):462-72.

81. Haupt Y, Maya R, Kazaz A, Oren M. Mdm2 promotes the rapid degradation of p53. nature. 1997;387(6630):296.
82. Stommel JM, Wahl GM. Accelerated MDM2 auto-degradation induced by DNA-damage kinases is required for p53 activation. The EMBO journal. 2004;23(7):1547-56.
83. Toledo F, Wahl GM. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. Nature Reviews Cancer. 2006;6(12):909.
84. Dei Tos AP, Doglioni C, Piccinin S, Sciot R, Furlanetto A, Boiocchi M, et al. Coordinated expression and amplification of the MDM2, CDK4, and HMGI-C genes in atypical lipomatous tumours. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland. 2000;190(5):531-6.
85. Somaiah N, Beird HC, Barbo A, Song J, Shaw KRM, Wang W-L, et al. Targeted next generation sequencing of well-differentiated/dedifferentiated liposarcoma reveals novel gene amplifications and mutations. Oncotarget. 2018;9(28):19891.
86. Bill KLJ, Seligson ND, Hays JL, Awasthi A, Demoret B, Stets CW, et al. Degree of MDM2 Amplification Affects Clinical Outcomes in Dedifferentiated Liposarcoma. The oncologist. 2019:theoncologist. 2019-0047.
87. Ricciotti RW, Baraff AJ, Jour G, Kyriss M, Wu Y, Liu Y, et al. High amplification levels of MDM2 and CDK4 correlate with poor outcome in patients with dedifferentiated liposarcoma: A cytogenomic microarray analysis of 47 cases. Cancer genetics. 2017;218:69-80.
88. Ware PL, Snow AN, Gvalani M, Pettenati MJ, Qasem SA. MDM2 copy numbers in well-differentiated and dedifferentiated liposarcoma: characterizing progression to high-grade tumors. American journal of clinical pathology. 2014;141(3):334-41.
89. Binh MBN, Sastre-Garau X, Guillou L, de Pinieux G, Terrier P, Lagacé R, et al. MDM2 and CDK4 immunostainings are useful adjuncts in diagnosing well-differentiated and dedifferentiated liposarcoma subtypes: a comparative analysis of 559 soft tissue neoplasms with genetic data. The American journal of surgical pathology. 2005;29(10):1340-7.
90. Italiano A, Bianchini L, Gjernes E, Keslair F, Ranchere-Vince D, Dumollard J-M, et al. Clinical and biological significance of CDK4 amplification in well-differentiated and dedifferentiated liposarcomas. Clinical cancer research. 2009:1078-0432. CCR-08-3185.
91. Sotillo R, Dubus P, Martn J, de la Cueva E, Ortega S, Malumbres M, et al. Wide spectrum of tumors in knock-in mice carrying a Cdk4 protein insensitive to INK4 inhibitors. The EMBO journal. 2001;20(23):6637-47.
92. Dei Tos AP, Doglioni C, Piccinin S, Maestro R, Mentzel T, Barbareschi M, et al. Molecular abnormalities of the p53 pathway in dedifferentiated liposarcoma. The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland. 1997;181(1):8-13.
93. Lee S, Park H, Ha SY, Paik KY, Lee SE, Kim JM, et al. CDK4 amplification predicts recurrence of well-differentiated liposarcoma of the abdomen. PloS one. 2014;9(8):e99452.
94. Hélias-Rodzewicz Z, Pédeutour F, Coindre JM, Terrier P, Aurias A. Selective elimination of amplified CDK4 sequences correlates with spontaneous adipocytic differentiation in liposarcoma. Genes, Chromosomes and Cancer. 2009;48(11):943-52.
95. Perez M, Muñoz-Galván S, Jiménez-García MP, Marín JJ, Carnero A. Efficacy of CDK4 inhibition against sarcomas depends on their levels of CDK4 and p16ink4 mRNA. Oncotarget. 2015;6(38):40557.
96. Italiano A, Bianchini L, Keslair F, Bonnafous S, Cardot-Leccia N, Coindre JM, et al. HMGA2 is the partner of MDM2 in well-differentiated and dedifferentiated liposarcomas

whereas CDK4 belongs to a distinct inconsistent amplicon. International journal of cancer. 2008;122(10):2233-41.
97. Meza-Zepeda LA, Berner JM, Henriksen J, South AP, Pedeutour F, Dahlberg AB, et al. Ectopic sequences from truncated HMGIC in liposarcomas are derived from various amplified chromosomal regions. Genes, Chromosomes and Cancer. 2001;31(3):264-73.
98. Wang X, Hulshizer RL, Erickson-Johnson MR, Flynn HC, Jenkins RB, Lloyd RV, et al. Identification of novel HMGA2 fusion sequences in lipoma: Evidence that deletion of let-7 miRNA consensus binding site 1 in the HMGA2 3' UTR is not critical for HMGA2 transcriptional upregulation. Genes, Chromosomes and Cancer. 2009;48(8):673-8.
99. Saâda-Bouzid E, Burel-Vandenbos F, Ranchère-Vince D, Birtwisle-Peyrottes I, Chetaille B, Bouvier C, et al. Prognostic value of HMGA2, CDK4, and JUN amplification in well-differentiated and dedifferentiated liposarcomas. Modern pathology. 2015;28(11):1404.
100. Xi Y, Shen W, Ma L, Zhao M, Zheng J, Bu S, et al. HMGA2 promotes adipogenesis by activating C/EBPβ-mediated expression of PPARγ. Biochemical and biophysical research communications. 2016;472(4):617-23.
101. Arlotta P, Tai AK, Manfioletti G, Clifford C, Jay G, Ono SJ. Transgenic mice expressing a truncated form of the high mobility group IC protein develop adiposity and an abnormally high prevalence of lipomas. Journal of Biological Chemistry. 2000.
102. Narita M, Narita M, Krizhanovsky V, Nuñez S, Chicas A, Hearn SA, et al. A novel role for high-mobility group a proteins in cellular senescence and heterochromatin formation. Cell. 2006;126(3):503-14.
103. Meltzer PS, Jankowski SA, Dal Cin P, Sandberg A, Paz I, Coccia M. Identification and cloning of a novel amplified DNA sequence in human malignant fibrous histiocytoma derived from a region of chromosome 12 frequently rearranged in soft tissue tumors. Cell Growth Differ. 1991;2(10):495-501.
104. Noble-Topham S, Kandel R, Burrow S, Bell R, Eppert K, Meltzer P, et al. SAS is amplified predominantly in surface osteosarcoma. Journal of orthopaedic research. 1996;14(5):700-5.
105. Forus A, Florenes V, Maelandsmo GM, Meltzer PS, Fodstad O, Myklebost O. Mapping of amplification units in the q13-14 region of chromosome 12 in human sarcomas: some amplica do not include MDM2. Cell growth & differentiation: the molecular biology journal of the American Association for Cancer Research. 1993;4(12):1065-70.
106. Creytens D, VAN GORP J, Speel E-J, Ferdinande L. Characterization of the 12q amplicons in lipomatous soft tissue tumors by multiplex ligation-dependent probe amplification-based copy number analysis. Anticancer research. 2015;35(4):1835-42.
107. Pawson T, Scott JD. Signaling through scaffold, anchoring, and adaptor proteins. Science. 1997;278(5346):2075-80.
108. Wang X, Asmann YW, Erickson-Johnson MR, Oliveira JL, Zhang H, Moura RD, et al. High-resolution genomic mapping reveals consistent amplification of the fibroblast growth factor receptor substrate 2 gene in well-differentiated and dedifferentiated liposarcoma. Genes, Chromosomes and Cancer. 2011;50(11):849-58.
109. Jing W, Lan T, Chen H, Zhang Z, Chen M, Peng R, et al. Amplification of FRS2 in atypical lipomatous tumour/well-differentiated liposarcoma and de-differentiated liposarcoma: a clinicopathological and genetic study of 146 cases. Histopathology. 2018;72(7):1145-55.

110. Zhang K, Chu K, Wu X, Gao H, Wang J, Yuan Y-C, et al. Amplification of FRS2 and activation of FGFR/FRS2 signaling pathway in high-grade liposarcoma. Cancer research. 2013;73(4):1298-307.
111. Park JH, Roeder RG. GAS41 is required for repression of the p53 tumor suppressor pathway during normal cellular proliferation. Molecular and cellular biology. 2006;26(11):4006-16.
112. Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, Decarolis PL, et al. Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. Nat Genet. 2010;42(8):715-21.
113. Pikor LA, Lockwood WW, Thu KL, Vucic EA, Chari R, Gazdar AF, et al. YEATS4 is a novel oncogene amplified in non–small cell lung cancer that regulates the p53 pathway. Cancer research. 2013.
114. Tao K, Yang J, Hu Y, Deng A. Knockdown of YEATS4 inhibits colorectal cancer cell proliferation and induces apoptosis. American journal of translational research. 2015;7(3):616.
115. Egan JB, Barrett MT, Champion MD, Middha S, Lenkiewicz E, Evers L, et al. Whole genome analyses of a well-differentiated liposarcoma reveals novel SYT1 and DDR2 rearrangements. PloS one. 2014;9(2):e87113.
116. Soini Y. Epigenetic and genetic changes in soft tissue sarcomas: a review. Apmis. 2016;124(11):925-34.
117. Singer S, Socci ND, Ambrosini G, Sambol E, Decarolis P, Wu Y, et al. Gene expression profiling of liposarcoma identifies distinct biological types/subtypes and potential therapeutic targets in well-differentiated and dedifferentiated liposarcoma. Cancer research. 2007;67(14):6626-36.
118. Tap WD, Eilber FC, Ginther C, Dry SM, Reese N, Barzan-Smith K, et al. Evaluation of well-differentiated/de-differentiated liposarcomas by high-resolution oligonucleotide array-based comparative genomic hybridization. Genes, Chromosomes and Cancer. 2011;50(2):95-112.
119. Mariani O, Brennetot C, Coindre J-M, Gruel N, Ganem C, Delattre O, et al. JUN oncogene amplification and overexpression block adipocytic differentiation in highly aggressive sarcomas. Cancer cell. 2007;11(4):361-74.
120. Beird HC, Wu C-C, Ingram DR, Wang W-L, Alimohamed A, Gumbs C, et al. Genomic profiling of dedifferentiated liposarcoma compared to matched well-differentiated liposarcoma reveals higher genomic complexity and a common origin. Molecular Case Studies. 2018;4(2):a002386.
121. Horvai AE, DeVries S, Roy R, O'donnell RJ, Waldman F. Similarity in genetic alterations between paired well-differentiated and dedifferentiated components of dedifferentiated liposarcoma. Modern Pathology. 2009;22(11):1477.
122. Casali P, Abecassis N, Bauer S, Biagini R, Bielack S, Bonvalot S, et al. Gastrointestinal stromal tumours: ESMO–EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology. 2018;29(Supplement_4):iv68–iv78.
123. Müller CR, Paulsen EB, Noordhuis P, Pedeutour F, Sæter G, Myklebost O. Potential for treatment of liposarcomas with the MDM2 antagonist Nutlin-3A. International journal of cancer. 2007;121(1):199-205.
124. Ray-Coquard I, Blay J-Y, Italiano A, Le Cesne A, Penel N, Zhi J, et al. Effect of the MDM2 antagonist RG7112 on the P53 pathway in patients with MDM2-amplified, well-

differentiated or dedifferentiated liposarcoma: an exploratory proof-of-mechanism study. The lancet oncology. 2012;13(11):1133-40.
125. Wagner AJ, Banerji U, Mahipal A, Somaiah N, Hirsch H, Fancourt C, et al. Phase I trial of the human double minute 2 inhibitor MK-8242 in patients with advanced solid tumors. Journal of Clinical Oncology. 2017;35(12):1304.
126. Jung J, Lee JS, Dickson MA, Schwartz GK, Le Cesne A, Varga A, et al. TP53 mutations emerge with HDM2 inhibitor SAR405838 treatment in de-differentiated liposarcoma. Nature communications. 2016;7:12609.
127. Dickson MA, Tap WD, Keohan ML, D'Angelo SP, Gounder MM, Antonescu CR, et al. Phase II trial of the CDK4 inhibitor PD0332991 in patients with advanced CDK4-amplified well-differentiated or dedifferentiated liposarcoma. Journal of clinical oncology. 2013;31(16):2024.
128. McGovern Y, Zhou CD, Jones RL. Systemic Therapy in Metastatic or Unresectable Well-Differentiated/Dedifferentiated Liposarcoma. Frontiers in oncology. 2017;7:292.
129. Dickson MA, Schwartz GK, Keohan ML, D'Angelo SP, Gounder MM, Chi P, et al. Progression-free survival among patients with well-differentiated or dedifferentiated liposarcoma treated with CDK4 inhibitor palbociclib: a phase 2 clinical trial. JAMA oncology. 2016;2(7):937-40.
130. Luke JJ, D'Adamo DR, Dickson MA, Keohan ML, Carvajal RD, Maki RG, et al. The cyclin-dependent kinase inhibitor flavopiridol potentiates doxorubicin efficacy in advanced sarcomas: preclinical investigations and results of a phase I dose-escalation clinical trial. Clinical Cancer Research. 2012;18(9):2638-47.
131. Infante JR, Cassier PA, Gerecitano JF, Witteveen PO, Chugh R, Ribrag V, et al. A phase I study of the cyclin-dependent kinase 4/6 inhibitor ribociclib (LEE011) in patients with advanced solid tumors and lymphomas. Clinical Cancer Research. 2016;22(23):5696-705.
132. Demetri GD, Fletcher CD, Mueller E, Sarraf P, Naujoks R, Campbell N, et al. Induction of solid tumor differentiation by the peroxisome proliferator-activated receptor-gamma ligand troglitazone in patients with liposarcoma. Proc Natl Acad Sci U S A. 1999;96(7):3951-6.
133. Debrock G, Vanhentenrijk V, Sciot R, Debiec-Rychter M, Oyen R, Van Oosterom A. A phase II trial with rosiglitazone in liposarcoma patients. British journal of cancer. 2003;89(8):1409.
134. Tseng WW, Somaiah N, Lazar AJ, Lev DC, Pollock RE. Novel systemic therapies in advanced liposarcoma: a review of recent clinical trial results. Cancers. 2013;5(2):529-49.
135. Tyler R, Dilworth M, James J, Blakeway D, Stockton J, Morton D, et al. The Molecular Landscape of Well Differentiated Retroperitoneal Liposarcoma. 2021.
136. Catalogue Of Somatic Mutations In Cancer. FOXD4L3 [Internet]. 2020 [Available from: https://cancer.sanger.ac.uk/cosmic/gene/analysis?ln=FOXD4L3.
137. Gene Expression Omnibus. FOXD4L3. 2017.
138. Perry JA, Seong BKA, Stegmaier K. Biology and therapy of dominant fusion oncoproteins involving transcription factor and chromatin regulators in sarcomas. Annual Review of Cancer Biology. 2019;3:299-321.
139. Uddin F, Rudin CM, Sen T. CRISPR gene therapy: applications, limitations, and implications for the future. Frontiers in Oncology. 2020;10:1387.
140. Cong L, Zhang F. Genome engineering using CRISPR-Cas9 system. Chromosomal mutagenesis: Springer; 2015. p. 197-217.

141. Thermofisher. TURBO DNA-free Kit [Internet]. 2019 [Available from: http://tools.thermofisher.com/content/sfs/manuals/cms_055740.pdf.
142. Covaris. truXTRAC® FFPE total NA Kit 2017 [Available from: https://covaris.com/wp/wp-content/uploads/resources_pdf/pn_010434.pdf.
143. UCSC. In-Silico PCR [Internet]. 2020 [Available from: https://genome.ucsc.edu/cgi-bin/hgPcr.
144. Oxford Nanopore. Ligation Sequencing Kit Protocol [Internet]. 2018 [Available from: https://store.nanoporetech.com/uk/media/wysiwyg/pdfs/SQK-LSK109/Genomic_DNA_by_Ligation_SQK-LSK109_-minion.pdf.
145. Thermofisher. Peirce BCA Protein Kit [Internet]. 2018 [Available from: https://www.thermofisher.com/order/catalog/product/23227#/23227.
146. Qiagen. QIAseq Targeted RNAscan Panels Handbook [Internet]. 2018 [Available from: https://www.qiagen.com/gb/resources/resourcedetail?id=22397db6-49fa-40cc-85c2-b58c4fa54a72&lang=en.
147. Brotherton P, Endicott P, Beaumont M, Barnett R, Austin J, Cooper A, et al. Single primer extension (SPEX) amplification to accurately genotype highly damaged DNA templates. Forensic Science International: Genetics Supplement Series. 2008;1(1):19-21.
148. Hiller K, Hangebrauk J, Jäger C, Spura J, Schreiber K, Schomburg D. MetaboliteDetector: comprehensive analysis tool for targeted and nontargeted GC/MS based metabolome analysis. Analytical chemistry. 2009;81(9):3429-39.
149. Joung J, Konermann S, Gootenberg JS, Abudayyeh OO, Platt RJ, Brigham MD, et al. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. Nature protocols. 2017;12(4):828-63.
150. Zymo Research. Zymo Midi Prep Plus Kit [Internet]. 2019 [Available from: https://files.zymoresearch.com/protocols/_d4075_quick-dna_midiprep_plus_kit.pdf.
151. New England Biolabs. Monarch DNA Gel Extraction KIt Protocol [Internet]. 2019 [Available from: https://international.neb.com/protocols/2015/11/23/monarch-dna-gel-extraction-kit-protocol-t1020.
152. Gene Cards. FOXD4L3 Gene [Internet]. 2018 [Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=FOXD4L3.
153. Gene Ontology Annotation (GOA) Database. FOXD4L3 [Internet]. 2018 [Available from: https://www.ebi.ac.uk/QuickGO/searchproducts/FOXD4L3.
154. The Human Protein Atlas. FOXD4L3 Tissue Atlas [Website]. 2018 [Available from: https://www.proteinatlas.org/ENSG00000187559-FOXD4L3/tissue.
155. Lam EW-F, Brosens JJ, Gomes AR, Koo C-Y. Forkhead box proteins: tuning forks for transcriptional harmony. Nature reviews Cancer. 2013;13(7):482-95.
156. Li J, Vogt PK. The retroviral oncogene qin belongs to the transcription factor family that includes the homeotic gene fork head. Proceedings of the National Academy of Sciences. 1993;90(10):4490-4.
157. Jackson BC, Carpenter C, Nebert DW, Vasiliou V. Update of human and mouse forkhead box (FOX) gene families. Human genomics. 2010;4(5):1-8.
158. Yang Y, Blee AM, Wang D, An J, Pan Y, Yan Y, et al. Loss of FOXO1 cooperates with TMPRSS2–ERG overexpression to promote prostate tumorigenesis and cell invasion. Cancer research. 2017;77(23):6524-37.
159. Hug N, Longman D, Cáceres JF. Mechanism and regulation of the nonsense-mediated decay pathway. Nucleic acids research. 2016;44(4):1483-95.

160. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007;446(7132):153-8.
161. Merid SK, Goranskaya D, Alexeyenko A. Distinguishing between driver and passenger mutations in individual cancer genomes by network enrichment analysis. BMC bioinformatics. 2014;15(1):308.
162. Jensen RA. Orthologs and paralogs-we need to get it right. Genome biology. 2001;2(8):interactions1002. 1.
163. Fan CM, Kuwana E, Bulfone A, Fletcher CF, Copeland NG, Jenkins NA, et al. "Expression patterns of two murine homologs of Drosophila single-minded suggest possible roles in embryonic patterning and in the pathogenesis of Down syndrome.". Mol Cell Neurosci. 1996;7(6):519.
164. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. Human molecular genetics. 2018;27(R2):R234-R41.
165. Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P, Snutch TP. MinION-based long-read sequencing and assembly extends the Caenorhabditis elegans reference genome. Genome research. 2018;28(2):266-74.
166. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, et al. The potential and challenges of nanopore sequencing. Nanoscience and technology: A collection of reviews from Nature Journals: World Scientific; 2010. p. 261-8.
167. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.
168. gnomAD browser. Single nucleotide variant: 9-70918189-A-T (GRCh37) [Internet]. 2018 [Available from: https://gnomad.broadinstitute.org/variant/9-70918189-A-T.
169. Deng N, Zhou H, Fan H, Yuan Y. Single nucleotide polymorphisms and cancer susceptibility. Oncotarget. 2017;8(66):110635.
170. Dewey FE, Pan S, Wheeler MT, Quake SR, Ashley EA. DNA sequencing: clinical applications of new DNA sequencing technologies. Circulation. 2012;125(7):931-44.
171. Oregon State University. Troubleshooting Guide to Sanger Sequencing 2020 [Available from: https://cgrb.oregonstate.edu/sites/cgrb.oregonstate.edu/files/files/CoreLab/ab_sanger_troubleshooting_guide.pdf.
172. Matthyssens LE, Creytens D, Ceelen WP. Retroperitoneal liposarcoma: current insights in diagnosis and treatment. Front Surg. 2015;2:4.
173. Kumaran M, Cass CE, Graham K, Mackey JR, Hubaux R, Lam W, et al. Germline copy number variations are associated with breast cancer risk and prognosis. Scientific reports. 2017;7(1):1-15.
174. Gazdar AF, Gao B, Minna JD. Lung cancer cell lines: Useless artifacts or invaluable tools for medical science? Lung cancer. 2010;68(3):309-18.
175. Depmap portal. 93T449 [(Internet)]. 2020 [Available from: https://depmap.org/portal/cell_line/ACH-001794?tab=mutation
.
176. Szumilas M. Explaining odds ratios. Journal of the Canadian academy of child and adolescent psychiatry. 2010;19(3):227.
177. Liang L, Lu G, Guogang Pan G, Deng Y, Liang J, Liang L, et al. A case-control study of the association between the SPP1 gene SNPs and the susceptibility to breast cancer in Guangxi, China. Frontiers in Oncology. 2019;9:1415.

178. Miao C, Liu D, Zhang F, Wang Y, Zhang Y, Yu J, et al. Association of FPGS genetic polymorphisms with primary retroperitoneal liposarcoma. Scientific Reports. 2015;5:9079.
179. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. Philosophical Transactions of the Royal Society B: Biological Sciences. 2013;368(1620):20120362.
180. Wilson JMG, Jungner G, Organization WH. Principles and practice of screening for disease. 1968.
181. McClurg JS. Gene and Operon Fusions. 2nd ed. Brenner SE, editor2013.
182. Tuna M, Amos CI, Mills GB. Molecular mechanisms and pathobiology of oncogenic fusion transcripts in epithelial tumors. Oncotarget. 2019;10(21):2095.
183. Shtivelman E, Lifshitz B, Gale RP, Canaani E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature. 1985;315(6020):550-4.
184. Parker BC, Annala MJ, Cogdell DE, Granberg KJ, Sun Y, Ji P, et al. The tumorigenic FGFR3-TACC3 gene fusion escapes miR-99a regulation in glioblastoma. The Journal of clinical investigation. 2013;123(2).
185. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. Nature Reviews Cancer. 2007;7(4):233-45.
186. Jain S, Xu R, Prieto VG, Lee P. Molecular classification of soft tissue sarcomas and its clinical applications. International journal of clinical and experimental pathology. 2010;3(4):416.
187. Li X, McGee-Lawrence ME, Decker M, Westendorf JJ. The Ewing's sarcoma fusion protein, EWS-FLI, binds Runx2 and blocks osteoblast differentiation. Journal of cellular biochemistry. 2010;111(4):933-43.
188. Brennan MF, Antonescu CR, Maki RG. Management of soft tissue sarcoma: Springer; 2013.
189. Narendra S, Valente A, Tull J, Zhang S. DDIT3 gene break-apart as a molecular marker for diagnosis of myxoid liposarcoma—assay validation and clinical experience. Diagnostic Molecular Pathology. 2011;20(4):218-24.
190. Brill E, Gobble R, Angeles C, Lagos-Quintana M, Crago A, Laxa B, et al. ZIC1 overexpression is oncogenic in liposarcoma. Cancer research. 2010;70(17):6891-901.
191. Sánchez-Martín M, González-Herrero I, Sánchez-García I. Soft tissue tumors: Liposarcoma: Myxoid liposarcoma. Atlas of Genetics and Cytogenetics in Oncology and Haematology. 2005.
192. Ensembl. SDHA [(Internet)]. 2018 [Available from: https://www.ensembl.org/Homo_sapiens/Gene/Summary?g=ENSG00000073578;r=5:218241-257082.
193. Renkema GH, Wortmann SB, Smeets RJ, Venselaar H, Antoine M, Visser G, et al. SDHA mutations causing a multisystem mitochondrial disease: novel mutations and genetic overlap with hereditary tumors. European Journal of Human Genetics. 2015;23(2):202-9.
194. Wagner AJ, Remillard SP, Zhang Y-X, Doyle LA, George S, Hornick JL. Loss of expression of SDHA predicts SDHA mutations in gastrointestinal stromal tumors. Modern Pathology. 2013;26(2):289-94.
195. Burnichon N, Briere J-J, Libé R, Vescovo L, Riviere J, Tissier F, et al. SDHA is a tumor suppressor gene causing paraganglioma. Human molecular genetics. 2010;19(15):3011-20.
196. Dwight T, Benn DE, Clarkson A, Vilain R, Lipton L, Robinson BG, et al. Loss of SDHA expression identifies SDHA mutations in succinate dehydrogenase–deficient gastrointestinal stromal tumors. The American journal of surgical pathology. 2013;37(2):226–33.

197. Italiano A, Chen C-L, Sung Y-S, Singer S, DeMatteo RP, LaQuaglia MP, et al. SDHA loss of function mutations in a subset of young adult wild-type gastrointestinal stromal tumors. BMC cancer. 2012;12(1):1-7.
198. Selak MA, Armour SM, MacKenzie ED, Boulahbel H, Watson DG, Mansfield KD, et al. Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF-α prolyl hydroxylase. Cancer cell. 2005;7(1):77-85.
199. Bardella C, Pollard PJ, Tomlinson I. SDH mutations in cancer. Biochimica et Biophysica Acta (BBA)-Bioenergetics. 2011;1807(11):1432-43.
200. Pérez-Pérez R, López JA, García-Santos E, Camafeita E, Gómez-Serrano M, Ortega-Delgado FJ, et al. Uncovering suitable reference proteins for expression studies in human adipose tissue with relevance to obesity. PLoS One. 2012;7(1):e30326.
201. van Nederveen FH, Gaal J, Favier J, Korpershoek E, Oldenburg RA, de Bruyn EM, et al. An immunohistochemical procedure to detect patients with paraganglioma and phaeochromocytoma with germline SDHB, SDHC, or SDHD gene mutations: a retrospective and prospective analysis. The lancet oncology. 2009;10(8):764-71.
202. Demichelis F, Fall K, Perner S, Andrén O, Schmidt F, Setlur S, et al. TMPRSS2: ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene. 2007;26(31):4596-9.
203. Nakano K, Takahashi S. Translocation-related sarcomas. International journal of molecular sciences. 2018;19(12):3784.
204. Crozat A, Åman P, Mandahl N, Ron D. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature. 1993;363(6430):640-4.
205. Nature Media. Differentially expressed promoters, as derived by RnBeads [Internet]. 2018 [Available from: https://media.nature.com/original/nature-assets/tp/journal/v6/n1/extref/tp2015210x2.pdf.
206. Annala M, Parker B, Zhang W, Nykter M. Fusion genes and their discovery using high throughput sequencing. Cancer letters. 2013;340(2):192-200.
207. Keegan NP, Wilton SD, Fletcher S. Breakpoint junction features of seven DMD deletion mutations. Human genome variation. 2019;6(1):1-7.
208. Yoshihara K, Wang Q, Torres-Garcia W, Zheng S, Vegesna R, Kim H, et al. The landscape and therapeutic relevance of cancer-associated transcript fusions. Oncogene. 2015;34(37):4845-54.
209. Oliver GR, Jenkinson G, Klee EW. Computational detection of known pathogenic gene fusions in a Normal tissue database and implications for genetic disease research. Frontiers in Genetics. 2020;11:173.
210. The Human Protein Atlas. SDHA 2020 [Available from: https://www.proteinatlas.org/ENSG00000073578-SDHA.
211. Tareen SH, Kutmon M, Adriaens ME, Mariman EC, de Kok TM, Arts IC, et al. Exploring the cellular network of metabolic flexibility in the adipose tissue. Genes & nutrition. 2018;13(1):17.
212. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, et al. Transcriptome sequencing to detect gene fusions in cancer. Nature. 2009;458(7234):97-101.
213. Heyer EE, Deveson IW, Wooi D, Selinger CI, Lyons RJ, Hayes VM, et al. Diagnosis of fusion genes using targeted RNA sequencing. Nature communications. 2019;10(1):1-12.
214. Liu J, Dang H, Wang XW. The significance of intertumor and intratumor heterogeneity in liver cancer. Experimental & molecular medicine. 2018;50(1):e416-e.

215. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl j Med. 2012;366:883-92.
216. Ibrahim A, Chopra S. Succinate Dehydrogenase–Deficient Gastrointestinal Stromal Tumors. Archives of Pathology & Laboratory Medicine. 2020;144(5):655-60.
217. Ravegnini G, Ricci R. Succinate dehydrogenase-deficient gastrointestinal stromal tumors: small steps toward personalized medicine? Epigenetics insights. 2019;12:2516865719842534.
218. DeBerardinis RJ, Mancuso A, Daikhin E, Nissim I, Yudkoff M, Wehrli S, et al. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. Proceedings of the National Academy of Sciences. 2007;104(49):19345-50.
219. Hensley CT, Faubert B, Yuan Q, Lev-Cohain N, Jin E, Kim J, et al. Metabolic heterogeneity in human lung tumors. Cell. 2016;164(4):681-94.
220. Ryan DG, Murphy MP, Frezza C, Prag HA, Chouchani ET, O'Neill LA, et al. Coupling Krebs cycle metabolites to signalling in immunity and cancer. Nature metabolism. 2019;1(1):16-33.
221. Jang C, Chen L, Rabinowitz JD. Metabolomics and isotope tracing. Cell. 2018;173(4):822-37.
222. Antoniewicz MR. A guide to 13 C metabolic flux analysis for the cancer biologist. Experimental & molecular medicine. 2018;50(4):1-13.
223. El-Science. How Does a GC-MS Work? [Internet]. 2016 [Available from: https://elscience.co.uk/our-lab/gcms-work.
224. Faubert B, Li KY, Cai L, Hensley CT, Kim J, Zacharias LG, et al. Lactate Metabolism in Human Lung Tumors. Cell. 2017;171(2):358-71 e9.
225. Maher EA, Marin-Valencia I, Bachoo RM, Mashimo T, Raisanen J, Hatanpaa KJ, et al. Metabolism of [U-13C]glucose in Human Brain Tumors In Vivo. NMR Biomed. 2012;25(11):1234-44.
226. Choudhury FK, Hackman GL, Lodi A, Tiziani S. Stable Isotope Tracing Metabolomics to Investigate the Metabolic Activity of Bioactive Compounds for Cancer Prevention and Treatment. Cancers. 2020;12(8):2147.
227. Lussey-Lepoutre C, Hollinshead KE, Ludwig C, Menara M, Morin A, Castro-Vega L-J, et al. Loss of succinate dehydrogenase activity results in dependency on pyruvate carboxylation for cellular anabolism. Nature communications. 2015;6:8784.
228. Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, et al. A roadmap for interpreting 13C metabolite labeling patterns from cells. Current opinion in biotechnology. 2015;34:189-201.
229. Rogatzki MJ, Ferguson BS, Goodwin ML, Gladden LB. Lactate is always the end product of glycolysis. Frontiers in neuroscience. 2015;9:22.
230. Rasola A, Neckers L, Picard D. Mitochondrial oxidative phosphorylation TRAP (1) ped in tumor cells. Trends in cell biology. 2014;24(8):455-63.
231. van Diepen JA, Robben JH, Hooiveld GJ, Carmone C, Alsady M, Boutens L, et al. SUCNR1-mediated chemotaxis of macrophages aggravates obesity-induced inflammation and diabetes. Diabetologia. 2017;60(7):1304-13.
232. Benit P, Letouze E, Rak M, Aubry L, Burnichon N, Favier J, et al. Unsuspected task for an old team: succinate, fumarate and other Krebs cycle acids in metabolic remodeling. Biochimica et Biophysica Acta (BBA)-Bioenergetics. 2014;1837(8):1330-7.

233. Mu X, Zhao T, Xu C, Shi W, Geng B, Shen J, et al. Oncometabolite succinate promotes angiogenesis by upregulating VEGF expression through GPR91-mediated STAT3 and ERK activation. Oncotarget. 2017;8(8):13174.
234. Sullivan LB, Gui DY, Hosios AM, Bush LN, Freinkman E, Vander Heiden MG. Supporting aspartate biosynthesis is an essential function of respiration in proliferating cells. Cell. 2015;162(3):552-63.
235. Sullivan LB, Luengo A, Danai LV, Bush LN, Diehl FF, Hosios AM, et al. Aspartate is an endogenous metabolic limitation for tumour growth. Nature cell biology. 2018;20(7):782-8.
236. Garcia-Bermudez J, Baudrier L, La K, Zhu XG, Fidelin J, Sviderskiy VO, et al. Aspartate is a limiting metabolite for cancer cell proliferation under hypoxia and in tumours. Nature cell biology. 2018;20(7):775-81.
237. Hay N. Reprogramming glucose metabolism in cancer: can it be exploited for cancer therapy? Nature Reviews Cancer. 2016;16(10):635.
238. Hui S, Ghergurovich JM, Morscher RJ, Jang C, Teng X, Lu W, et al. Glucose feeds the TCA cycle via circulating lactate. Nature. 2017;551(7678):115-8.
239. Betz MJ, Slawik M, Lidell ME, Osswald A, Heglind M, Nilsson D, et al. Presence of brown adipocytes in retroperitoneal fat from patients with benign adrenal tumors: relationship with outdoor temperature. The Journal of Clinical Endocrinology & Metabolism. 2013;98(10):4097-104.
240. Olsen JM, Sato M, Dallner OS, Sandström AL, Pisani DF, Chambard J-C, et al. Glucose uptake in brown fat cells is dependent on mTOR complex 2–promoted GLUT1 translocation. Journal of Cell Biology. 2014;207(3):365-74.
241. Hasan O, Nguyen M, Son H, Puri S. Role of FDG PET/CT in detecting unsuspected brain metastases during routine staging work up in oncology practice and its impact on patient management. Journal of Nuclear Medicine. 2014;55(supplement 1):1303-.
242. Tateishi U, Yamaguchi U, Seki K, Terauchi T, Arai Y, Hasegawa T. Glut-1 expression and enhanced glucose metabolism are associated with tumour grade in bone and soft tissue sarcomas: a prospective evaluation by [18 F] fluorodeoxyglucose positron emission tomography. European journal of nuclear medicine and molecular imaging. 2006;33(6):683.
243. Amelio I, Cutruzzolá F, Antonov A, Agostini M, Melino G. Serine and glycine metabolism in cancer. Trends in biochemical sciences. 2014;39(4):191-8.
244. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014;343(6166):84-7.
245. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annual review of biochemistry. 2010;79:181-211.
246. Depmap portal. Data downloads [Internet]. [Available from: https://depmap.org/portal/.
247. Wang B, Wang M, Zhang W, Xiao T, Chen C-H, Wu A, et al. Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. Nature protocols. 2019;14(3):756-80.
248. Yamauchi T, Masuda T, Canver MC, Seiler M, Semba Y, Shboul M, et al. Genome-wide CRISPR-Cas9 screen identifies leukemia-specific dependence on a pre-mRNA metabolic pathway regulated by DCPS. Cancer Cell. 2018;33(3):386-400. e5.
249. Online Mendelian Inheritance in Man. An Online Catalog of Human Genes and Genetic Disorders [Internet]. 2019 [Available from: https://www.omim.org.

250. Cai K, Tonelli M, Frederick RO, Markley JL. Human mitochondrial ferredoxin 1 (FDX1) and ferredoxin 2 (FDX2) both bind cysteine desulfurase and donate electrons for iron–sulfur cluster biosynthesis. Biochemistry. 2017;56(3):487-99.
251. Han W, Xin Z, Zhao Z, Bao W, Lin X, Yin B, et al. RNA-binding protein PCBP2 modulates glioma growth by regulating FHL3. The Journal of clinical investigation. 2013;123(5):2103-18.
252. Di X, Jin X, Ma H, Wang R, Cong S, Tian C, et al. The oncogene IARS2 promotes non-small cell lung cancer tumorigenesis by activating the AKT/MTOR pathway. Frontiers in Oncology. 2019;9:393.
253. Ruan C, Ouyang X, Liu H, Li S, Jin J, Tang W, et al. Sin1-mediated mTOR signaling in cell growth, metabolism and immune response. National Science Review. 2019;6(6):1149-62.
254. Ghosh D, Srivastava GP, Xu D, Schulz LC, Roberts RM. A link between SIN1 (MAPKAP1) and poly (rC) binding protein 2 (PCBP2) in counteracting environmental stress. Proceedings of the National Academy of Sciences. 2008;105(33):11673-8.
255. Irminger-Finger I, Ratajska M, Pilyugin M. New concepts on BARD1: Regulator of BRCA pathways and beyond. The international journal of biochemistry & cell biology. 2016;72:1-17.
256. Comprehensive resource of mammalian protein complexes. [Internet]. [Available from: https://mips.helmholtz-muenchen.de/corum/.
257. University of Texas. Tumour Supressor Gene Database [Internet]. 2020 [Available from: https://bioinfo.uth.edu/TSGene/index.html?csrt=7365306011986851951.
258. Dallosso A, Øster B, Greenhough A, Thorsen K, Curry T, Owen C, et al. Long-range epigenetic silencing of chromosome 5q31 protocadherins is involved in early and late stages of colorectal tumorigenesis through modulation of oncogenic pathways. Oncogene. 2012;31(40):4409-19.
259. Bikle DD. Vitamin D metabolism, mechanism of action, and clinical applications. Chemistry & biology. 2014;21(3):319-29.
260. Mughal AA, Grieg Z, Skjellegrind H, Fayzullin A, Lamkhannat M, Joel M, et al. Knockdown of NAT12/NAA30 reduces tumorigenic features of glioblastoma-initiating cells. Molecular cancer. 2015;14(1):160.
261. Chemaly ER, Troncone L, Lebeche D. SERCA control of cell death and survival. Cell Calcium. 2018;69:46-61.
262. Zhuo M, Zhu C, Sun J, Weis WI, Sun Z. The β-catenin binding protein ICAT modulates androgen receptor activity. Molecular endocrinology. 2011;25(10):1677-88.
263. Mutter AC, Tyryshkin AM, Campbell IJ, Poudel S, Bennett GN, Silberg JJ, et al. De novo design of symmetric ferredoxins that shuttle electrons in vivo. Proceedings of the National Academy of Sciences. 2019;116(29):14557-62.
264. Sheftel AD, Stehling O, Pierik AJ, Elsässer H-P, Mühlenhoff U, Webert H, et al. Humans possess two mitochondrial ferredoxins, Fdx1 and Fdx2, with distinct roles in steroidogenesis, heme, and Fe/S cluster biosynthesis. Proceedings of the National Academy of Sciences. 2010;107(26):11775-80.
265. Frey AG, Nandal A, Park JH, Smith PM, Yabe T, Ryu M-S, et al. Iron chaperones PCBP1 and PCBP2 mediate the metallation of the dinuclear iron enzyme deoxyhypusine hydroxylase. Proceedings of the National Academy of Sciences. 2014;111(22):8031-6.
266. Jung M, Mertens C, Tomat E, Brüne B. Iron as a central player and promising target in cancer progression. International Journal of Molecular Sciences. 2019;20(2):273.

267. Yu Y, Wong J, Lovejoy DB, Kalinowski DS, Richardson DR. Chelators at the cancer coalface: desferrioxamine to Triapine and beyond. Clinical Cancer Research. 2006;12(23):6876-83.
268. Zhang X, Fryknäs M, Hernlund E, Fayad W, De Milito A, Olofsson MH, et al. Induction of mitochondrial dysfunction as a strategy for targeting tumour cells in metabolically compromised microenvironments. Nature communications. 2014;5(1):1-14.
269. Frias MA, Thoreen CC, Jaffe JD, Schroder W, Sculley T, Carr SA, et al. mSin1 is necessary for Akt/PKB phosphorylation, and its isoforms define three distinct mTORC2s. Current Biology. 2006;16(18):1865-70.
270. Sarbassov DD, Guertin DA, Ali SM, Sabatini DM. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. Science. 2005;307(5712):1098-101.
271. Li S, Oh Y-T, Yue P, Khuri FR, Sun S-Y. Inhibition of mTOR complex 2 induces GSK3/FBXW7-dependent degradation of sterol regulatory element-binding protein 1 (SREBP1) and suppresses lipogenesis in cancer cells. Oncogene. 2016;35(5):642-50.
272. Starheim KK, Gromyko D, Evjenth R, Ryningen A, Varhaug JE, Lillehaug JR, et al. Knockdown of human Nα-terminal acetyltransferase complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization. Molecular and cellular biology. 2009;29(13):3569-81.
273. Benavides-Serrato A, Lee J, Holmes B, Landon KA, Bashir T, Jung ME, et al. Specific blockade of Rictor-mTOR association inhibits mTORC2 activity and is cytotoxic in glioblastoma. PLOS one. 2017;12(4):e0176599.
274. Valenta T, Hausmann G, Basler K. The many faces and functions of β-catenin. The EMBO journal. 2012;31(12):2714-36.
275. Kahn M. Can we safely target the WNT pathway? Nature reviews Drug discovery. 2014;13(7):513-32.
276. Liu J, Pan S, Hsieh MH, Ng N, Sun F, Wang T, et al. Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. Proceedings of the National Academy of Sciences. 2013;110(50):20224-9.
277. Jung Y-S, Park J-I. Wnt signaling in cancer: therapeutic targeting of Wnt signaling beyond β-catenin and the destruction complex. Experimental & Molecular Medicine. 2020:1-9.
278. Li J, Luco A-L, Ochietti B, Fadhil I, Camirand A, Reinhardt TA, et al. Tumoral vitamin D synthesis by CYP27B1 1-α-hydroxylase delays mammary tumor progression in the PyMT-MMTV mouse model and its action involves NF-κB modulation. Endocrinology. 2016;157(6):2204-16.
279. Jeon S-M, Shin E-A. Exploring vitamin D metabolism and function in cancer. Experimental & molecular medicine. 2018;50(4):1-14.
280. Clinckspoor I, Hauben E, Verlinden L, Van den Bruel A, Vanwalleghem L, Vander Poorten V, et al. Altered expression of key players in vitamin D metabolism and signaling in malignant and benign thyroid tumors. Journal of Histochemistry & Cytochemistry. 2012;60(7):502-11.
281. Lopes N, Sousa B, Martins D, Gomes M, Vieira D, Veronese LA, et al. Alterations in Vitamin D signalling and metabolic pathways in breast cancer progression: a study of VDR, CYP27B1 and CYP24A1 expression in benign and malignant breast lesions Vitamin D pathways unbalanced in breast lesions. BMC cancer. 2010;10(1):483.

282. Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang C-Z, Ben-David U, et al. Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. Cancer discovery. 2016;6(8):914-29.
283. Lopez ER, Regulla K, Pani MA, Krause M, Usadel K-H, Badenhoop K. CYP27B1 polymorphisms variants are associated with type 1 diabetes mellitus in Germans. The Journal of steroid biochemistry and molecular biology. 2004;89:155-7.
284. Bosse KR, Diskin SJ, Cole KA, Wood AC, Schnepp RW, Norris G, et al. Common variation at BARD1 results in the expression of an oncogenic isoform that influences neuroblastoma susceptibility and oncogenicity. Cancer research. 2012;72(8):2068-78.
285. Yau EH, Rana TM. Next-generation sequencing of genome-wide CRISPR screens. Next Generation Sequencing: Springer; 2018. p. 203-16.
286. Kampmann M, Bassik MC, Weissman JS. Functional genomics platform for pooled screening and generation of mammalian genetic interaction maps. Nature protocols. 2014;9(8):1825.
287. Klein A, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Weitz; David, A.; Kirschner; Marc, W. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161:1187-201.
288. Lun A. Overcoming systematic errors caused by log-transformation of normalized single-cell RNA sequencing data. BioRxiv. 2018:404962.
289. 10xGenomics. How is “Log2foldchange” calculated [Internet]. 2020 [Available from: https://kb.10xgenomics.com/hc/en-us/articles/360007388751-How-is-Log2-fold-change-calculated-.
290. Kolset SO, Pejler G. Serglycin: a structural and functional chameleon with wide impact on immune cells. The journal of Immunology. 2011;187(10):4927-33.
291. Zhang Z, Deng Y, Zheng G, Jia X, Xiong Y, Luo K, et al. SRGN-TGFβ2 regulatory loop confers invasion and metastasis in triple-negative breast cancer. Oncogenesis. 2017;6(7):e360-e.
292. Rutkowska A, Preuss I, Gessier F, Sailer AW, Dev KK. EBI2 regulates intracellular signaling and migration in human astrocyte. Glia. 2015;63(2):341-51.
293. Barata JT, Durum SK, Seddon B. Flip the coin: IL-7 and IL-7R in health and disease. Nature Immunology. 2019;20(12):1584-93.
294. Laurent GJ, Shapiro SD. Encyclopedia of respiratory medicine: Academic Press; 2006.
295. The Human Protein Atlas. CXCR4 [Internet]. [Available from: https://www.proteinatlas.org/ENSG00000121966-CXCR4.
296. Lippitz BE. Cytokine patterns in patients with cancer: a systematic review. The lancet oncology. 2013;14(6):e218-e28.
297. Zhao H, Guo L, Zhao H, Zhao J, Weng H, Zhao B. CXCR4 over-expression and survival in cancer: a system review and meta-analysis. Oncotarget. 2015;6(7):5022.
298. Chen BJ, Chapuy B, Ouyang J, Sun HH, Roemer MG, Xu ML, et al. PD-L1 expression is characteristic of a subset of aggressive B-cell lymphomas and virus-associated malignancies. Clin Cancer Res. 2013;19(13):3462-73.
299. Campbell JJ, Murphy KE, Kunkel EJ, Brightling CE, Soler D, Shen Z, et al. CCR7 expression and memory T cell diversity in humans. The Journal of Immunology. 2001;166(2):877-84.
300. Pavlasova G, Mraz M. The regulation and function of CD20: an “enigma” of B-cell biology and targeted therapy. haematologica. 2020;105(6):1494-506.

301. Hardison RC. Evolution of hemoglobin and its genes. Cold Spring Harbor perspectives in medicine. 2012;2(12):a011627.
302. Chung J, Anderson SA, Gwynn B, Deck KM, Chen MJ, Langer NB, et al. Iron regulatory protein-1 protects against mitoferrin-1-deficient porphyria. Journal of Biological Chemistry. 2014;289(11):7835-43.
303. GeneCards. GYPC [(Internet)]. 2017 [Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=GYPC.
304. Argraves WS, Greene LM, Cooley MA, Gallagher WM. Fibulins: physiological and disease perspectives. EMBO reports. 2003;4(12):1127-31.
305. Moll F, Katsaros D, Lazennec G, Hellio N, Roger P, Giacalone P-L, et al. Estrogen induction and overexpression of fibulin-1C mRNA in ovarian cancer cells. Oncogene. 2002;21(7):1097-107.
306. Forti S, Scanlan MJ, Invernizzi A, Castiglioni F, Pupa S, Agresti R, et al. Identification of breast cancer-restricted antigens by antibody screening of SKBR3 cDNA library using a preselected patient's serum. Breast cancer research and treatment. 2002;73(3):245-56.
307. Nakao Y, Mitsuyasu T, Kawano S, Nakamura N, Kanda S, Nakamura S. Fibroblast growth factors 7 and 10 are involved in ameloblastoma proliferation via the mitogen-activated protein kinase pathway. International journal of oncology. 2013;43(5):1377-84.
308. Jayachandran A, Anaka M, Prithviraj P, Hudson C, McKeown SJ, Lo P-H, et al. Thrombospondin 1 promotes an aggressive phenotype through epithelial-to-mesenchymal transition in human melanoma. Oncotarget. 2014;5(14):5782.
309. Sauzay C, Voutetakis K, Chatziioannou A, Chevet E, Avril T. CD90/Thy-1, a cancer-associated cell surface signaling molecule. Frontiers in cell and developmental biology. 2019;7:66.
310. Kumar A, Bhanja A, Bhattacharyya J, Jaganathan BG. Multiple roles of CD90 in cancer. Tumor Biology. 2016;37(9):11611-22.
311. Koumas L, Phipps RP. Differential COX localization and PG release in Thy-1+ and Thy-1− human female reproductive tract fibroblasts. American Journal of Physiology-Cell Physiology. 2002;283(2):C599-C608.
312. Chen W-C, Chang Y-S, Hsu H-P, Yen M-C, Huang H-L, Cho C-Y, et al. Therapeutics targeting CD90-integrin-AMPK-CD133 signal axis in liver cancer. Oncotarget. 2015;6(40):42923.
313. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Molecular & Cellular Proteomics. 2014;13(2):397-406.
314. Paul JD, Coulombe KL, Toth PT, Zhang Y, Marsboom G, Bindokas VP, et al. SLIT3–ROBO4 activation promotes vascular network formation in human engineered tissue and angiogenesis in vivo. Journal of molecular and cellular cardiology. 2013;64:124-31.
315. Svensson KJ, Long JZ, Jedrychowski MP, Cohen P, Lo JC, Serag S, et al. A secreted Slit2 fragment regulates adipose tissue thermogenesis and metabolic function. Cell metabolism. 2016;23(3):454-66.
316. Kundel DW, Stromquist E, Greene AL, Zhdankin O, Regal RR, Rose-Hellekant TA. Molecular characterizations of Nop16 in murine mammary tumors with varying levels of c-Myc. Transgenic research. 2012;21(2):393-406.
317. Tansey WP. Mammalian MYC proteins and cancer. New Journal of Science. 2014;2014.

318. Mai S, Hanley-Hyde J, Fluri M. c-Myc overexpression associated DHFR gene amplification in hamster, rat, mouse and human cell lines. Oncogene. 1996;12(2):277.
319. Sirvent N, Coindre J-M, Maire G, Hostein I, Keslair F, Guillou L, et al. Detection of MDM2-CDK4 amplification by fluorescence in situ hybridization in 200 paraffin-embedded tumor samples: utility in diagnosing adipocytic lesions and comparison with immunohistochemistry and real-time PCR. The American journal of surgical pathology. 2007;31(10):1476-89.
320. Zhou W-Y, Zheng H, Du X-L, Yang J-L. Characterization of FGFR signaling pathway as therapeutic targets for sarcoma patients. Cancer biology & medicine. 2016;13(2):260.
321. Moseti D, Regassa A, Kim W-K. Molecular regulation of adipogenesis and potential anti-adipogenic bioactive molecules. International journal of molecular sciences. 2016;17(1):124.
322. Darlington GJ, Ross SE, MacDougald OA. The role of C/EBP genes in adipocyte differentiation. Journal of Biological Chemistry. 1998;273(46):30057-60.
323. Brey CW, Nelder MP, Hailemariam T, Gaugler R, Hashmi S. Krüppel-like family of transcription factors: an emerging new frontier in fat biology. International journal of biological sciences. 2009;5(6):622.
324. Wu Z, Wang S. Role of kruppel-like transcription factors in adipogenesis. Developmental biology. 2013;373(2):235-43.
325. Birsoy K, Chen Z, Friedman J. Transcriptional regulation of adipogenesis by KLF4. Cell metabolism. 2008;7(4):339-47.
326. MacDougald OA, Lane MD. Transcriptional regulation of gene expression during adipocyte differentiation. Annual review of biochemistry. 1995;64(1):345-73.
327. Li Z, Gilbert JA, Zhang Y, Zhang M, Qiu Q, Ramanujan K, et al. An HMGA2-IGF2BP2 axis regulates myoblast proliferation and myogenesis. Developmental cell. 2012;23(6):1176-88.
328. Wang Y, Sul HS. Pref-1 regulates mesenchymal cell commitment and differentiation through Sox9. Cell metabolism. 2009;9(3):287-302.
329. Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, et al. Assessing tumor infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the International Immuno-Oncology Biomarkers Working Group: Part 2: TILs in melanoma, gastrointestinal tract carcinomas, non-small cell lung carcinoma and mesothelioma, endometrial and ovarian carcinomas, squamous cell carcinoma of the head and neck, genitourinary carcinomas, and primary brain tumors. Advances in anatomic pathology. 2017;24(6):311.
330. Cibrián D, Sánchez-Madrid F. CD69: from activation marker to metabolic gatekeeper. European journal of immunology. 2017;47(6):946-53.
331. Tseng WW, Malu S, Zhang M, Chen J, Sim GC, Wei W, et al. Analysis of the intratumoral adaptive immune response in well differentiated and dedifferentiated retroperitoneal liposarcoma. Sarcoma. 2015;2015.
332. Tanaka A, Sakaguchi S. Targeting Treg cells in cancer immunotherapy. European journal of immunology. 2019;49(8):1140-6.
333. Bai X, Yi M, Jiao Y, Chu Q, Wu K. Blocking TGF-β signaling to enhance the efficacy of immune checkpoint inhibitor. OncoTargets and therapy. 2019;12:9527.
334. Kerkelä E, Lahtela J, Larjo A, Impola U, Mäenpää L, Mattila P. Exploring transcriptomic landscapes in red blood cells, in their extracellular vesicles and on a single-cell level. 2020.

335. Bastos RN, Volloch Z, Aviv H. Messenger RNA population analysis during erythroid differentiation: a kinetical approach. Journal of molecular biology. 1977;110(2):191-203.
336. Kunicki TJ. Platelet membrane glycoproteins and their function: an overview. Blut. 1989;59(1):30-4.
337. Yun S-H, Sim E-H, Goh R-Y, Park J-I, Han J-Y. Platelet activation: the mechanisms and potential biomarkers. BioMed research international. 2016;2016.
338. Kim HS, Won YJ, Shim JH, Kim HJ, Kim J, Hong HN, et al. Morphological characteristics of vasculogenic mimicry and its correlation with EphA2 expression in gastric adenocarcinoma. Scientific reports. 2019;9(1):1-13.
339. Hashimoto T, Shibasaki F. Hypoxia-inducible factor as an angiogenic master switch. Frontiers in pediatrics. 2015;3:33.
340. Riera-Sans L, Behrens A. Regulation of αβ/γδ T cell development by the activator protein 1 transcription factor c-Jun. The Journal of Immunology. 2007;178(9):5690-700.
341. Lynn RC, Weber EW, Sotillo E, Gennert D, Xu P, Good Z, et al. c-Jun overexpression in CAR T cells induces exhaustion resistance. Nature. 2019;576(7786):293-300.
342. The Human Protein Atlas. MDM2 [Internet]. 209 [Available from: https://www.proteinatlas.org/ENSG00000135679-MDM2/tissue.
343. Atlas THP. CDK4 [Internet]. [Available from: https://www.proteinatlas.org/ENSG00000135446-CDK4.
344. Mohseny AB, Hogendoorn PC. Concise review: mesenchymal tumors: when stem cells go mad. Stem cells. 2011;29(3):397-403.
345. Willems SM, Mohseny AB, Balog C, Sewrajsing R, Briaire-de Bruijn IH, Knijnenburg J, et al. Cellular/intramuscular myxoma and grade I myxofibrosarcoma are characterized by distinct genetic alterations and specific composition of their extracellular matrix. J Cell Mol Med. 2009;13(7):1291-301.
346. Matushansky I, Hernando E, Socci ND, Matos T, Mills J, Edgar MA, et al. A developmental model of sarcomagenesis defines a differentiation-based classification for liposarcomas. The American journal of pathology. 2008;172(4):1069-80.
347. Huang T, Wang L, Liu D, Li P, Xiong H, Zhuang L, et al. FGF7/FGFR2 signal promotes invasion and migration in human gastric cancer through upregulation of thrombospondin-1. International Journal of Oncology. 2017;50(5):1501-12.
348. Paul C, Sardet C, Fabbrizio E. The Wnt-target gene Dlk-1 is regulated by the Prmt5-associated factor Copr5 during adipogenic conversion. Biology open. 2015;4(3):312-6.
349. Skubitz KM, Cheng EY, Clohisy DR, Thompson RC, Skubitz APN. Differential gene expression in liposarcoma, lipoma, and adipose tissue. Cancer investigation. 2005;23(2):101-14.
350. Kypta RM, Waxman J. Wnt/β-catenin signalling in prostate cancer. Nature Reviews Urology. 2012;9(8):418.
351. Khalil DN, Smith EL, Brentjens RJ, Wolchok JD. The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. Nature reviews Clinical oncology. 2016;13(5):273-90.
352. Kanwal M, Ding X-J, Ma Z-H, Li L-W, Wang P, Chen Y, et al. Characterization of germline mutations in familial lung cancer from the Chinese population. Gene. 2018;641:94-104.
353. Chantrel-Groussard K, Gaboriau F, Pasdeloup N, Havouis R, Nick H, Pierre J-L, et al. The new orally active iron chelator ICL670A exhibits a higher antiproliferative effect in

human hepatocyte cultures than O-trensox. European journal of pharmacology. 2006;541(3):129-37.
354. Becton DL, Bryles P. Deferoxamine inhibition of human neuroblastoma viability and proliferation. Cancer Research. 1988;48(24 Part 1):7189-92.
355. Zhong Z, Sepramaniam S, Chew XH, Wood K, Lee MA, Madan B, et al. PORCN inhibition synergizes with PI3K/mTOR inhibition in Wnt-addicted cancers. Oncogene. 2019;38(40):6662-77.

Appendix I: SHARPS patient information sheet and consent form

Patient information sheet V1.6

PARTICIPANT INFORMATION SHEET
SARCOMA STUDY

The following information describes a study which we are inviting you to take part in, the results of which will be used in part towards a PhD qualification for the lead researcher. This study has been reviewed and accepted by the Berkshire B Research Ethics Committee.

What is the study?

This is a study to better understand the way liposarcoma cells create and use energy to grow and divide. We have looked at the entire genetic code of these cancers using a technology called whole genome sequencing. We have found a fault whereby part of one gene (called SDHA) sticks to another gene (known as a gene fusion), causing SDHA to stop working. The SDHA gene is part of a process by which we produce energy, and faults in this gene have been shown to cause other types of cancer. We are interested in looking at the consequences of the SDHA fusion gene on the different pathways the cell uses to create energy.

Our aim is to give patients, going for surgery to remove a liposarcoma, a form of 'labelled' glucose which will be used by cells as an energy source. Once the tumour is removed, we will use samples from it to see which abnormal methods the cancer is using to create energy. This is through a process called mass spectroscopy which relies on the 'labelled' glucose infusion to do so. These abnormal pathways are targets for drug therapies, limiting the ability of the cancer cell to grow and divide. If we can achieve this, any surgery would potentially have a higher success rate and a lower risk of complications.

This study will be funded by the University of Birmingham and aims to recruit patients from the Midlands Abdominal and Retroperitoneal Sarcoma Unit (MARSU) ran at the Queen Elizabeth Hospital, Birmingham.

Why are you doing it?

The Queen Elizabeth Hospital is home to one of the largest sarcoma units in the country allowing us to research rare types of sarcoma. We are carrying out this study with the University of Birmingham, who have the facilities to look for the 'labelled' glucose using a technique called mass spectroscopy.

Why am I being invited to take part in this study?

We are inviting you because you have been diagnosed with a liposarcoma, the tumour type this study hopes to better understand. After discussion with your surgeon, you have decided to undergo surgery to remove this tumour. You are therefore eligible to take part in the study.

What would I have to do?

We need your permission to take part in the study. This will mean giving you an infusion of the labelled glucose into a vein when you are having the operation. Once the tumour is taken out, we will take biopsies from it to use for testing. The rest of the tumour will be examined by a pathologist as is standard practice. Medical records and data may be accessed by researchers, regulatory authorities or the NHS trust. There are no extra appointments. The tissue we use will be stored in the onsite biobank. The experiments we perform will use up all the samples.

Will being in the study change my treatment at all?

Whether you take part in the study or not will not influence the treatment you receive in any way. You are free to stop being in the study at any point and your treatment will remain exactly the same.

Potential Harm or Benefit to Participants

There are no harmful effects of using a labelled glucose infusion or any benefits. It simply acts as a tracer for us to identify in the tumour sample later on. The remaining glucose will be passed in urine or broken down by your body. The anaesthetist will be responsible for the glucose infusion during the operation and may choose to stop it if deemed necessary. We only take a very small biopsy of the tumour sample so this will not interfere with the final pathology report.

What happens if I withdraw from the study?

You are entitled to withdraw from the study at any point. Any samples that have been taken at that point are stored in the onsite Biobank (HBRC) and will be destroyed in accordance with the Human Tissue Authority's Code.

How will my personal information be kept confidential?

University Hospitals Birmingham NHS Foundation Trust (UHBfT) is the sponsor for this study based in England. We will be using information from you and your medical records in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. UHBfT will keep identifiable information about you for 5 years after the study has finished.

Your rights to access, change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally identifiable information possible. You can find out more about how we use your information at www.research.uhb.nhs.uk.

Researchers at UHBfT will use your name and contact details to contact you about the research study, and make sure that relevant information about the study is recorded for your care, and to oversee the quality of the study. Administrators from UHBfT and regulatory organisations may look at your medical and research records to check the accuracy of the research study.

The biobank on site at the Queen Elizabeth will anonymise samples that leave the hospital grounds for analysis. Samples will be labelled with a randomly generated number that contain none of your personal information. The only identifiable information we use will be kept within the hospital, on a password controlled laptop, locked in the sarcoma department. This information cannot be accessed outside of the hospital.

Can I get the results of the study?

If you are interested in the results of the study a summary can be sent to you once all of the analyses are complete.

Funding

This study is funded by University of Birmingham and has been approved by the relevant ethics committee.

What do I need to do if I want to be in the study?

All you need to do to take part is to agree and sign the consent form. It is important to take the time to read this information and ask any questions you have. Once you have agreed to be in the study the information will be collected. You are within your rights to withdraw consent for the study at any point in your treatment.

Thank you very much for reading this information and we hope you will be part of the study. If you wish to contact our research team to discuss this study, please do not hesitate to call 0121 371 5884.

If you would like general, impartial advice on taking part in research studies we would suggest contacting the PALS office on 0121 371 2550

Mr. Desai (Consultant Surgeon)
Mr. Tyler (Research Registrar)

Patient Consent Form V 1.2

CONSENT FORM FOR RESEARCH STUDY

Investigating tumour metabolism in Well Differentiated Retroperitoneal Liposarcoma

Name of researchers: Mr. Robert Tyler, Dr. Andrew Beggs, Mr. Anant Desai

	Please initial to
• I confirm that I have read and understood the information sheet version ……Date…….	☐
• I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
• I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.	☐
• I understand that relevant sections of any of my medical notes and data collected during th study may be looked at by researchers, from regulatory authorities or from the NHS Trust, where it is relevant to my taking part in the research. I give permission for these individuals to have access to my records.	☐
	☐

- I agree to take part in the above research study and consent to my tumour samples being used for research.

Name of Patient	Date	Signature
Name of Person taking consent (if different from the researcher)	Date	Signature
Researcher	Date	Signature
Translator	Date	Signature

Appendix II

List of genes from Figure 6.5. Group 2 (Green) are strongly positively selected at Day 14 and weakly selected in the control group. Group 4 (Orange) are strongly negatively selected at Day 14 and weakly selected in the control group.

Group	Gene Name	Protein Function	Beta Score	FDR
2	PLN	Phospholamban - regulates the calcium pump in cardiac muscle cells	3.7826	0.885213
2	TMEM14B	Phosphorylates G1/S cell cycle proteins	2.6255	0.543729
2	NME8	Associated with ciliary function	2.3657	0.543729
2	SRD5A1	Steroid 5α-reductase, converts testosterone to dihydrotestosterone	2.1727	0.791138
2	PCDHGC3	Protocadherin gamma-C3 involved in cell adhesion	2.0199	0.839378
2	CHRNA7	Encodes nicotinic acetylcholine receptor	1.9819	0.923352
2	NBPF6	Neuroblastoma breaking point family 6 - function unknown	1.9561	0.926984
2	SYTL3	Peripheral membrane protein, involved in vesicular tracking	1.8449	0.685644
2	FAM50B	Retrotransposon	1.6163	0.81264
2	CEP78	Centrosomal protein required for ciliogenesis	1.3538	0.82268
4	CGB7	Beta 7 subunit of chorionic gonadotrophin	-6.623	0.572426
4	POTEM	Member of ankyrin domain involved in neural development	-6.6093	0.579857
4	KRTAP2-4	Hair keratin-associated proteins	-6.2704	0.660559
4	FGD6	Role not fully elucidated, may exchange bound GTP for free GTP	-4.8849	0.089286
4	CYP27B1	1-α-hydroxylase, converts calcifediol to calcitriol (active Vitamin D)	-4.3694	0.15709
4	HIST2H4B	Histone	-3.6029	0.983649
4	DEFB103B	Microbicidal and cytotoxic peptide secreted by neutrophils	-2.8425	0.909078
4	SAFB	DNA-binding protein with specificity for matrix attachment region DNA elements	-2.771	0.190294
4	NAA30	Catalytic subunit of N-terminal acetyltransferase complex	-2.7103	0.387305
4	RPL21	Ribosomal protein component of 60S subunit	-1.3517	0.65215