Identification of Subtype-specific Pathogenetic Aberrations in Soft Tissue Sarcoma utilising High-resolution Genomic Copy Number Analysis

BY

Abdulazeez Temitope Salawu

A thesis submitted to the Medical School, University of Sheffield
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Supervisor: Dr. Karen Sisley

Department of Oncology,
CR-UK/YCR Sheffield Cancer Research Centre
Medical School, University of Sheffield

May 2014

ACKNOWLEDGEMENTS

Dr. Karen Sisley, my supervisor has my deepest and most sincere gratitude for her constant support, advice and patience. Professor Malcolm Reed for continued belief in me, his motivation and support for all my endeavours. Drs Dave Hammond and Aliya Ul-Hassan for welcoming me into the scientific community, patiently showing me the ropes and for many hours of long discussion.

Ms Lynda Wyld for all her help with clinical aspects of the project and for the sound career advice. Dr Malee Fernando for always being there with immense support, advice and enthusiasm. Drs David Hughes and John Goepel for help with everything Pathology. All the patients approached who happily gave consent for the study. The Sarcoma Clinical Nurse Specialists and Sarcoma clinic staff for making me feel very welcome. All members of the Sheffield Sarcoma MDT, Operating Theatres, Histopathology Department and Surgical Secretaries.

Sarcoma Trust/Sarcoma UK and the University of Sheffield for providing the funding for this project. Professor Penella Woll, Dr Martin Nicklin and Dr Gordon Cooper for believing in me and for guidance in the right direction. All members (past and present) of the Rare Tumour Research Group who were helpful in more ways than they can imagine. The technical team of the Oncology Department who were extremely supportive and generous.

The best parents and siblings in the world for whom words alone cannot express my eternal gratitiude. My loving wife for her superhuman patience, for keeping me sane through the most difficult parts of the project and for the greatest gift I could ever hope for. My friends who made Sheffield feel like home and everyone else who I have not mentioned in person for lack of space, but have my appreciation and gratitude

Finally, I dedicate this work to the Creator who made everything possible.

ABSTRACT

Soft tissue sarcomas (STS) are a group of malignant tumours that arise in mesenchymal tissues. Their exact cause remains unknown but the strongest associations are inherited cancer susceptibility and radiation exposure. They are very diverse, comprising over 50 clinico-biologic subtypes based on phenotypic and genetic information. The majority of STS subtypes however still have no known specific diagnostic markers and fewer still possess therapeutic targets. This is in spite of pervasive chromosomal instability evidenced by multiple seemingly random somatic copy number aberrations (SCNA).

One aim of this study was the utilisation of microarray-based comparative genomic hybridisation (aCGH) for high resolution mapping of genome-wide SCNA in order to identify patterns that are specific for three STS subtypes - Leiomyosarcoma (LMS), Undifferentiated Pleomorphic Sarcoma (UPS) and Gastrointestinal Stromal Tumours (GIST). A total of seventy-four cases were analysed and frequent and/or subtype-specific SCNA with pathogenetic potential were identified using multiple statistical methods and biologic context correlation. Secondly, the study aimed to develop a suitable disease model for functional studies to confirm the pathogenetic implication of identified candidate genes. Primary tissue cultures were therefore set up from 47 tumours.

Over 20 candidate pathogenetic genes and molecular pathways were identified among the three STS subtypes with potential roles in cancer hallmark acquisition, genomic instability and mesenchymal lineage differentiation. Confluent cell cultures were established for more than half of primary tumours, including 7 stable long-term cultures (15% of cases). One candidate gene with a potential role in tumour progression, *TRIO* was frequently amplified in UPS. The pattern of amplification was confirmed using fluorescence in situ hybridisation (FISH) on primary tumour metaphases and its over-expression confirmed by Immunohistochemistry. Finally, preliminary studies using small molecule targeted inhibition of the *TRIO* gene product showed significant inhibition of *in vitro* proliferation and invasion of primary UPS cells.

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES

LIST OF TABLES

LIST OF ABBREVIATIONS

aCGH	Array-based Comparative Genomic Hybridisation
Angio	Angiosarcoma
ASPS	Alveolar Soft Part Sarcoma
ATCC	American Type Culture Collection
BAC	Bacterial Artificial Chromosomes
cBAC	Combined Bacterial Artificial Chromosomes
CEP	Chromosome Enumeration Probe
CIN	Chromosomal Instability
CNV	Copy Number Variation
COGcell	Children's Oncology Group Cell Culture And Xenograft Repository
DAPI	4', 6-Diamidino-2-Phenylindole
DAVID	Database For Annotation, Visualisation And Integrated Discovery
ddH_2O	Distilled De-Ionised Water
DDLPS	Dedifferentiated Liposarcoma
dH_2O	Distilled Water
DLRS	Derivative Log Ratio Spread
DMSO	Dimethyl sulphoxide
DoL	Degree Of Labelling
DSB	Double Strand Break
DSMZ	German Collection Of Microorganisms And Cell Cultures
EBV	Ebstein Barr Virus
ECM	Extracellular Matrix
EDTA	Ethylene Diamine Tetraacetic Acid
EGFR	Epidermal Growth Factor Receptor
EMCS	Extraskeletal Myxoid Chondrosarcoma
FASST2	Fast Adaptive States Segmentation Technique 2
FDR	False Detection Rate
FF	Fresh Frozen
FFPE	Formalin Fixed Paraffin Embedded
FISH	Fluorescence In-Situ Hybridisation
FNCLCC	Fédération Nationale Des Centres De Lutte Contre Le Cancer

LIST OF ABBREVIATIONS

GIST	Gastrointestinal Stromal Tumour
GISTIC	Genomic Identification Of Significant Targets In Cancer
H+E	Haematoxylin And Eosin
HHV8	Human Herpes Virus 8
HIER	Heat-Induced Epitope Retrieval
HMM	Hidden Markov Model
HSR	Homogenously Staining Region
IMT	Inflammatory Myofibroblastic Tumour
JCRB	Japanese Collection Of Research Bioresources
Kb	Kilobase
KEGG	Kyoto Encyclopaedia Of Genes And Genomes
KSHV	Kaposi Sarcoma-Related Herpes Virus
LB	Luria Bertani
LGMFS	Low Grade Myofibroblastic Sarcoma
LMS	Leiomyosarcoma
LOH	Loss Of Heterozygosity
MAPK	Mitogen-Activated Protein Kinase
Mb	Megabase
MCR	Minimal Common Region
MFS	Myxofibrosarcoma
MIN	Microsatellite Instability
MLPS	Myxoid Liposarcoma
MMR	Mismatch Repair
MPNST	Malignant Peripheral Nerve Sheath Tumour
MTT	3- (4, 5-Dimethylthiazol-2-Yl)-2, 5-Diphenyltetrazolium Bromide
NBF	Neutral Buffered Formalin
NCI	National Cancer Institute
NER	Nucleotide Exchange Repair
NHEJ	Non Homologous End-Joining
NIN	Nucleotide Instability
PBS	Phosphate Buffered Saline

LIST OF ABBREVIATIONS

PDGFRB	Platelet-Derived Growth Factor Receptor B
PIK3	Phospho-Inositol-3-Kinase
PLPS	Pleomorphic Liposarcoma
PrEST	Protein Epitope Signature Tag
PRMS	Pleomorphic Rhabdomyosarcoma
RhoGEF	Rho Guanine Nucleotide-GTPase Exchange Factor
RTRG	Rare Tumour Research Group
SCNA	Somatic Copy Number Aberration
SFT	Solitary Fibrous Tumour
SLE	Systemic Lupus Erythematosus
SMMHC	Smooth Muscle Myosin Heavy Chain
SSC	Saline Sodium Citrate
STAC	Significance Testing Of Aberrant Copy Number
STR	Short Tandem Repeat
STS	Soft Tissue Sarcoma
STS	Soft Tissue Sarcoma
SUMO	Small Ubiquitin-Like Modifier
TAE	Tris-Acetate-EDTA
ULS	Universal Linkage System
UPS	Undifferentiated Pleomorphic Sarcoma
WDLPS	Well-Differentiated Liposarcoma
WHO	World Health Organisation

CHAPTER ONE

INTRODUCTION

1.1 Soft Tissues of the Body and Sarcoma

Human body tissues may be divided into four main groups – epithelial tissues (parenchyma), which comprise specialised organ-specific functional cells; the mesenchyme, which provides functional and structural support for the epithelial tissues in various organ systems; the haemato-lymphoid, and the nervous systems (Knowles and Selby 2006). The mesenchyme is derived from the embryological mesoderm and neuro-ectoderm and includes connective tissue (fibroblasts and extracellular matrix), adipose tissue, bone, cartilage, muscle, lymphatic and blood vessels (Figure 1.1). With the exception of bone, these are referred to as the soft tissues of the body (Knowles and Selby 2006).

Sarcomas are described as malignant neoplasms that arise in these mesenchymal tissues (Knowles and Selby 2006). They constitute a genetically, pathologically and clinically diverse group of tumour with grave life- and/or limb-threatening consequences (Thway 2009). Since all organs have a specialised epithelium as well as supporting mesenchymal tissue, soft tissue sarcomas (STS) can in theory arise in virtually any organ of the body. They however occur most frequently in the extremities (up to 75%) followed by the trunk (10%) and retroperitoneum (10%) (Fletcher, Bridge et al. 2013, Francis, Dennis et al. 2013).

1.2 Epidemiology

Compared with carcinomas that characteristically arise in parenchymal tissues, soft tissue sarcomas are relatively rare and constitute only about 1% of adult cancers and 1% of all soft tissue neoplasms (Toro, Travis et al. 2006, Fletcher, Bridge et al. 2013, Francis, Dennis et al. 2013). However, up to 15% of these tumours occur in children and adolescents, with specific subtypes e.g. embryonal rhabdomyosarcoma, occurring almost exclusively in this age group (Thway 2009, Fletcher, Bridge et al. 2013).

The World Health Organization (WHO) estimates the worldwide annual incidence of STS at around 50 cases per million population with no evidence of significant geographical variation (Fletcher, Bridge et al. 2013). In the United Kingdom, annual incidence estimates are in keeping with worldwide figures with approximately 3300 new cases diagnosed in 2010 (Francis, Dennis et al. 2013). Like most other malignancies the frequency of STS increases with age, with incidence rising up to 230 cases per million among men aged >85 years. Overall, the median age at diagnosis is around 65 years with a slight male preponderance (Fletcher, Bridge et al. 2013). However, the gender distribution varies with age group. While the male-female ratio at age >85 is approximately 1.9:1 in the UK, there is a female preponderance among individuals aged 45 – 59 years attributed to an increased number of gynaecological tract sarcomas (Francis, Dennis et al. 2013).

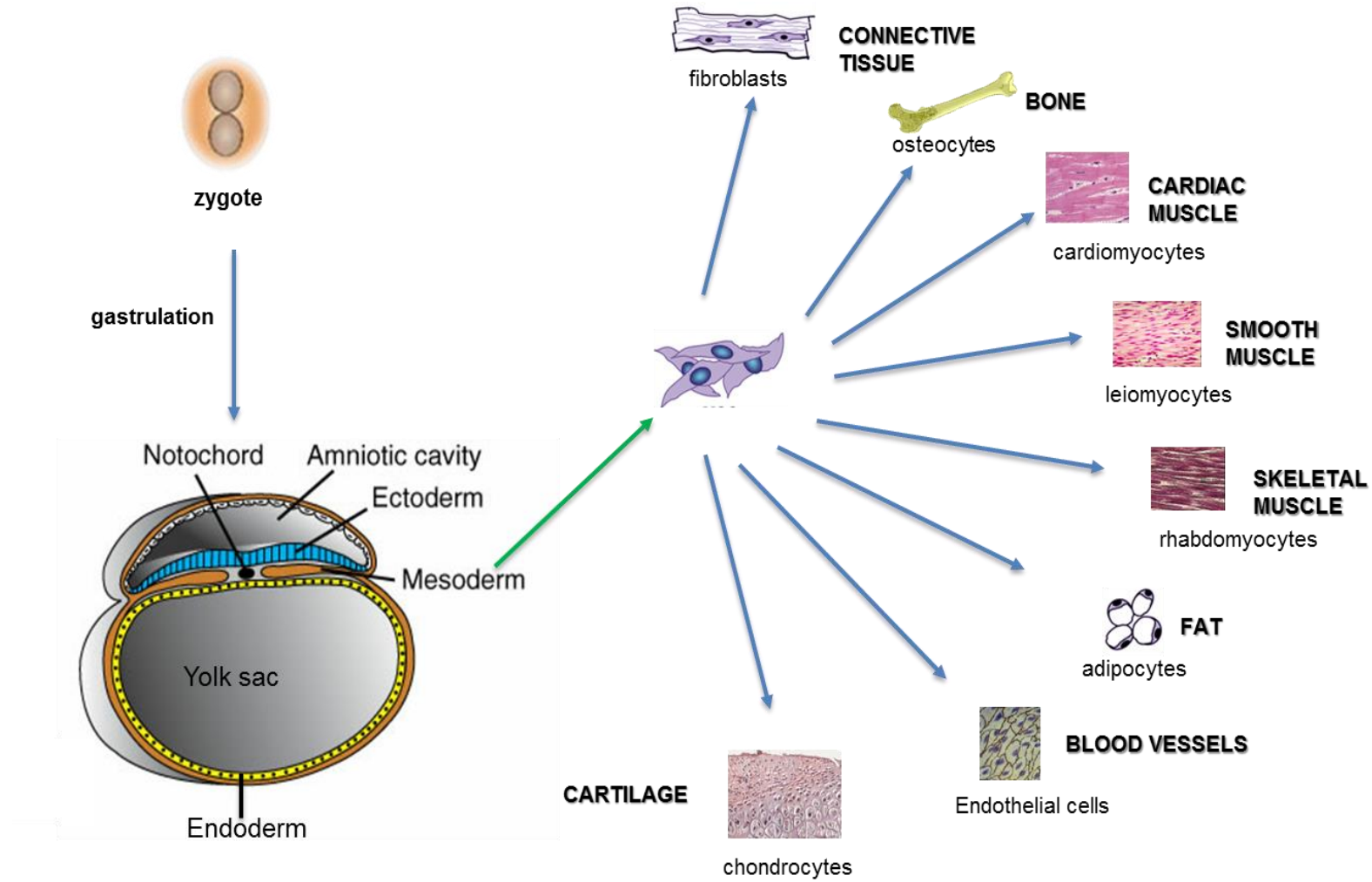

Figure 1.1: Illustration of the Embryological Origin of Mesenchymal Tissues

Mesenchymal stem cells derived from the mesoderm and neuro-ectoderm differentiate into various adult mesenchymal cell lineages. With the exception of bone, these are collectively referred to as soft tissues.

1.3 Aetiology and Predisposing Factors

Like most other cancers, the exact cause of STS is unknown. While most arise sporadically, several inheritable cancer-predisposing syndromes, as well as environmental factors have been implicated.

1.3.1 Familial Cancer-predisposing Syndromes

Increased risk of sarcoma development has been identified among individuals with certain familial cancer syndromes, all of which result from autosomal recessive mutations in (or deletion of) tumour suppressor genes. Some of these inheritable mutations have also been demonstrated in tumour cells from sporadic cases of STS and even in tumour cells in other cancers (Malkin, Li et al. 1990).

1.3.1.1 Li Fraumeni Syndrome

First described in a retrospective study of families with a high incidence of childhood rhabdomyosarcoma, a follow-up study found that STS were the second commonest tumours among members of the index four families (Li and Fraumeni 1982). The syndrome is characterized by a germ-line mutation in the *TP53* tumour-suppressor gene that causes an increased lifetime risk of developing malignant tumours of a wide variety, at a young age and at multiple primary sites. In decreasing order of frequency, other cancers observed in the syndrome include pre-menopausal breast cancer, osteosarcoma, brain tumours, adrenocortical tumours, lung cancer, leukaemia, lymphoma, renal cancer, colorectal cancer and gonadal germ cell tumours (Bougeard, Sesboue et al. 2008).

1.3.1.2 Neurofibromatosis Type I

Another heritable disease, neurofibromatosis Type I is caused by an autosomal dominant mutation of the tumour-suppressor gene, *NF1* and characterized by the development of skin lesions, skeletal dysplasia, and benign neurofibromas. Neurofibromatosis patients however have a higher frequency of development of malignant peripheral nerve sheath tumour (MPNST) subtype of STS than the normal population such that it is regarded as a hallmark complication of the disease (Korf 2000). Patients have been noted to occasionally develop other STS subtypes, Gastrointestinal stromal tumours, leukaemias and phaeochromocytomas (King, Debaun et al. 2000).

1.3.1.3 Familial Retinoblastoma

Characterised by the development of malignant tumours of the retina in infancy or early childhood, familial retinoblastoma patients carry a germ-line mutation in one copy of the tumour suppressor *RB1* gene. Damage to the other normal allele of the gene (‘a second hit’) has been implicated in tumour development (Knudson 1971, Wong, Boice et al. 1997). Several years (10 or more) after radiotherapy for their ocular tumours, these patients have a 20-fold increased risk compared to the normal population, of developing second malignant tumours the most common of which are bone and soft tissue sarcomas (Marees, Moll et al. 2008). The second tumours occur with a frequency higher than that for patients with sporadic retinoblastoma and while the majority occur within the radiation field, up to a quarter arise outside the field suggesting a radiation-independent aetiology (Kleinerman, Schonfeld et al. 2012).

1.3.2 Viral Infections

1.3.2.1 Kaposi Sarcoma-associated Herpes Virus (KSHV)

Formerly referred to as the Human Herpes Virus Type 8 (HHV8), this virus was initially identified in AIDS-related form of Kaposi sarcoma (Chang, Cesarman et al. 1994). It was also detected in all other forms of Kaposi sarcoma and follow-up studies showed that antibodies to KSHV were present in blood even before development of tumour (Weiss, Whitby et al. 1998), suggesting that HHV-8 plays a role in the pathogenesis of Kaposi sarcoma, while immunosuppression determines its clinical course (Fletcher, Unni et al. 2002).

1.3.2.2 Ebstein-Barr Virus

Similarly, leiomyosarcomas are rare in immunocompetent children, but significantly more common among HIV-positive ones. Detection of Epstein Barr Virus (EBV) in leiomyomas and leiomyosarcomas from HIV-positive children, but not HIV-negative ones led researchers to conclude that EBV infection of smooth muscle in AIDS patients contributes to the pathogenesis of benign and malignant tumours in these patients (McClain, Leach et al. 1995). Further, EBV infection in a small percentage (<1%) of post-transplant immunosuppressed patients results in mesenchymal tumours that arise in vascular smooth muscle (Jonigk, Laenger et al. 2012).

1.3.3 Ionizing Radiation

Following **radiation exposure**, usually as adjuvant therapy for cancer e.g. retinoblastoma or breast cancer, a significantly higher than normal frequency of occurrence of sarcomas has been noted (Warren and Sommer 1936, Robinson, Neugut et al. 1988). The majority of the recorded tumours are un-differentiated pleomorphic sarcomas and the relative risk increases with the dose of radiation the patients were exposed to (Sheppard and Libshitz 2001). Comprising 3 - 5% of all STS, radiation-related sarcomas were found to carry a poor prognosis (Hazard ratio 1.7) compared to other STS (Gladdy, Qin et al. 2010, Bjerkehagen, Smastuen et al. 2012).

1.3.4 Chronic Lymphoedema

Stewart-Treves syndrome was first described in 1948 after women who had chronic lymphoedema as a complication of axillary lymph node surgery were noted to develop lymphangiosarcoma (Stewart and Treves 1948). Similar tumours have subsequently been described in patients with patients with lymphoedema from other causes, but rare in patients without lymphoedema (Sordillo, Chapman et al. 1981). Some authors have attributed this STS development to regional immunosuppression in the limbs of these patients (Schreiber, Barry et al. 1979).

1.3.5 Other Environmental Factors

Exposure to certain chemicals such as vinyl chloride and thorotrast has been associated with development of hepatic angiosarcoma (Makk, Creech et al. 1974, Kaick, Dalheimer et al. 1999). Yet other STS subtypes have been linked (with limited evidence) to occupational exposure to chemicals including dioxin (Fingerhut, Halperin et al. 1991), phenoxyherbicides and chlorophenols (Hoppin, Tolbert et al. 1998).

1.4 DIAGNOSIS

Soft tissue sarcomas most commonly present as painless lumps, the size of which generally depends on the anatomic location. While superficial tumours and those in the distal extremities or head/neck are often discovered early and at a small size; deep-seated (e.g. retroperitoneal) tumours or those in proximal extremities may grow to a large size, or occasionally become symptomatic (due to compression of nerves, vessels or other organs) before they are discovered (Cormier and Pollock 2004, Clark, Fisher et al. 2005). Many tumours are discovered after trauma to nearby structures, and this leads to initial misdiagnosis as benign trauma-related pathology e.g. haematoma or fat necrosis. In order to prevent late diagnoses, current guidelines by the British Sarcoma Group recommend the investigation by a specialist sarcoma centre, of any soft tissue mass that is increasing in size, greater than 5cm in diameter, or deep to the fascia (Grimer, Judson et al. 2010).

Radiological imaging by Ultrasonography, Plain radiography, Computed Tomography, Magnetic Resonance Imaging, is typically the first modality of investigation of soft tissue masses. In experienced hands, radiologic features are able to rule out malignancy in the majority of cases. Imaging is often necessary to guide biopsy for histological diagnosis as well as determination of tumour size and extent for staging the disease and for determination of appropriate treatment modalities. It is also necessary for follow-up; early identification of recurrence or metastasis, and prognostication (Cormier and Pollock 2004).

While initial histological characterisation of sarcomas is usually by cellular morphology with most tumours composed of spindle cells, epitheloid cells, pleomorphic cells or small round cells, accurate diagnosis typically requires the employment of ancillary methods such as immunohistochemistry, cytogenetic and molecular genetic techniques (Thway 2009). Being rare as well as very morphologically diverse (with >100 histological subtypes), soft tissue tumours tend to constitute a significant diagnostic challenge for many pathologists (Bovée and Hogendoorn 2009). In addition, it is now known that STS with similar morphologic appearance may actually represent distinct pathologic entities with different clinical course and response to treatment.

The major subtypes of malignant soft tissue tumours according to the 2013 WHO classification are summarized in Table 1.1. While the current classification retains the traditional broad tumour classes that are based upon the adult mesenchymal tissues that they resemble the most, the WHO Working Group for soft tissue tumour classification have since 2002 incorporated the molecular and genetic features of these tumours in addition to histopathology (Fletcher, Unni et al. 2002, Fletcher, Bridge et al. 2013). This is because molecular and genetic information has vastly increased the understanding of the biology of these tumours leading to improved classification and more accurate prediction of their clinical course (Fletcher, Bridge et al. 2013, Fletcher 2014).

Table 1.1: Major Subtypes of Soft Tissue Sarcomas

Adipocytic Tumours	*Intermediate (locally aggressive)* • Atypical lipomatous tumour / Well differentiated liposarcoma *Malignant* • Dedifferentiated liposarcoma • Myxoid liposarcoma • Pleomorphic liposarcoma • Liposarcoma, not otherwise specified
Fibroblastic / Myofibroblastic Tumours	*Intermediate (rarely metastasizing)* • Dermatofibrosarcoma protuberans • Solitary fibrous tumour (malignant) • Inflammatory myofibroblastic tumour • Myxoinflammatory fibroblastic sarcoma • Infantile fibrosarcoma *Malignant* • Adult fibrosarcoma • Myxofibrosarcoma • Low grade fibromyxoid sarcoma • Sclerosing epithelioid fibrosarcoma
Smooth Muscle Tumours	*Malignant* • Leiomyosarcoma
Skeletal Muscle Tumours	*Malignant* • Embryonal rhabdomyosarcoma • Alveolar rhabdomyosarcoma • Pleomorphic Rhabdomyosarcoma • Spindle Cell/Sclerosing Rhabdomyosarcoma
Vascular Tumours	*Intermediate (rarely metastasizing)* • Kaposi sarcoma *Malignant* • Epithelioid haemangioendothelioma • Angiosarcoma of soft tissue
Chondro-osseous Tumours	*Malignant* • Extraskeletal mesenchymal Chondrosarcoma • Extraskeletal osteosarcoma
Gastrointestinal Stromal Tumours	*Malignant* • Malignant Gastrointestinal Stromal Tumour
Nerve Sheath Tumours	*Malignant* • Malignant Peripheral Nerve Sheath Tumour • Epitheloid malignant Peripheral Nerve Sheath Tumour • Malignant Triton Tumour • Malignant Granular Cell Tumour • Ectomesenchymoma
Tumours of uncertain differentiation	*Intermediate (rarely metastasising)* • Angiomatoid fibrous histiocytoma *Malignant* • Synovial sarcoma • Epithelioid sarcoma • Alveolar soft part sarcoma • Clear cell sarcoma of soft tissue • Extraskeletal myxoid chondrosarcoma • Malignant mesenchymoma • Desmoplastic small round cell tumour • Extrarenal rhabdoid tumour • Neoplasms with perivascular epithelioid cell differentiation (PEComa) • Intimal sarcoma
Undifferentiated/Unclassified Sarcomas	*Malignant* • Undifferentiated spindle Cell Sarcoma • Undifferentiated Pleomorphic Sarcoma • Undifferentiated round cell sarcoma • Undifferentiated epitheloid sarcoma • Undifferentiated sarcoma NOS

Subtype classification is based on 2013 WHO Classification of Bone and Soft Tissue Tumours (Fletcher, Bridge et al., 2013). Only non-benign soft tissue sarcoma subtypes are shown.

1.5 PROGNOSIS AND TREATMENT

Histologic diagnosis by itself may not be sufficient to predict the clinical course in most STS (Fletcher, Bridge et al. 2013). Other than histology, factors that have been shown to influence prognosis include the tumour grade, depth, size and patient's age (Grobmyer and Brennan 2003).

1.5.1 Tumour Grading

Tumour histopathologic grade is generally regarded as the most important factor for prediction of distant metastasis in soft tissue sarcomas (Coindre 2006, Thway 2009) and the most widely-accepted STS grading system as noted in the 2013 WHO classification was developed by the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) and based on the Trojani system (Trojani, Contesso et al. 1984), a name by which it is still commonly referred. As summarised in Tables 1.2, it totals scores assigned to three independent prognostic factors including - histological subtype/degree of differentiation (Table 1.3); extent of tumour necrosis; and mitotic index (Coindre, Terrier et al. 1996).

Table 1.2: Grading System for Soft Tissue Sarcomas

Histological Parameter		Definition
1. Tumour differentiation	Score 1	Sarcomas closely resembling normal adult mesenchymal tissue (e.g. well-differentiated liposarcoma, well-differentiated leiomyosarcoma)
	Score 2	Sarcomas for which histological typing is certain (e.g., myxoid Liposarcoma, myxofibrosarcoma).
	Score 3	Embryonal and undifferentiated sarcomas, sarcomas of doubtful type, synovial sarcomas.
2. 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
3. Tumour necrosis	Score 0	No necrosis
	Score 1	<50% tumour necrosis
	Score 2	50% tumour necrosis
Histological grade	***Grade 1***	***Total score 2,3***
	Grade 2	***Total score 4,5***
	Grade 3	***Total score 6, 7, 8***

** HPF – High Powered Field*

Grading System was developed by the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) and based on the Trojani system (Table adapted from Coindre 2006).

Table 1.3: Differentiation Scores for Soft Tissue Sarcoma Subtypes in the FNCLCC grading system

Histological Type	Differentiation Score
Well-differentiated liposarcoma	1
Well-differentiated leiomyosarcoma	
Malignant Neurofibroma	
Well-differentiated fibrosarcoma	
Myxoid Liposarcoma	2
Conventional leiomyosarcoma	
Conventional MPNST	
Conventional fibrosarcoma	
Myxofibrosarcoma	
Myxoid Chondrosarcoma	
Conventional angiosarcoma	
*High-grade Myxoid (Round Cell) Liposarcoma	3
Pleomorphic liposarcoma	
*Dedifferentiated Liposarcoma	
*Embryonal/alveolar/pleomorphic rhabdomyosarcoma	
Poorly-differentiated/pleomorphic leiomyosarcoma	
Poorly-differentiated/epithelioid angiosarcoma	
Poorly-differentiated MPNST	
Malignant triton tumour	
Synovial Sarcoma	
Extraskeletal Osteosarcoma	
*Extraskeletal Ewing Sarcoma	
Mesenchymal chondrosarcoma	
*Clear cell sarcoma	
*Epithelioid sarcoma	
*Alveolar Soft Part Sarcoma	
Malignant Rhabdoid Tumour	
Undifferentiated (spindle cell and pleomorphic) sarcoma	

Grading System was developed by the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) and based on the Trojani system (Table adapted from Coindre 2006).

The FNCLCC (Trojani) system is however of very limited use in determining likelihood of local tumour recurrence, which is believed to depend mainly on surgical resection margins (discussed in detail below. It is therefore not recommended for use in tumours that rarely metastasize, as well as specific STS subtypes including dedifferentiated and round cell liposarcoma, Rhabdomyosarcoma, Epitheloid sarcoma, Ewing sarcoma, Alveolar soft part Sarcoma and clear cell sarcoma where histologic type alone is more predictive of metastatic potential.

1.5.2 Tumour Staging

While tumours that arise in superficial tissues such as atypical lipomatous tumours in subcutaneous tissue, are generally considered to be less aggressive since they rarely metastasize, deep-seated tumours tend to follow a more aggressive clinical course (Billing, Mertens et al. 2008, Thway 2009). Metastasis in majority of STS is via the haematogenous route and unlike carcinoma, lymph node spread is rare (<5%), with few exceptions including epitheloid sarcomas, synovial sarcomas, embryonal rhabdomyosarcoma, clear cell sarcoma and angiosarcoma (Fong, Coit et al. 1993).

In common clinical use for prognostication is the TNM staging system that combines systems developed by the UICC and AJCC and incorporates tumour size, depth, nodal and distant metastases with tumour grade. The features of this system are summarised in Table 1.4.

Table 1.4: TNM Staging System for of Soft Tissue Sarcomas

Parameter	Definition				
Primary tumour (T)[1]	Tx	Primary tumour cannot be assessed			
	T0	No evidence of primary tumour			
	T1	Tumour < 5cm in greatest dimension T1a: superficial tumour T1b: deep tumour			
	T2	Tumour > 5cm in greatest dimension T2a: superficial tumour T2b: deep tumour			
Regional lymph nodes (N)[2]	Nx	Regional lymph nodes cannot be assessed			
	N0	No regional lymph node metastasis			
	N1	Regional lymph node metastasis			
Distant metastasis (M)	M0	No distant metastasis			
	M1	Distant metastasis			
Grading	Low Grade	FNCLCC Grade 1			
	High Grade	FNCLCC Grade 2 and 3			
TNM Stage	**Stage IA**	T1a	N0, Nx	M0	Low grade
		T1b	N0, Nx	M0	Low grade
	Stage IB	T2a	N0, Nx	M0	Low grade
		T2b	N0, Nx	M0	Low grade
	Stage IIA	T1a	N0, Nx	M0	High grade
		T1b	N0, Nx	M0	High grade
	Stage IIB	T2a	N0, Nx	M0	High grade
	Stage III	T2b	N0, Nx	M0	High grade
	Stage IV	Any T	N1	M0	Any grade
		Any T	Any N	M1	Any grade

1 – Superficial and deep tumours are exclusively superficial and deep to the deep fascia, respectively and tumours that are retroperitoneal, mediastinal and pelvic in location are regarded as deep tumours

2 – Since regional node involvement is rare, N0 instead of NX or pNX is assigned in cases where nodal status is not assessed either clinically or pathologically

Table adapted from 2013 WHO Classification of Bone and Soft Tissue Tumours (Fletcher, Bridge et al., 2013)

1.5.3 Treatment Strategy

Due to differences in sensitivity to various therapeutic modalities, accurate histological diagnosis, grading and staging are essential for determining appropriate regimes. However, wide surgical resection of the primary tumour remains the mainstay of treatment for the majority of localised STS in adults in the UK (Grimer, Judson et al. 2010). Histological assessment of resection margins is then carried out as this has been found to be the best parameter for assessing the likelihood of local recurrence and probably influences overall survival (Novais, Demiralp et al. 2010). In general, excision with a normal soft tissue margin of 1cm or equivalent (e.g. normal fascia) is desirable and in appropriate cases, further excision may be carried out in order to achieve the desired safe margin. Surgical excision may be also considered as a palliative procedure in some cases of metastatic disease or as a definitive procedure for solitary or few metastatic tumour deposits (Grimer, Judson et al. 2010).

Post-operative radiotherapy is the standard adjuvant modality for intermediate and high-grade tumours, although pre-operative radiotherapy may be used especially in radio-sensitive subtypes such as myxoid liposarcoma (de Vreeze, de Jong et al. 2008) to reduce tumour size (down-stage) and increase the probability of achieving safe surgical margins. Chemotherapy has not been shown to be effective in the adjuvant setting, but neoadjuvant chemotherapy is integral to treatment of certain STS subtypes such as Ewing's tumours, alveolar and embryonal Rhabdomyosarcoma and may be used for down staging in selected cases. For the majority of sarcomas, chemotherapy is reserved as palliation for advanced metastatic disease. Targeted molecular therapies are in use for specific STS subtypes e.g. Imatinib for Gastrointestinal Stromal Tumours and Dermatofibrosarcoma protuberans (discussed later).

1.5.4 Follow-up and Survival

High grade STS tend to relapse within 2 – 3 years, while low-grade tumours relapse later if at all (Casali, Jost et al. 2008). For intermediate and high grade therefore, current UK and European guidelines recommend 3 - 4 monthly follow-up with clinical examination and appropriate imaging for the first 2 – 3 years; bi-annual follow up till 5 years and subsequent annual follow up. While in patients with low-grade tumours, 4-6 monthly follow up for 3 – 5 years and subsequent yearly follow up is regarded as sufficient (Casali, Jost et al. 2008, Grimer, Judson et al. 2010).

Average 5-year survival for STS ranges between 50 and 65% and one third to half of patients develop lung metastases and/or local tumour recurrence despite the current standard of therapy (Grimer, Judson et al. 2010, Fletcher, Bridge et al. 2013, Francis, Dennis et al. 2013). There is therefore an urgent need for development of effective targeted therapies for use in metastatic disease and like for most cancers, the key to this is believed to lie in the genetic and molecular mechanisms that drive these tumours.

1.6 A Genetic Basis for Cancer

Initial proposals for genomic abnormalities being responsible for cancer were made in the early twentieth century by Boveri following observations of abnormal chromosomes in malignant cells that were published in 1914 and recently translated and reviewed by Henry Harris (Boveri 1914, Boveri 2008). With the establishment of DNA as the hereditary material of cells and developments in cytogenetics, specific chromosomal aberrations in many malignant tumours, particularly in haematological malignancies and certain mesenchymal tumours were demonstrated (Mertens, Panagopoulos et al. 2009). Further evidence was provided by the demonstration of oncogenic transformation of normal mouse fibroblast cells by transfection with total genomic DNA from cancer cell lines (Krontiris and Cooper 1981).

It is now generally accepted that while certain germ-line mutations are responsible for familial predisposition to cancer, acquired changes in the genome of cancer cells are responsible for the causation and pathogenesis of cancer (Weir, Zhao et al. 2004, Stratton, Campbell et al. 2009). Various changes have been described at different levels within the genome of cancer cells, ranging from loss/gain of entire chromosomes to subtle nucleotide base mutations. These changes generally lead to alterations to the normal expression of genes that regulate cell proliferation and differentiation and result in the malignant phenotype of cancer cells. While some of these changes are believed play a role in the initiation of tumourigenesis ('drivers'), others ('passengers') are acquired as tumours progress and do not necessarily contribute to pathogenesis (Stratton, Campbell et al. 2009).

DNA sequence mutations involving one or a few nucleotides in the coding region of oncogenes or tumour suppressor genes have been associated with many cancers (Stratton, Campbell et al. 2009). These base substitutions, deletions or insertions can be 'mis-sense' or 'non-sense' mutations and disrupt the function of the products of these oncogenes or tumour-suppressor genes resulting in uncontrolled cell proliferation. Typical examples include k-*RAS* mutations in pancreatic cancer, c-*KIT* mutations in gastrointestinal tumours and *p53* mutations in a wide range of cancers (Lengauer, Kinzler et al. 1998, Stratton, Campbell et al. 2009).

Numerical alterations in the normal diploid chromosome complement (aneuploidy) due to gains or losses in whole chromosomes are a feature of most human cancers (Lengauer, Kinzler et al. 1998). Chromosome numbers between 60 and 90 are common among cancer cells with significant intra-tumour heterogeneity as well as structural abnormalities such as inversions, deletions and duplications (Rajagopalan and Lengauer 2004). Structural rearrangements in chromosomal DNA such as translocations are also notable in cancer pathogenesis. They can be balanced, in which the rearrangement changes the gene order, but does not duplicate or delete any part of the DNA; or unbalanced, where there is an associated loss or gain of chromosomal DNA.

Balanced translocations have been described in most types of tumours (Lengauer, Kinzler et al. 1998, Mitelman, Johansson et al. 2003). Many of these are associated with, and even occasionally are pathognomonic of specific tumour types. They also correlate with clinical course and are of prognostic significance, in some tumours (Mitelman, Johansson et al. 2007). The molecular consequence of most recurrent, specific balanced translocations is the rearrangement of genetic

sequences and formation of fused genes, which results in either altered expression of the involved genes or formation of abnormal proteins. In addition, *in vivo* experimental studies of these gene fusion constructs in transgenic animals results in neoplasia (Heisterkamp, Jenster et al. 1990, Perez-Losada, Pintado et al. 2000), while silencing of the fusion transcripts in *in vitro* systems reverses their neoplastic characteristics (Thomas, Gessner et al. 2005, Thomas, Greil et al. 2006). The sum total of this evidence overwhelmingly suggests that gene rearrangements are not only pathogenetically significant, but probably represent important events in tumour initiation.

Somatic copy number aberrations (SCNA), in which restricted regions of chromosomes are either lost (deletions) or gained (amplifications), with consequent alteration in the expression of oncogenes and tumour suppressor genes are another feature of many tumour genomes (Albertson, Collins et al. 2003). They differ from copy number variations (CNV), which are variations in DNA copy number seen in the germ-line DNA of different individuals and are of no pathological significance (Beroukhim, Mermel et al. 2010). They can be detected by molecular cytogenetic methods e.g. Fluorescence-*in-situ*-Hybridization (FISH) and comparative genomic hybridization (CGH) and have led to the identification of cancer-related genes and potential therapeutic targets (Zender, Spector et al. 2006, Lahortiga, De Keersmaecker et al. 2007, The Cancer Genome Atlas Research Network 2008).

1.7 Genetics of Soft Tissue Sarcomas

Genomic alterations ranging from subtle sequence changes to possession of multiples of the normal chromosome complement have been described among STS. Significant correlation between molecular and genetic features of these tumours and their clinico-biologic properties have led to better classification, clinical management and identification of potential therapeutic targets (Mertens, Panagopoulos et al. 2009). On the basis of their known genetic changes, most STS may be allocated four main groups as discussed in the following sections.

1.7.1 Sarcomas with Specific Oncogenic Mutations

Gene sequence mutations have been found to be specific to certain sarcoma subtypes, most notably the Gastro-Intestinal Stromal Tumours (GIST). Originally classified histologically as smooth muscle tumours and then immunophenotypically as neural tumours (Fletcher, Berman et al. 2002), the discovery of these gene mutations and their consequent molecular alterations have not only provided further insight to the cells of origin and pathogenesis of these tumours, but been translated into actual and clinically-effective molecular therapy. Hirota et al, in 1998 published a study in which they demonstrated that GISTs carried a specific sequence mutation in the *cKIT* gene, which leads to its constitutional activation, independent of its ligand (Hirota, Isozaki et al. 1998). Subsequent studies showed that in a subset of GISTs that were negative for the *KIT* mutation, mutations with similar consequences in another tyrosine kinase, platelet-derived growth factor receptor gene, *PDGFRA* were present (Heinrich, Corless et al. 2003).

The demonstration of KIT by immunohistochemistry (positive in up to 95% of cases) has proven to be a sensitive diagnostic tool for this previously poorly categorized tumour. Furthermore and in concordance with the pathogenetic evidence, the clinical use of tyrosine kinase inhibitor treatments –

Imatinib and Sunitinib as first and second-line therapies has led to clinical response in patients with metastatic and inoperable disease and greatly improved what used to be a dire prognosis (Liegl-Atzwanger, Fletcher et al. 2010).

1.7.2 Sarcomas with Characteristic Translocations and Fusion Genes

This group, which constitutes approximately 20% of cases of STS, are characterized by specific chromosomal aberrations, mostly balanced translocations, some of which are even considered to be pathognomonic (Mitelman, Johansson et al. 2007, Mertens, Panagopoulos et al. 2009). Most of these gene fusions and other molecular consequences are believed to be important steps in tumorigenesis. At present, this group includes more than a dozen sarcoma sub-types and about 30 known translocations, most of which result in fusion genes as summarised in Table 1.5.

Based on the current understanding of genomics in tumour development, the specificity of fusion genes for sarcoma subtypes makes it likely that they are 'driver' aberrations. This is because they are consistently found in these sarcomas wherein they frequently are the sole karyotypic abnormality. It is therefore presumable that they are important in tumour initiation and promote cell proliferation with accumulation of further 'passenger' mutations that confer more malignant characteristics as the tumour progresses (Mitelman, Johansson et al. 2007). This concept is further given credence by the results of studies that show that like in other cancers, expression of these fusion gene constructs *in vitro* and in experimental animals leads to the development of tumours e.g. leukaemia, liposarcoma that have the phenotypic characteristics of the human tumours (Daley, Van Etten et al. 1990, Perez-Losada, Pintado et al. 2000). In addition, the clinical success of therapies that specifically target fusion transcripts or their molecular effects, as demonstrated in Philadelphia chromosome-positive Chronic Myeloid Leukaemia make them potential targets for future therapies (Deininger, Buchdunger et al. 2005).

The observation that sarcomas may arise following relatively few specific genetic events may partially explain why sarcomas in children and young patients tend to belong to this group, while sarcomas without characteristic translocations appear to require the accumulation of genetic alterations over time and are thus mostly seen in older patients (Thway 2009, Ordonez, Osuna et al. 2010). The molecular effects of fusion genes that elucidate how they contribute to pathogenesis has been demonstrated in many STS subtypes and on this basis, translocation sarcomas may be categorized based on the molecular alterations that result from their characteristic fusion genes. These will now be discussed using relevant examples.

Table 1.5: Characteristic chromosomal translocations and fusion genes in Soft Tissue Sarcoma Subtypes

Sarcoma Subtype	Genomic Aberration	*Genetic Consequence*
Fusion Genes resulting in Aberrant Transcription Factors		
Ewing's'/PNET	t(11;22)(q24;q12) t(21;22)(q22;q12) t(7;22)(p22;q12) t(17;22)(q12;q12) t(2;22)(q33;q12) t(16;21)(p11;q22) t(2;16) inv (22)	*EWSR1-FLI1* *EWSR1-ERG* *EWSR1-ETV1* *EWSR1-ETV4* *EWSR1-FEV* *FUS-ERG* *FUS-FEV* *EWSR1-ZSG*
Myxoid/round cell liposarcoma	t(12;16)(q13;p11) t(12;22)(q13;q12)	*FUS–DDIT3* *EWSR1–DDIT3*
Desmoplastic small round cell tumour	t(11;22)(p13;q12)	*EWSR1–WT1*
Clear cell sarcoma	t(2;22)(q33;q12) t(12;22)(q13;q12)	*EWSR1–CREB1* *FUS–ATF1* *EWSR1–ATF1*
Angiomatoid fibrous histiocytoma	t(2;22)(q33;q12) t(12;16)(q13;p11) t(12;22)(q13;q12)	*EWSR1–CREB1* *FUS–ATF1* *EWSR1–ATF1*
Extraskeletal myxoid chondrosarcoma	t(3;9)(q12;q22) t(9;15)(q22;q21) t(9;17)(q22;q11) t(9;22)(q22;q12)	*TFG–NR4A3* *TCF12–NR4A3* *TAF15–NR4A3* *EWSR1–NR4A3*
Low-grade fibromyxoid sarcoma	t(7;16)(q33-34;p11) t(11;16)(p11;p11)	*FUS–CREB3L2* *FUS–CREB3L1*
Alveolar soft part sarcoma	t(X;17)(p11;q25)	*ASPSCR1–TFE3*
Alveolar rhabdomyosarcoma	t(1;13)(p36;q14) t(2;2)(p23;q36) t(2;13)(q36;q14) t(X;2)(q13;q36)	*PAX7–FOXO1A* *PAX3–NCOA1* *PAX3–FOXO1A* *PAX3–FOXO4*
Solitary fibrous tumour	inv (12)(q13q13)	*NAB2-STAT6*
Epitheloid Haemangioepithelioma	t(1;3)(p36.3;q25)	*WWTR1-CAMTA1*
Mesenchymal chondrosarcoma	t(8;8)(q21.1;q13.3)	*HEY1-NCOA2*
Fusion Genes resulting in Dysregulated tyrosine Kinase Genes		
Infantile fibrosarcoma	t(12;15)(p13;q25)	*ETV6–NTRK3*
Inflammatory myofibroblastic tumour	t(1;2)(q21;p23) inv(2)(p23q35) t(2;2)(p23;q13) t(2;4)(p23;q21) t(2;11)(p23;p15) t(2;17)(p23;q23) t(2;19)(p23;p13)	*TPM3–ALK* *ATIC–ALK* *RANBP2–ALK* *SEC31A–ALK* *CARS–ALK* *CLTC–ALK* *TPM4–ALK*
Fusion Genes involving Chromatin Remodelling Genes		
Synovial sarcoma	t(X;18)(p11;q11) t(X;20)(p11;q13)	*SS18–SSX1,* *SS18–SSX2* *SS18L1–SSX1* *SS18–SSX4*
Fusion Genes resulting in Growth Factor Dysregulation		
Dermatofibrosarcoma protuberans	Ring chromosome, t(17;22)(q21;q13), der(22)t(17;22)	*COL1A1–PDGFB*

Subtypes and genetic data are based on 2013 WHO Classification of Bone and Soft Tissue Tumours (Fletcher, Bridge et al., 2013). Only non-benign soft tissue sarcoma subtypes are shown.

1.7.2.1 Fusions Resulting In Aberrant Transcription Factors

About half of the fusion proteins that have been demonstrated in STS so far involve products of genes that encode transcription factors (Jain, Xu et al. 2010). Most important among these genes are those that encode the **TET** family of transcription proteins that include the Translocated in Liposarcoma gene ***TLS*** (also known as *FUS*); Ewing's' Sarcoma gene ***EWSR1***; and TATA-binding Associated Factor gene ***TAFII68*** (also known as *TAF15*). The members of this family display considerable redundancy in their pathogenetic roles in various STS subtypes (Jain, Xu et al. 2010). Apart from the TET family, other transcription factor genes are also involved in the fusion genes specific for some sarcoma subtypes (Table 1.5).

The fusion gene products typically consist of the replacement of the RNA-binding domain of one transcription factor with the DNA-binding domain of its fusion partner and this is believed to result in alterations in specificity and transcription activity of the target gene. Resultant effects that have been demonstrated with some of these fusion transcripts include increased cell proliferation, improved cell survival, reduced apoptosis, induction of angiogenesis, invasion and metastases (Bovée and Hogendoorn 2009).

Examples of STS subtypes with fusion genes that have been demonstrated to have this pathogenetic mechanism are shown on Table 1.5. Interestingly, some of these fusion genes are shared by STS subtypes e.g. the t (12;16)(q13;p11) that fuses *EWRS1* with *ATF1* that is characteristic in both Clear Cell Sarcoma and Angiomatoid Fibrous Histiocytoma. This observation of identical genetic defects in morphologically and clinically distinct tumours has led authors to suggest that these gene fusions play an early role in sarcomagenesis by inducing transformation in mesenchymal progenitor cells, which then go on to differentiate along distinct pathways (Antonescu, Dal Cin et al. 2007, Iwasaki, Nabeshima et al. 2009).

1.7.2.2 Fusion Genes Resulting In Deregulated Receptor Tyrosine Kinases

Fusion genes have been described in STS that result in constitutive activation of receptor tyrosine kinases in a manner analogous to the *BCR- ABL* fusion of Chronic Myeloid Leukaemia. Examples of these STS include congenital fibrosarcomas (Knezevich, McFadden et al. 1998, Krishnan, Khanna et al. 2008) and Inflammatory Myofibroblastic Tumour (IMT) (Lawrence, Perez-Atayde et al. 2000, Patel, Murphy et al. 2007, Mertens, Panagopoulos et al. 2009).

1.7.2.3 Fusion Genes Involving Chromatin Remodelling Genes

In synovial sarcomas, the SS18 *(SYT)* gene located at 18q11 and one of three *SSX* genes in the Xp11 band forms one of three fusion genes – *SS18-SSX1* (Clark, Rocques et al. 1994)*; SS18-SSX2* (de Leeuw, Balemans et al. 1995)*;* or *SS18-SSX4* (Skytting, Nilsson et al. 1999). The SSX genes which are believed to regulate transcription via epigenetic mechanisms (histone modification and DNA methylation) are juxtaposed to form a dysfunctional chimeric gene (de Bruijn, Nap et al. 2007).

1.7.2.5 Fusion Genes Resulting in Growth Factor Dysregulation

Dermatofibrosarcoma protuberans is characterized by the fusion of the platelet-derived growth factor *β* gene (*PDGFB*) at 22q13 and the *α*1 chain of type 1 collagen gene (*COL1A1*) at 17q22, which results in dysregulation of autocrine growth signalling by PDGFB (Simon, Pedeutour et al. 1997). The *COL1A1-PDGFB* chimeric gene is the result of either ring chromosomes that contain low-level amplified sequences from 17q and 22q (Minoletti, Miozzo et al. 1995, Naeem, Lux et al. 1995, Kiuru-Kuhlefelt, El-Rifai et al. 2001), or an unbalanced translocation (17;22)(q22;q13) (Pedeutour, Lacour et al. 1996, Dobin, Diaz et al. 1999).

As would be expected with this pathogenetic mechanism, tumours carrying the *COL1A1-PDGFB* fusion genes show clinical response to tyrosine kinase inhibitors e.g. imatinib mesylate (McArthur, Demetri et al. 2005). Copy number gains of this fusion gene have also been described in the rare transformation of DFSP into high-grade fibrosarcoma (Abbott, Erickson-Johnson et al. 2006).

1.7.3 Sarcomas with Characteristic Somatic Copy Number Changes

Somatic copy number abnormalities (SCNA), mostly amplifications have been detected in most soft tissue sarcomas (Myllykangas, Böhling et al. 2007). Some amplifications are seen with such high frequency in certain sarcoma subtypes, that they are regarded as characteristic. While they are not in principle specific, because they are also seen with significantly high frequency in other distinct sarcoma subtypes, they can play a valuable role in achieving correct diagnosis, particularly in cases where morphological characteristics are equivocal (Bovée and Hogendoorn 2009). In addition, these amplifications are believed to harbour clues to the pathogenesis of the sarcomas and thus potential therapeutic targets (Coindre, Pedeutour et al. 2009).

Atypical Lipomatous Tumour (ALT)/Well-differentiated Liposarcomas (WDLPS) and Dedifferentiated Liposarcomas (DDLPS), which constitute 20-25% of all sarcomas, represent the most important example of sarcomas in this category. The presence of a supernumerary ring/giant 'marker' chromosome that always contains amplified sequences from the long arm of chromosome 12 (12q14-15) is characteristic of these tumours. While this usually represents the sole karyotypic aberration in ALT/WDLPS, secondary aberrations such as double minute chromosomes and multiple copies of the marker chromosome with significant intratumour heterogeneity are sometimes observed in DDLPS (Pedeutour, Suijkerbuijk et al. 1994, Italiano, Bianchini et al. 2009).

Among important target genes in the amplified regions is *MDM2* located at 12q15, which is over-expressed in up to 100% of tumours and whose product inhibits p53 thus promoting cell survival. Other important potential oncogenes include *CDK4*, which is involved in the G1-S phase cell cycle checkpoint and *HMG*, which is involved in chromatin organization during transcription (Iwasaki, Nabeshima et al. 2009). Amplifications of the 1p32 and 6q23 bands in DDLPS lead to the amplification of *JUN* and *ASK1*, respectively. These genes are also amplified in mouse models of liposarcoma. Both gene products inhibit peroxisome proliferator-activated receptor gamma (PPAR-γ), a major player in adipocytic differentiation and may thus play a role in the dedifferentiation of liposarcomas and represent potential therapeutic targets (Mertens, Panagopoulos et al. 2009).

Despite the fact that they are not exactly tumour specific, identification of these karyotypic changes or gene amplifications (*MDM2* and *CDK4*) is playing an increasing role in diagnosis of these tumours (Mertens, Panagopoulos et al. 2009). They may be used to differentiate ALT/WDLPS from morphologically similar, but benign adipocytic tumours or for the establishment of a lipogenic origin for dedifferentiated sarcomas. It has been suggested that majority of pleomorphic sarcomas that arise in the retroperitonum are dedifferentiated liposarcomas, since they carry these gene amplifications (Coindre, Mariani et al. 2003).

Another STS with a characteristic copy number change is the **Embryonal Rhabdomyosarcoma**, in which 70-100% of cases have allele loss at 11p15.5. Other characteristic changes include trisomy of chromosomes 2 and 8 (Anderson, Gordon et al. 1999, Mertens, Panagopoulos et al. 2009). This copy number alteration is also seen among Wilms' tumours and hepatoblastoma and though it may not be important from a diagnostic point of view, it is believed to be the site of pathogenetically important genomic change, since it is consistently detected (Anderson, Gordon et al. 1999, Xia, Pressey et al. 2002). Molecular studies have identified a number of cell cycle related genes in the region as candidates. They include *H19, IGF2, CDKN1C* and *BWR1A,* but whether or not they are actually involved in tumour pathogenesis is yet to be fully elucidated (Anderson, Gordon et al. 1999).

1.7.4 Sarcomas with Complex Karyotypic Aberrations

This group comprises the majority of STS (about 50%) and includes most adult tumours that exhibit spindle cell and pleomorphic morphology. Highly complex karyotypes with multiple numerical and structural aberrations are characteristic and marked intra-tumour heterogeneity is seen in these tumours with metaphase aberrations observed in virtually all cells and no two cells with an identical chromosome complement in many cases. In addition, all these tumours possess numerous recurrent and non-random somatic copy number alterations that have been reported to involve virtually every chromosome in the genome (Mandahl, Mertens et al. 2004).

Tumours that belong to this group include leiomyosarcoma (LMS); myxofibrosarcoma (MFS); pleomorphic subtypes of liposarcoma (PLPS) and rhabdomyosarcoma (PRMS); undifferentiated pleomorphic sarcomas (UPS or pleomorphic sarcoma UPS – not otherwise specified). Others are Malignant Peripheral Nerve Sheath Tumours and relatively rare and less well-studied STS like adult fibrosarcoma, angiosarcoma and extraskeletal osteosarcoma (Guillou and Aurias 2009).

Typically there is no known correlation between the chromosome breakpoints and clinico-pathological subtypes (Guillou and Aurias 2009, Iwasaki, Nabeshima et al. 2009). There is however, considerable overlap in the chromosomal aberration patterns across the tumour subtypes (Mertens, Panagopoulos et al. 2009). Both conventional and array CGH analyses of these tumours showed that LMS and UPS share a similar genomic profile (up to 80% similarity) of frequently amplified/lost sequences, some of which even have prognostic significance (Larramendy, Gentile et al. 2008, Carneiro, Francis et al. 2009). Similarly, MFS and PLPS are reported to show very similar genomic aberrations that are distinct from those in LMS. However, frequent copy number changes in certain regions are common to all four STS subtypes, including deletion in 10q and 13q, as well as amplifications in the short arm of chromosome 5 (Idbaih, Coindre et al. 2004).

Generally, it is believed that these genomic regions are amplified and deleted frequently because they contain oncogenes and tumour suppressor genes, respectively that are involved in tumour pathogenesis (Helman and Meltzer 2003). In a few regions, molecular pathogenetic consequences of the copy number changes have been demonstrated, but many marker chromosomes and their target genes remain elusive because of dearth of investigated cases and the karyotypic complexity. Some of these as yet unidentified unbalanced chromosomal aberrations may even include pathogenetically-important gene fusions (Mertens, Panagopoulos et al. 2009).

1.8 Cellular Origin of Sarcomas

The exact cell from which sarcomas originate remains unknown. Despite nomenclature that is based on morphological resemblance to adult mesenchymal cell lineages, a step-wise progression model from normal cell through dysplasia to malignancy (as established in carcinomas) does not appear to apply to sarcomas (Thway 2009). STS that exhibit features of specific mesenchymal lineage differentiation such as Rhabdomyosarcoma are known to arise in areas of the body where skeletal muscle tissue is sparse or even absent (Thway 2009) and many STS with similar histologic appearance have been shown to represent distinct clinic-biologic and molecular genetic entities (Fletcher, Bridge et al. 2013)

A concept of differentiation is therefore proposed for sarcomagenesis whereby malignant transformation occurs in mesenchymal stem cells which retain a limited capacity for differentiation into the various lineages (Xiao, Mohseny et al. 2013). The main problem with this concept however is that the identity of human mesenchymal stem cells remains controversial and as yet, they have not been categorically characterised or isolated with most preparation being a heterogenous mixture of cells, some of which express putative stem cell markers (Lin, Wang et al. 2011). However, ectopic expression of the EWS/FLI1 chimeric protein in human bone marrow derived mesenchymal progenitor cells reportedly led to their transformation with immunophenotypic and transcriptomic characteristics that closely resembled Ewing's tumours (Miyagawa, Okita et al. 2008).

Recent studies in transgenic mice where mesenchymal stem cells are better characterised, showed that myogenic progenitor cells were able to give rise to both myogenic and non-myogenic sarcomas depending on which genes were disrupted (Blum, Ano et al. 2013). These studies support the differentiation theory for sarcomagenesis but also raise the question of a possible role for specific genomic aberrations in the determination of the mesenchymal lineage differentiation of STS subtypes.

1.9 GENOMIC INSTABILITY IN SOFT TISSUE SARCOMAS

Genomic instability has been described as an enabling characteristic that facilitates the acquisition of the hallmarks of malignancy (Hanahan and Weinberg 2011). It is believed to account for the progressive accumulation of mutations in the cancer cell genome with the resultant effect that many tumour cells are characterised by multiple genomic changes at various levels and there is significant intra-tumour heterogeneity (Negrini, Gorgoulis et al. 2010). Instability has been described at the level of nucleotide sequences and microsatellites, where they are believed to be due to defects in Nucleotide Exchange Repair (NER) and Mismatch Repair (MMR) mechanisms for DNA damage during replication, respectively. Nucleotide Instability (NIN), in particular can result in drastic phenotypic changes, but these forms of instability, are mostly seen in hereditary forms of cancer and are thus relatively rare (Lengauer, Kinzler et al. 1998). Microsatellite instability (MIN) with associated mismatch repair defects has been well characterised in a subset of colorectal cancers and detected in up to 40% of STS in some studies (Martin, Grear Hurt et al. 1998, Boland and Goel 2010, Monument, Lessnick et al. 2012).

Chromosomal instability (CIN) in which cells accumulate chromosomal abnormalities (whole and/or segmental amplifications and deletions) at a higher than normal rate is however the major form of genomic instability in sporadic cancers (Negrini, Gorgoulis et al. 2010). In various cancers, CIN has been quantified using the number of SCNA present in the genome; the proportion of the genome affected; or the rate of Loss of Heterozygosity (LOH) and studies showed that it is a prominent feature of many STS, particularly among those without characteristic chromosomal abnormalities (Johnson, Gettings et al. 2007, Barretina, Taylor et al. 2010, Beroukhim, Mermel et al. 2010). Although the exact mechanisms of chromosomal instability are still far from being full understood, two main theories – the mutator hypothesis and oncogene-induced DNA replication stress model are well-established (Negrini, Gorgoulis et al. 2010).

DNA damage in the form of Double Strand Breaks (DSBs) occurs frequently (up to 10 events per day in normal cells) from a variety of intrinsic and extrinsic causes (Lieber 2010). In the normal cell, intact cell cycle checkpoints and DNA repair pathways are required to stop the mitotic cell cycle and repair the damage (or failing this, eliminate the damaged cell by initiating apoptosis) thus maintaining the integrity of the genome and preventing the transmission of this damage in mitosis. This involves a very complex interplay between proteins coded for by genes known collectively as ‘caretaker genes’ including prominent examples such as *TP53, BRCA1, BRCA2, CDNK2A* and *ATM* (Lengauer, Kinzler et al. 1998, Negrini, Gorgoulis et al. 2010). The mutator hypothesis states that mutations that lead to inactivation of the ‘caretaker’ genes result in accumulation of replication errors and further driver and passenger mutations.

While the mutator hypothesis is well established with regard to hereditary cancers such as those seen in Li-Fraumeni syndrome and Ataxia Telangiectasia that are characterised by germ-line mutations of *TP53* and *ATM* respectively, observations in sporadic cancers that only a third of tumours that exhibit CIN possess caretaker gene mutations suggest it may not be solely responsible for their CIN. Other mechanisms of caretaker gene inactivation such as copy number deletion or

epigenetic regulation may however play a role in these sporadic cancers (Negrini, Gorgoulis et al. 2010). Mutations and/or altered gene expression involving the *TP53* gene have been shown in numerous genetic and expression studies to be common among soft tissue sarcomas with complex genomic profiles (Toguchida and Nakayama 2009, Barretina, Taylor et al. 2010). Zhang et al demonstrated frequent association between *ATM* defects and rhabdomyosarcomas (Zhang, Bhakta et al. 2003) and previous study carried out in the Rare Tumour Research Group revealed frequent copy number losses involving the *ATM* gene in both GIST and leiomyosarcoma (Ul-Hassan, Sisley et al. 2009).

Reviewed by Halzonetis et al, the second model used to explain CIN is derived from evidence in cancerous and precancerous lesions and proposes that abnormal proliferation driven by activated oncogenes deregulates the entry of tumour cells into the mitotic cycle resulting in DNA replication stress and accumulation of double strand breaks at specific genomic sites called “common fragile sites” (Halazonetis, Gorgoulis et al. 2008). In both models, the genomic instability leads to the disruption of more caretaker genes that exacerbates the genomic instability and sets up a vicious cycle of progressive accumulation of genomic abnormalities.

More recently, whole genome sequencing studies in cancer have led to the description of a phenomenon in which a cancer cell undergoes a single ‘cataclysmic’ genomic event termed chromothripsis that results in a large number of genomic rearrangements (Kloosterman, Hoogstraat et al. 2011, Stephens, Greenman et al. 2011). This mechanism for chromosomal instability and possibly even tumour initiation is believed to occur in around 2 – 3% of cancers in general, but bone sarcomas in particular (Stephens, Greenman et al. 2011). Another study suggested that a subset of malignant melanomas that showed complex genomic rearrangements suggestive of chromothripsis were associated with a poor prognosis (Hirsch, Kemmerling et al. 2013).

Telomere dysfunction is another mechanism that has been demonstrated in cancers and believed to contribute to CIN in conjunction with the two models described above (Murnane 2012). In normal cells, telomeres are short repeat sequences with associated proteins present at the ends of chromosomes to prevent them from being seen by the DNA repair mechanisms as Double Strand Breaks (DSBs). They also participate in cellular aging as their progressive shortening and attrition eventually leads to senescence (Jefford and Irminger-Finger 2006). As demonstrated in STS, telomere loss leads to erroneous chromosomal end joining and a series of bridge–fusion-breakage cycles (Figure 1.2) that result in aneuploidy, chromosomal translocations, amplifications or deletions (Artandi, Chang et al. 2000, Artandi and DePinho 2000, Gisselsson, Pettersson et al. 2000, Murnane 2012). While telomere loss promotes the initiation of genomic instability, tumour cells have been shown to subsequently activate telomerase maintenance mechanisms such as telomerase and alternative lengthening of telomeres in order to remain immortal (Begus-Nahrmann, Hartmann et al. 2012). Abnormal maintenance of telomere length has been demonstrated in various STS subtypes (Montgomery, Argani et al. 2004, Johnson, Gettings et al. 2007, Heaphy, Subhawong et al. 2011) and clinical trials evaluating telomere maintenance as a therapeutic target in cancer are on-going (Xu, Li et al. 2013).

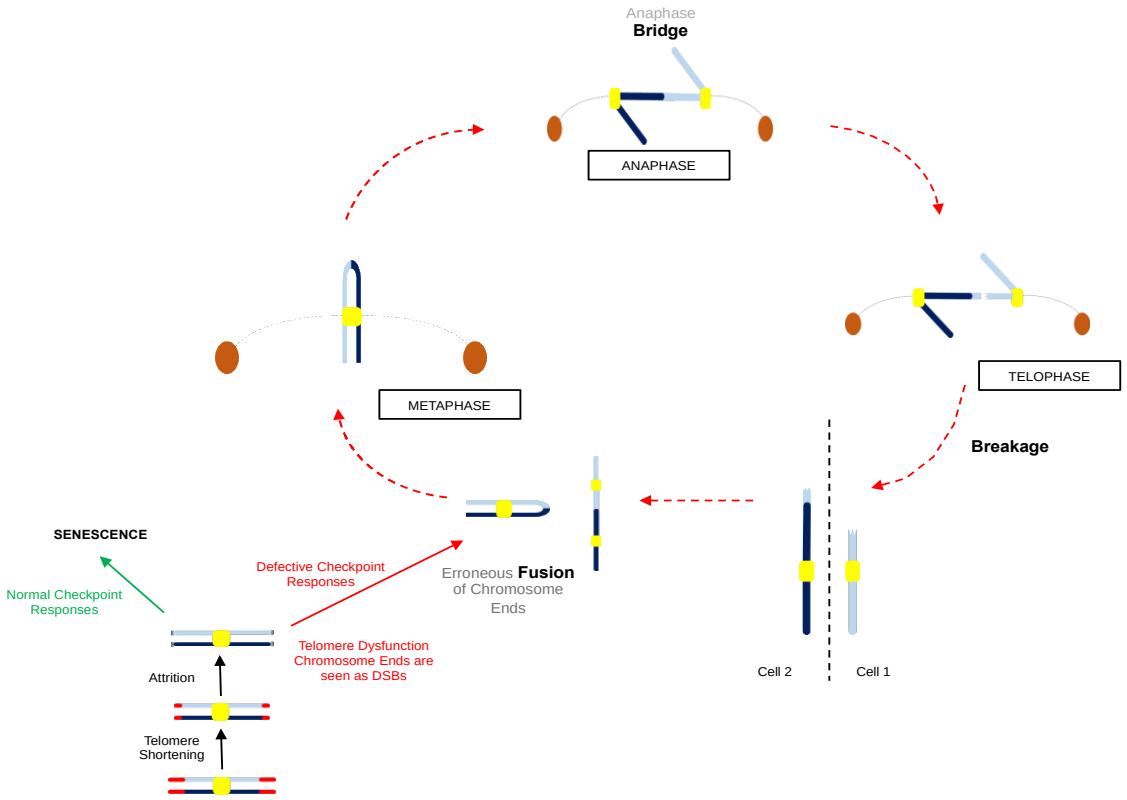

Figure 1.2: Illustration of Breakage–Fusion–Bridge (BFB) Cycle contribution to Chromosomal Instability

Telomere Dysfunction and Defective Cell Cycle Checkpoint Responses in tumour cells lead to erroneous fusion of chromosome ends that form a bridge between chromatid ends during anaphase with subsequent breakage of one of the chromatids in telophase. Following cytokinesis, one daughter cell contains an extra copy of the affected chromosome segment (amplification) with a corresponding deletion in the other daughter cell. The ends of the abnormal chromosomal appear as double strand breaks (DSBs), which may become fused and enter another BFB cycle.

1.10 Copy Number Mapping in Cancer

Copy number aberrations are an important pathogenetic mechanism in cancer and understanding their patterns can lead to identification of therapeutic targets (Taylor, Barretina et al. 2008). High-resolution mapping of copy number aberrations in cancer genomes is a valuable way of identifying recurrent genomic changes. Coupled with epigenetic, expression and functional data, this can further our understanding of the molecular basis of cancers and markers can be identified that can be targeted for tumour diagnosis, prognostic sub-classification or pharmacologic/biologic therapy (Taylor, Barretina et al. 2011). To this end, large-scale projects to catalogue genomic copy number aberrations in large cohorts of specific cancers e.g. the Cancer Genome Atlas Project have been established and over time have yielded important insights in a number of cancers, including glioblastoma multiforme and ovarian carcinoma (The Cancer Genome Atlas Research Network 2008, Mankoo, Shen et al. 2011, The Cancer Genome Atlas Research Network 2011).

Oligonucleotide micro-array based comparative genomic hybridization (array CGH or aCGH) is a high-resolution technique that is commonly utilised for the simultaneous detection of all SCNA across the entire genome. It is therefore highly suited for the study of certain STS subtypes, which are characterised by multiple complex SCNA as it allows the detection, accurate mapping and subsequent analysis of aberrations with a view to identification of pathogenetic abnormalities.

1.11 Hypothesis and Study Aims

A large proportion of STS are characterised by genomic instability that is evidenced by pervasive, multiple seemingly random genomic copy number aberrations. Aberrations that occur with a higher frequency than can be explained by random chance is likely to have been specifically targeted by tumours for a survival benefit. The following hypothesis was therefore developed for this PhD project:

"***Soft Tissue Sarcomas that are characterized by complex karyotypes bear genomic abnormalities that are subtype-specific, but as yet unidentified"***

Further, strong associations exist between STS and DNA-damaging ionising radiation and inheritable conditions where DNA-repair mechanisms are defective. It is therefore plausible that defects in cell cycle regulatory and DNA repair genes are important in STS pathogenesis.

This PhD Project therefore specifically aims:

1. To carry out high-throughput genome-wide mapping of somatic copy number aberrations (SCNA) in primary soft tissue sarcomas using Oligonucleotide Array Comparative Genomic Hybridisation

2. To identify patterns of SCNA that are specific for STS subtypes as well as those involved in DNA Repair and Cell Cycle Regulation and Mesenchymal Cell Differentiation

3. To develop a suitable disease model for the functional study of candidate pathogenetic SCNA and affected genes

4. To further characterise candidate pathogenetic SCNA and affected genes in terms of protein expression and functional consequences

CHAPTER TWO

MATERIALS AND METHODS

2.1 PATIENTS AND TUMOUR SAMPLES

2.1.1 Ethics Statement

National Research Ethics Committee approval was obtained for the collection and use of fresh and archival tissue samples (reference numbers 09/H1313/52 and 09/H1313/30, respectively). Patients gave written informed consent prior to the collection of fresh tissue samples and all data from archival samples was analysed anonymously. Tumour and normal tissue was collected and stored according to the principles of the Declaration of Helsinki and the use of tissue was in compliance with the Human Tissue Act, 2004.

2.1.2 Tumour Samples

Tumour samples were collected as fresh specimens and/or archival FFPE blocks from STS cases collected between 1994 and 2012 at the Histopathology Department of the Royal Hallamshire Hospital, Sheffield, UK. Details of all 88 STS cases included in this study are summarised in Appendix 1.

2.1.2.1 Fresh Tumour Samples

Fresh tumour samples were obtained from patients receiving surgical treatment for STS at Sheffield Teaching Hospitals within 30 minutes of resection. STS cases comprised various subtypes and tumours were macroscopically sampled by an experienced sarcoma pathologist prior to fixing in 10% neutral-buffered formalin (NBF). Samples for tissue culture were collected in sterile phosphate-buffered saline (PBS) and additional tumour samples were snap-frozen and stored in liquid nitrogen or at -80ºC until DNA extraction or other molecular analysis. Normal tissue where available, was obtained from tumour-free margins of excised specimens and snap-frozen in liquid nitrogen.

2.1.2.2 FFPE Samples

Representative FFPE tumour sections stained by routine haematoxylin and eosin (H+E) were examined by an experienced sarcoma pathologist and areas comprised of >70% tumour cells were outlined. Where necessary, the outlined areas were then scraped off subsequent 20µm sections for DNA extraction, or specifically targeted on 5µm sections for Fluorescence In-situ Hybridisation and/or immunohistochemistry.

2.2 MATERIALS

2.2.1 General Laboratory Reagents and Equipment

General laboratory reagents and equipment used in this study are summarised in Appendix 2. Method-specific reagents are described in subsequent sections.

2.2.2 Tissue Culture and Chromosome Preparation

Culture Medium: Primary STS tumour cells were maintained in RPMI 1640 culture medium (Lonza), supplemented as outlined in table 2.1 below:

Table 2.1: Tissue culture medium supplements

Supplement	%(v/v)
Fetal Calf Serum (Lonza)	20%
L-Glutamine (200mM in 0.85% NaCl; Lonza)	1%
Amphotericin B (Lonza)	1%
Penicillin/Streptomycin (10kU/ml; Lonza)	1%
D+ glucose (45% solution, Sigma)	0.4%

Media was then stored at 4^{o}C and warmed to 37^{o}C prior to use.

Trypsin-EDTA: 0.4% Trypsin-EDTA solution (Lonza-BioWhitaker) was stored at -20^{o}C in 50ml aliquots then thawed gently in a 37^{o}C incubator prior to use.

Phosphate-buffered Saline: Sterile Dulbecco's Phosphate Buffered Saline (Lonza-BioWhitaker) containing 0.0095M PO_4 without Ca^{2+} or Mg^{2+} was stored at room temperature.

Colcemid: 10µg/ml KaryoMAX® Colcemid (Gibco), was stored at 4ºC

Hypotonic Solution: 0.075M potassium chloride (KCl) was prepared by dissolving 2.235mg KCl in 400ml dH_2O, autoclaved then stored at 4ºC and warmed to 37ºC prior to use.

Fixative Solution: A mixture of methanol and glacial acetic acid in a ratio of 3:1 respectively was prepared and used within 30 minutes.

2.2.3 Giemsa Banding

Sorenson's Buffer: 9.47g of disodium hydrogen orthophosphate (Na_2HPO_4) and 9.08g potassium di-hydrogen orthophosphate (KH_2PO_4) made up to 1000ml with ddH_2O and stored at room temperature for up to one month.

Banding Trypsin: 0.6g of Trypsin 1:250 powder (Difco) in 250ml of Sorensen's buffer then stored in 10ml aliquots at -20ºC.

Gurr's Buffer: One Gurr's buffer tablet (BDH) was dissolved in 1000ml dH_2O and adjusted to pH 6.8 then stored at room temperature.

Giemsa (Leishman) Stain: Leishman's Stain (BDH) and Gurr's buffer were mixed in a 1:5 ratio and used within 15 minutes of preparation

2.2.4 DNA Extraction

Sodium thiocyanate: 1M NaSCN (Sigma) was prepared by dissolving 12.3g NaSCN powder in distilled water (dH_2O) and stored at room temperature for up to one month

DNA Extraction Kit: DNeasy® Blood and Tissue Kit (Qiagen) comprising

- Proteinase K
- Tissue Lysis Buffer (Buffer ATL),
- Lysis Buffer (Buffer AL),
- Wash Buffers (Buffers AW1 and AW2),
- Elution Buffer (Buffer AE)
- DNeasy® mini spin columns and collection tubes.

2.2.5 Agarose Gel Electrophoresis

Agarose: Agarose powder (Fisher) was stored at room temperature and gels were prepared just before use.

Running Buffer: 1X TAE (Tris-Acetate-EDTA) was diluted from 50X stock solution prior to use. Stock solution comprised 242g Tris base; 57.1ml Glacial Acetic Acid and 18.6g EDTA dissolved in 1000ml dH_2O and stored at room temperature.

Electrophoresis Unit: Multi sub choice electrophoresis unit (Geneflow), comprises

- Gel casting tray (15 × 7cm)
- Sample comb (for 20 samples)
- Electrophoresis tank

Power Source: Power-Pac 3000 basic power supply for electrophoresis (BioRad)

Ethidium bromide: 10mg/ml Ethidium bromide prepared by dissolving 1g ethidium bromide in 100ml dH_2O and stored at 4ºC

DNA Ladder: 1kB DNA ladder (Promega) stored at 4ºC

Loading Buffer: 6X Loading buffer - 0.25% (w/v) bromophenol blue and 30% (v/v) glycerol prepared by adding 25mg bromophenol blue to 3ml glycerol and making up to 10ml with dH_2O then stored at 4ºC

2.2.6 Array Comparative Genomic Hybridisation

Restriction Digestion Enzymes: Restriction enzymes and other reagents (Promega)

- *Alu 1* (10 U/µl)
- *Rsa 1* (10 U/µl)
- 10X Buffer C
- Acetylated Bovine Serum Albumin (10µg/ml)

All stored at -20°C.

Random Priming with Exo-Klenow Labelling Kit: Agilent Genomic DNA Enzymatic Labelling Kit (Agilent) comprising

- Random Primers,
- 5X Buffer,
- 10X dNTP,
- Cyanine 3-dUTP (1.0 mM),
- Cyanine 5-dUTP (1.0 mM), and
- Exo-Klenow Fragment

All stored at -20ºC.

Labelled DNA Purification: Amicon® Ultra 0.5ml 30kDA filters and 1.5ml microfuge collection tubes (Millipore)

ULS Labelling Kit: Agilent ULS Labelling Kit (Agilent) stored at 4ºC comprising

- ULS-Cy 3 reagent
- ULS-Cy 5 reagent
- 10X Labelling solution
- Agilent-KREA*pure*® purification columns with collection tubes

Cot-1 DNA: 1mg/ml Cot-1 DNA Pre-Hybridisation (Invitrogen) stored at -20ºC

Hybridisation Kit: Oligo aCGH Hybridisation Kit (Agilent) comprising

- 2X Oligo aCGH Hybridisation solution
- 10X Blocking Agent

Stored at room temperature.

Blocking Solution: CGH*block*® (Agilent) stored at -20ºC

Hybridisation Assembly: Microarray hybridisation assembly comprising

- SurePrint® G3 Human CGH Microarray Slide - 4 × 180K
- Hybridisation Gasket Slide
- Hybridisation Chamber Kit - SureHyb® enabled, Stainless Steel

All obtained from Agilent, Stockport, UK.

Hybridisation Oven: Microarray Hybridisation Oven (Agilent) equipped with removable rotator rack

Wash Buffer Kit: Oligo aCGH/ChIP-on chip Wash Buffer Kit comprising

- Oligo aCGH Wash Buffer 1, and
- Oligo aCGH Wash Buffer 2

Scanner: SureScan High-Resolution Microarray Scanner (Agilent)

2.2.7 Fluorescence In-situ Hybridisation

Commercial Probes: Commercial FISH probes (Vysis) were stored at -20°C in the dark. The commercial probes used in this study are summarised in Table 2.2 below.

Table 2.2: Commercial FISH Probes used in this study

Probe Name	Target	Locus	Fluorophore	Volume for 5 slides (Efficiency)*
Vysis LSI ATM	*ATM*	11q22.3	SpectrumOrange®	1.5µl
Vysis LSI 13 (RB1)	*RB1*	13q14	SpectrumGreen®	1.5µl
CEP 3	*Chr 3 Centromere*	3p11.1-q11.1	SpectrumOrange®	1.0µl
CEP 8	*Chr 8 Centromere*	8p11.1-q11.1	SpectrumOrange®	1.0µl
CEP 11	*Chr 11 Centromere*	11p11.11-q11	SpectrumGreen®	1.0µl

* - based on user experience within the Rare Tumour Research Group

Bacterial Artificial Chromosome (BAC) Clones: Clones were obtained from two BAC libraries (Invitrogen) as *E. coli* bacterial glycerol stocks and stored at -80ºC. Their characteristics are summarised on Table 2.3 below.

Table 2.3: BAC Clones used in this study

Clone ID	Target	Locus	Library	Size	Vector	Host
RP11-586K23	*TRIO (5' end)*	5p15.2	RPCI-11	219kB	pBACe3.6	DH10B *E. coli*
CTD-2505B6	*TRIO (3' end)*	5p15.2	CITB-Human D1	183kB	pBeloBAC11	HS996 *E.coli*

Bacterial Culture Media: Luria-Bertani (LB) media (Sigma) was prepared as solid media or broth according to manufacturers' instructions and stored at room temperature

Selection Antibiotic: Chloramphenicol (Sigma) stored in stock concentration of 20mg/ml at 4ºC in the dark and added to LB media at a working concentration of 12.5µg/ml.

Solution P1: prepared as outlined below, filter-sterilized and stored at 4°C

- 50mM Tris, pH 8
- 10mM EDTA
- 100µg/ml RNase A

Solution P2: prepared as outlined below, filter sterilized and stored at room temperature

- 0.2N NaOH
- 1% SDS

Solution P3: 3M Potassium Acetate (KOAc) at pH 5.5; autoclaved and then stored at 4°C.

BAC DNA Labelling Kit: Nick Translation Kit (Abbott Molecular) stored at -20°C, comprising

- Nick translation enzyme mix
- 10X nick translation buffer
- dATP, dCTP, dGTP, and dTTP solutions

- Nuclease-free water

Labelled Nucleotides: 1mM SpectrumGreen-labelled dUTP (Abbott Molecular) diluted to 0.2mM with nuclease-free water and stored in 10µl aliquots at -20ºC in the dark.

Hybridisation buffers: LSI/WCP and CEP Hybridisation buffers (AbbottMolecular) stored at -20ºC

Tissue Pre-treatment: Zymed 'Spotlight' Tissue Pretreatment Kit (Invitrogen) stored at 4ºC, comprising

- Pretreatment solution (pH 7.0), and
- Enzyme reagent

Counterstain: VECTASTAIN® mounting medium with DAPI (4', 6-diamidino-2-phenylindole) obtained from Abbott Molecular and stored in the dark at 4°C

SSC Buffer: 20X Saline Sodium Citrate (SSC) stock prepared as from components outlined below in 1L dH_2O, adjusted to pH 7.0 and stored at room temperature

- 3M NaCl
- 300mM $NaHCO_3$

SSCT1 - 0.4xSSC/0.3% Tween 20 (200ml 2XSSC + 1.5ml Tween 20 made up to 500ml with deionised water)

SSCT2 - 2XSSC/0.1% Tween 20 (500ml 2XSSC + 0.5ml Tween 20)

PBS with MgCl2 - PBS + 50mM MgCl2 (50ml 1M MgCl2 + 950ml PBS) stored at room temperature for up to one month

Pepsin – Pepsin (Sigma) stored at -20ºC in 25µl aliquots at 100mg/ml stock concentration and diluted 1:2000 in 0.01N HCl for working concentration.

Fixative Solution – 37% Formaldehyde (Sigma) stored at room temperature in the dark and diluted to 1% in PBS/MgCl2 as working solution

2.2.8 Immunochemistry

Peroxidase Quenching Solution: 0.3% Hydrogen peroxide (H_2O_2) freshly prepared by diluting stock 30% H_2O_2 stock solution (Sigma) in 100% methanol.

Blocking Serum: Normal horse serum (Vector) was stored at 4°C and diluted in PBS to give a 10%working concentration

Tris-EDTA Buffer: 1.21g Tris base and 0.37g EDTA in 1L dH2O, adjusted to pH 9.0 + 0.5ml Tween 20

Citrate Buffer: 1.92g anhydrous citric acid dissolved in 1L dH2O adjusted to pH 6.0 + 0.5ml Tween 20

Primary Antibodies: Primary antibodies were stored at 4°C and diluted in 2% normal serum (diluted in PBS). The various antibodies used in this study are summarised in Table 2.4 below.

Table 2.4: A summary of primary antibodies used in this study and the corresponding reagents

Antigen (Ag) Target	Manufacturer	Source	Type	Class	Working Dilution	Blocking Serum	Ag Retrieval Buffer
Trio	Santa Cruz Sigma	Goat Rabbit	Polyclonal Polyclonal	IgG IgG	1:150 1:50	Horse	Tris-EDTA (pH 9.0)
cKIT	Dako	Mouse	Monoclonal	IgG	1:200	Horse	Citrate (pH 6.0)
α-smooth muscle actin	Dako	Mouse	Monoclonal	IgG	1:100	Horse	N/A

Secondary Antibodies: Secondary antibodies (Vector) were stored at 4°C and diluted 1:200 in 2% blocking serum to give working solution.

Avidin/Biotin Peroxidase Kit: VECTASTAIN Elite ABC Kit (Vector) stored at 4ºC and used according to manufacturers' instructions

Peroxidase Substrate Kit: DAB Peroxidase Substrate Kit (Vector) stored at 4ºC and used according to manufacturers' instructions

DePex: DPX mountant (Sigma) stored at room temperature and used in a fume hood

2.2.9 Cell Proliferation Assay

Protein Inhibitor: ITX3 (Sigma), a specific inhibitor of the n-terminal domain of the Trio RhoGEF protein dissolved to 1mM stock concentration in DMSO and stored at room temperature.

Tetrazolium reagent: 3- (4, 5-Dimethylthiazol-2-yl)-2, 5-diphenyltetrazolium bromide (MTT) obtained from Sigma was dissolved to 1mg/ml stock concentration in sterile PBS and stored at 4ºC.

2.2.10 Cell Migration/Invasion Assay

Boyden Chamber: 48-well micro chemotaxis chamber AP-48 (Neuroprobe)

Filter Membrane: Polycarbonate track etch membrane with 8µm pore size (Neuroprobe) was used for all cell lines

Extracellular Matrix Mix: Extracellular matrix (ECM) components were diluted and stored at 4ºC. The ECM mix was freshly prepared just before use as summarised in Table 2.5 below.

Table 2.5: Extracellular Matrix Mix Components

ECM Protein	Source	Manufacturer	Stock Concentration	Working Concentration (diluted in PBS)
Collagen	Human Placenta	Sigma	2.0mg/ml (in 0.5M Acetic Acid)	15µg/ml
Fibronectin	Human Fibroblasts		2.0mg/ml (in sterile dH_2O)	
Laminin	Human Fibroblasts		1.0 mg/ml (in sterile PBS)	

2.3 METHODS

2.3.1 Tissue Culture

2.3.1.1 Fresh Tumour Processing

STS tumour specimens were obtained as described in Section 2.1.2. A small fresh tissue sample (typically around 10mm × 10mm × 5mm) was placed in a sterile petri dish with a few drops of culture media and minced with a sterile scalpel until very fine. An additional 10ml culture media was then added and the suspended tissue was transferred to a 25ml universal tube and centrifuged at 1000rpm for 5 minutes at room temperature. The supernatant was then removed, the tissue re-suspended in fresh media and transferred to T_{25} flasks and 1ml slopes, which were subsequently placed in a 5% CO_2 incubator at 37^{o}C in 95% humidified air (37°C/5%CO_2 incubator). Media in the flasks and slopes was changed as required.

2.3.1.2 Sub-cultures

Washes were set up when it appeared that some viable cells remained unattached to the flask or slope. The media in the culture was transferred to a 25ml universal tube, and replaced with warm fresh media. The tube was then centrifuged at 1000rpm for 5 minutes and the supernatant removed. The pellet was re-suspended in warm fresh media and transferred to a sterile T_{25} flask, which was placed in the incubator.

Cells were generally sub-cultured when they reached 80 – 90% confluence to provide room for further growth. For a T_{25} flask, the media was removed and the cells rinsed twice with sterile PBS. Trypsin (2.5mls) was then added and the cells incubated at 37^{o}C and monitored on the inverted microscope until dissociated. Trypsinisation was then halted by adding 2.5ml of fresh culture media. The cells (now suspended) were transferred to a 25ml universal tube and centrifuged at 1000rpm for 5 minutes at room temperature. The supernatant was removed, the cell pellet re-suspended in fresh media and split among the appropriate number of flasks, where the media was made up to 5mls before being placed in the incubator. For cells that were dividing faster, cultures were transferred to T_{75} flasks and sub-cultured as described using twice as much media and trypsin at each step.

2.3.2 DNA Extraction

Qiagen DNeasy® Blood and Tissue Kit was used to extract DNA from tissues and cultured cells. Briefly, the kit works on the principles of alkaline lysis, binding, washing and elution. The cells or tissues are lysed then loaded onto a silica-based membrane, which selectively adsorbs DNA in the presence of high concentrations of chaotropic salts. Centrifugation and two wash steps serve to remove contaminants and the pure DNA is then eluted.

2.3.2.1 Cell/Tissue Preparation and Lysis

FFPE Tissue

Tissue was macro-dissected where necessary as described in Section 2.1.2.2. Approximately 4mm^3 of tissue (the equivalent of two 20µm-thick sections measuring 10mm × 10mm) was placed in a 1.5ml microfuge tube , then 480µl PBS and 20µl 10% Tween 20 (diluted in nuclease free water) were added. The tube was then placed on a heat block at 90ºC for ten minutes to de-paraffinise the tissue.

This was followed by centrifugation at 10,000 × *g* for 15 minutes at room temperature and immediate transfer to ice for 2 minutes. The wax disc formed on the surface was carefully removed with tweezers and the supernatant discarded without disturbing the pellet. Absolute ethanol (1ml) was added followed by centrifugation at 10,000 × *g* for 5 minutes at room temperature. The ethanol was removed and the lid left open to allow the pellet allowed to air-dry completely.

Cross-links (DNA:DNA and DNA:protein) were then broken by adding 1M NaSCN (400µl) to each sample tube and overnight incubation at 37°C. Samples were then centrifuged at 10,000 × *g* for 20 minutes at room temperature and the supernatant carefully removed. PBS (400µl) was added followed by centrifugation at 10,000 × *g* for a further 20 minutes at room temperature and removal of the supernatant. Buffer ATL (360µl) and Proteinase K (40µl) were then added and the sample incubated on a heat block at 56^{o}C for at least 24 hours, with periodic vortexing and addition of fresh proteinase K every 6 to 8 hours until the tissue was completely dissolved.

The sample was then allowed to cool to room temperature and RNase A (8µl) was added and briefly vortexed then incubated at room temperature for 2 minutes. Buffer AL (400µl) was added and the tube incubated on a hot block at 70°C for 10 minutes. Absolute ethanol (440µl) was then added to the tube and its contents transferred to 2 separate labelled DNeasy® mini spin columns (about 660µl each).

Fresh Frozen Tissue

Frozen Tissue (10 – 20mg) was transferred to a clean nuclease-free 1.5ml microfuge tube and Proteinase K (20µl) and Tissue Lysis Buffer ATL (180µl) added and mixed by vortexing. The tube was then incubated on a heat block at 56^{o}C for up to 24 hours with periodic vortexing and addition of fresh Proteinase K every 6-8 hours until the tissue was completely dissolved. A mixture of Buffer AL (200 µl) and absolute ethanol (200µl) was then added to each sample tube, immediately vortexed and the sample transferred to a single labelled DNeasy® mini spin column.

Cultured Cells

Cultured cells were dissociated with trypsin and centrifuged as described in Section 2.3.1.2. The cells were then re-suspended in PBS at a density of approximately 2 × 10^{7}cells/ml. 200µl of the cell suspension was then transferred to a 1.5ml microfuge tube. Proteinase K (20µl) and Buffer AL (200µl) were added, the mixture vortexed and incubated at 56^{o}C on a heat block until the tissue was completely dissolved (usually within 30 minutes). Absolute ethanol (200µl) was then added to each sample tube, vortexed and the contents transferred to a single labelled DNeasy® mini spin column.

2.3.2.2 DNA Adsorption

The DNeasy® mini spin column loaded with lysate (FFPE tissue, fresh tissue or cultured cells) was then placed in a collection tube and centrifuged at 6,000 × *g* for 1 minute. The collection tube and contents were discarded and the spin column placed in a fresh collection tube.

2.3.2.3 Washing

Buffer AW1 (500µl) was added to the each spin column and centrifuged at 6,000 × *g* for 1 min. The flow-through and collection tube were again discarded and the spin column transferred to a fresh collection tube.

The subsequent wash step depended on the source of DNA. For fresh frozen tissue or cultured cells, buffer AW2 (500µl) was added, while for FFPE samples it was substituted with an equivalent volume of 80% ethanol. Samples were then centrifuged at 16,000 × *g* for 3 minutes.

2.3.2.4 DNA Elution

DNA from fresh frozen tissue or cultured cells was eluted in Buffer AE or 1X TE, while that from FFPE tissue was eluted in nuclease-free water. The volume of eluent used depended on the desired DNA yield and concentration. In general, smaller eluent volumes gave higher eluate DNA concentrations, but lower overall DNA yields and *vice versa*. Eluent volumes therefore ranged between 50µl (for FFPE samples) and 200µl (for cultured cells).

Following the second wash step, spin columns were placed in a fresh, labelled 1.5ml microfuge tube. An appropriate volume of eluent was added and then incubated upright for 1 minute at room temperature followed by centrifugation at 6,000 × *g* for 1 minute. The elution was repeated in the same or a fresh microfuge tube if an even larger overall DNA yield was required. The spin column was then discarded and the microfuge tube containing eluted DNA stored at 4°C (in Buffer AE or 1X TE) or -20°C (in nuclease free water).

2.3.3 DNA Quantification and Purity Assessment

DNA was quantified and assessed for purity by UV-VIS Spectrophotometry using the NanoDrop® ND-1000 (ThermoScientific) instrument set-up for measurement of double-stranded DNA. The instrument automatically calculates DNA concentration from small sample volumes (as little as 1µl) using a modified Beer-Lambert equation:

$$c = (A * \varepsilon)/b$$

Where, c is DNA concentration in ng/µl; A is absorbance at 260nm, ε is the wavelength-dependent absorbance coefficient (50ng·cm/µl for double-stranded DNA) and b is the path length in cm. The ratios of absorbance at 260nm *vs.* 280nm (A_{260}/A_{280}) and 260nm *vs.* 230nm (A_{260}/A_{230}) were used to assess DNA purity. Generally, A_{260}/A_{280}values between 1.80 and 2.00 indicated the absence of protein contamination, while A_{260}/A_{230} values greater than 1.90 indicated the absence of contamination by other organic compounds.

Briefly, the optical surfaces of the Nanodrop® instrument were cleaned with lint-free wipes and 1 - 2µl of nuclease-free water loaded to initialize the instrument. Similar volumes of DNA eluent and samples were then loaded to serve as a blank and for measurement respectively, cleaning the optical surfaces between samples. DNA samples that were of low concentration were vacuum-centrifuged to the desired concentration, while those samples with low A_{260}/A_{230} were re-precipitated for purification (Section 2.3.4)

2.3.4 DNA Precipitation

Sodium Acetate-Ethanol precipitation was used to purify DNA samples that had low A_{260}/A_{230} values indicative of contamination with organic substances such as guanidine or ethanol that could potentially inhibit DNA labelling reactions. Precipitation was also used to clean labelled BAC DNA for FISH experiments (Section 2.3.8.4).

The DNA sample was placed in a 1.5ml microfuge tube and 3M sodium acetate (0.1 volumes) added, followed by absolute ethanol (2.5 volumes). The reaction was then mixed well by pipetting up and down and incubated at -80ºC for 30 minutes. The sample was then centrifuged at 12,000 × *g* for 30 minutes at -4ºC. The supernatant was carefully removed leaving the DNA pellet intact and 70% ethanol (200µl) carefully added and then removed to wash the pellet (this step was omitted for BAC DNA). The tube was then left open under a laminar flow hood to allow residual ethanol to evaporate completely at room temperature. The precipitated DNA was then reconstituted with 50µl of 1X TE and its concentration and purity re-assessed as described (Section 2.3.3).

2.3.5 Agarose Gel Electrophoresis

Agarose gels used in this study were prepared in 1% concentration by dissolving agarose (1g) in 1X TAE (100ml) in a conical flask and heating until fully dissolved in a microwave oven. The solution was t allowed to cool for about 3 minutes before addition of Ethidium bromide (5µl), pouring into the gel-casting tray with combs in position and then left to set as it cools. The comb was subsequently removed and the gel submerged in an electrophoresis tank containing 1X TAE as running buffer.

Each DNA sample (5µl) was mixed with 6x loading buffer (1µl) and then loaded into a single noted well of the agarose gel. Standard 1kB DNA ladder (5µl) was also loaded into a separate well for DNA fragment size determination. Electrophoresis was carried out at 90V for 90 minutes and bands were visualized using a UV trans-illuminator and photographed.

2.3.6 Array-based Comparative Genomic Hybridisation

Oligonucleotide Array-based Comparative Genomic Hybridisation was used to detect copy number aberrations in tumour DNA. Briefly, tumour and matched reference gDNA were labelled with different fluorescent dyes and co-hybridised to an array of 180,000 oligonucleotide probes covering the entire genome at high resolution. The array was then washed and scanned, the differential intensity of the fluorescent dyes at each probe serving as a surrogate for the ratio of copy numbers of probe sequence in the tumour vs. reference genome. The main steps of this method are shown schematically in Figure 2.1.

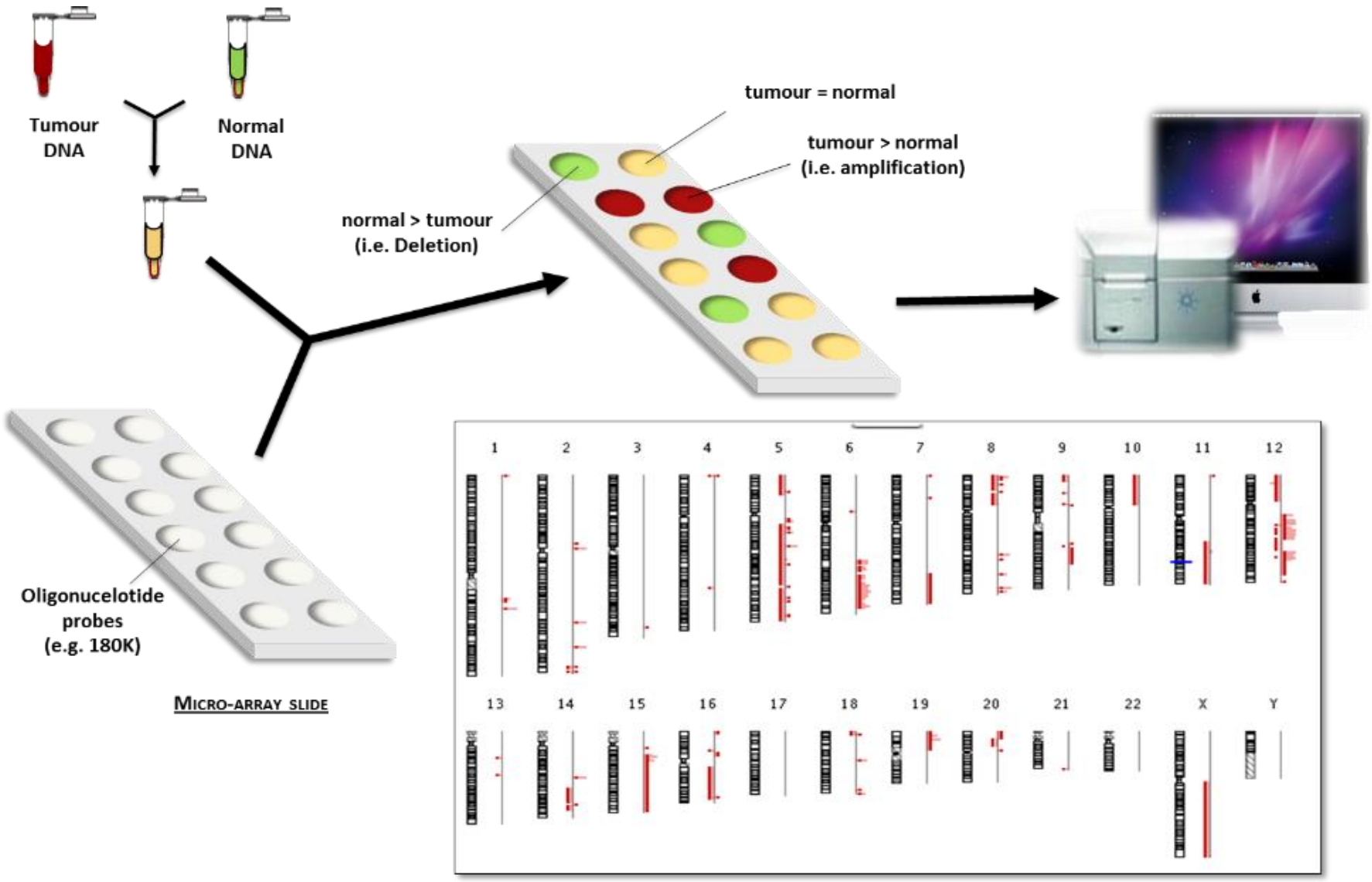

Figure 2.1: Schematic Representation of Oligonucleotide Microarray-based Genomic Hybridisation

2.3.6.1 DNA Labelling (Random Priming with Exo-Klenow)

Whole genomic DNA obtained from fresh frozen tissue, cultured cells or commercially available reference DNA was labelled by random priming with Exo-Klenow (enzymatic method) carried out according to manufacturers' instructions. Briefly, genomic DNA is digested at specific restriction sites, denatured and then using random primers, a DNA polymerase (Exo-Klenow fragment of E. coli DNA polymerase) is used to amplify the DNA while fluorophore-labelled dUTP nucleotides are simultaneously incorporated at the 3' ends of newly-synthesized DNA fragments. Any unincorporated fluorescent nucleotides are then removed from the system in order to minimise background fluorescent signal in array experiments.

For a typical aCGH experiment, 0.5 - 1µg of each DNA sample was labelled. Samples of tumour DNA were matched with equivalent amounts of normal DNA (either extracted from normal tissue from the same patient or commercial reference DNA) and labelled simultaneously in separate reactions. Prior to labelling, the concentration and purity of DNA samples were determined as described in Section 2.3.3.

Restriction Digestion

For each 4 × 180K microarray, 8 DNA samples (4 tumour and 4 matched normal) of volumes equivalent to the starting amount of DNA were placed in clean 0.2ml thin-walled PCR tubes and made up to 20.2µl with nuclease free water, where necessary.

A Digestion Master Mix, sufficient for 10 reactions was then made up as outlined on Table 2.6 below.

Table 2.6: Restriction Digestion Master Mix Components

	Per tube (µl)	*For 10 tubes (µl)
Nuclease-free Water	2.0	20
10X buffer	2.6	26
Acetylated BSA (10µg/µl)	0.2	2
Alu-1 (10µg/µl)	0.5	5
Rsa-1 (10µg/µl)	0.5	5
Total	5.8	58

To each reaction tube, Digestion Master Mix (5.8µl) was then added, making the volume up to 26µl and the tubes transferred to a thermocycler with a heated lid, programmed for incubation at 37ºC for 2 hours for the digestion reaction, followed by 65ºC for 20 minutes to kill the enzymes. The reaction tubes were then held at 4ºC until ready for labelling.

Enzymatic Labelling Reaction

Random Primers (5µl) were added to each reaction tube (containing restriction digested DNA), making the volume up to 31µl and incubated at 95ºC for 3 minutes, then held at 4ºC for at least 5 minutes.

During this incubation step, Labelling Master Mixes were prepared using Cy3-dUTP and Cy5-dUTP in separate tubes, on ice and in dim conditions, as outlined on Table 2.7 below. Generally, tumour DNA was labelled with Cy5 while normal DNA was labelled with Cy3. However, fluorophores may be swapped provided tumour and normal DNA are differentially labelled and the fluorophore used for each is noted.

Table 2.7: Exo-Klenow Labelling Master Mix Components

	Per tube (µl)	***For 5 tubes (µl)**
5X buffer	10.0	50
10X dNTP	5.0	25
Cy3-dUTP or Cy5-dUTP	3.0	15
Exo-klenow fragment	1.0	5
Total	19.0	95

The reaction tubes were briefly centrifuged and the appropriate labelling master mix (19µl) was added to each tube to make a total volume of 50µl and mixed well by pipetting up and down. The tubes were then transferred to the thermocycler where they were incubated at 37ºC for 2 hours for the labelling reaction then 65ºC for 10 minutes to inactivate the enzyme. Samples were then held at 4ºC until clean up.

Clean-up of Exo-Klenow Labelled DNA

Unincorporated fluorophore-labelled nucleotides and other reaction components were removed using Amicon® 30kDa filters, which trap the labelled genomic DNA fragments on the basis of their large size. Each labelling reaction was mixed with 1X TE (430µl) and transferred to a labelled Amicon® filter placed in a collection tube and centrifuged at 14, 000 *x g* for 10 minutes at room temperature. The flow-through was removed and discarded and further 1X TE (480µl) was added to each filter followed by a second centrifugation at 14, 000 *x g* for 10 minutes. The flow-through and collection tube were discarded and the filters inverted into fresh collection tubes.

The inverted filters were centrifuged at 1, 000 *x g* for 1 minute at room temperature to yield a flow-through volume of approximately 21µl clean labelled DNA. When necessary, sample volumes were adjusted to 21µl by addition of 1X TE or vacuum centrifugation).

2.3.6.2 DNA Labelling (Universal Linkage System)

The Universal Linkage System (ULS) is a chemical reaction, which directly incorporates platinum-conjugated fluorescent dyes into DNA molecules. Optimised for use in FFPE DNA, it was used to label tumour and matched normal DNA for array CGH. Starting amounts of tumour DNA ranged between 0.5 and 1µg (usually 0.8µg), depending on the amount of DNA available and this was matched with equivalent amounts of normal DNA (as with Exo-Klenow labelling).

Heat Fragmentation

All DNA samples were visualised on agarose gels (Section 2.3.5) to determine average fragment size. For samples with relatively large DNA fragments e.g. commercial reference DNA or some FFPE DNA samples, heat fragmentation was required prior to labelling.

For a single 4 × 180K array experiment, the corresponding volumes of eight DNA samples were placed in thin-walled 0.2ml PCR tubes and made up to 8µl with nuclease-free water. Fragmentation was then carried out at 95ºC and the duration of incubation depended on the average DNA fragment size, as summarised on Table 2.8 below. Incubation was carried out on a thermocycler with a heated lid and the reaction tubes were subsequently held at 4ºC until ready for labelling.

Table 2.8: Duration for Heat Fragmentation by DNA source and Fragment size

Average DNA Fragment Size (Kb)	DNA source	Incubation Time (at 95ºC)
>10.0	Commercial DNA Fresh frozen tissue	10 minutes
7.0 – 10.0	Some FFPE tissues	5 minutes
<7.0	Most FFPE tissues	No fragmentation

ULS Labelling Reaction

During the heat fragmentation step, ULS-Cy3 and ULS-Cy5 dye master mixes were prepared on ice and in dim conditions. The volume of dye used per reaction depended on the amount of DNA in the reaction. For most experiments, about 1µl dye and 0.8µg of DNA, corresponding to a ratio of 1.25µl/µg was used, as shown on Table 2.9 below.

Table 2.9: ULS Labelling Master Mix Components

Component	Per tube (µl)	*For 4 tubes (µl)
10X Labelling solution	1.0	4.5
ULS-Cy3 or ULS-Cy5	1.0*	4.5
Total	2.0	9.0

** - the volume of dye used depends on the amount of DNA used. A typical ratio was 1µl dye for 0.8µg DNA (1.25:1µl/µg). When smaller volumes of dye were required, the total volume per reaction was made up with nuclease- free water.*

Following a pulse centrifugation to drive the contents of the reaction tube to its bottom, appropriate dye master mix (2µl) was added to each tube to make the reaction volume up to 10µl, and mixed well by pipetting up and down. The tubes were then placed in a thermocycler with a heated lid programmed for incubation at 85ºC for 30 minutes for the labelling reaction and then held at 4ºC until ready for clean-up.

ULS Clean-up

Unreacted dye can lead to excess background fluorescence and this was removed using KREApure® filters (Agilent). Towards the end of the labelling reaction, one KREApure® filter per reaction was briefly vortexed, placed in a collection tube and centrifuged at 17, 000 × *g* for 1 minute at room temperature. The flow-through and lid were then discarded and 300µl of nuclease-free water applied to each filter. A further centrifugation step at 17, 000 × *g* for 1 minute was then carried out and the collection tubes and flow-through discarded. The filters were then transferred to clean labelled 1.5ml microfuge tubes.

As soon as labelling was completed, the reaction tubes were pulse centrifuged to drive contents to the bottom and 10µl of nuclease-free water added, making the volume up to 20µl in each. Each total reaction was then transferred to a KREApure® filter placed in the corresponding labelled microfuge tube and centrifuged at 17, 000 × *g* for 1 minute at room temperature. The filters were then discarded and clean ULS-labelled DNA samples collected in the microfuge tube were stored at 4ºC in the dark until required.

2.3.6.3 Measurement of DNA Labelling Efficiency

UV-VIS Spectrophotometry was used to determine the fluorescent activity of labelled DNA using the NanoDrop® ND-1000. The procedure used was similar to that described for DNA quantification (in Section 2.3.3) and the instrument simultaneously calculates the concentration of DNA, Cy3- and Cy5-dyes based on absorbance measurements at wavelengths 260, 550 and 650nm, respectively. The absorbance coefficient was set for double-stranded DNA at 50ng·cm/µl and appropriate blank solutions were used (1X TE and 1X labelling solution for Exo-Klenow and ULS labelled samples, respectively). From each labelled DNA sample, 1.5µl was taken for spectrophotometry.

Exo-klenow labelling reactions involve amplification of the DNA while ULS labelling reactions do not. The two methods therefore had different parameters for evaluating the efficiency of labelling. For Exo-Klenow labelled samples, the final DNA yield and the amount of dye per microgram of DNA (Specific Activity) were used. ULS labelling efficiency was assessed using the number of molecules of dye per molecule of DNA (Degree of labelling) expressed as a percentage. The calculation methods for these parameters are shown below with optimal values summarised in Table 2.10.

$$\boldsymbol{Dye\ Specific\ Activity}\ (\boldsymbol{pmol/\mu g}) = \frac{Dye\ concentration\ (pmol/\ \mu L)}{DNA\ concentration\ (\mu g/\mu L)}$$

$$\boldsymbol{DNA\ Yield} = DNA\ concentration\ (\mu g/\mu l) \times volume\ (\mu l)$$

$$\boldsymbol{Degree\ of\ Labelling}\ (\%) = \frac{340 \times Dye\ concentration\ (pmol/\mu l)}{DNA\ concentration\ (\mu g/\mu l)} \times 100\%$$

Table 2.10: Expected Values for DNA labelling efficiency parameters

	Exo-Klenow		ULS	
	Cy3	***Cy5***	***Cy3***	***Cy5***
Dye Specific Activity *(pmol/µg)*	25 - 55	20 - 40	-	-
Degree of Labelling (*%*)	-	-	1.75 - 3.50	0.75 – 2.50
DNA Yield (*µg*)	5 - 10		Same as input amount*	

Values are based on starting amounts of DNA between 0.5 and 1.0µg

** - ULS labelling reactions do not amplify input DNA*

If labelling was optimal, tumour and matched normal DNA samples (19.5µl each for enzymatic labelling and 18.5µl each for ULS labelling) were then combined. For ULS-labelled samples, the combined volume was reduced from 37µl to 22µl by vacuum centrifugation in the dark and then prepared for hybridisation. Enzymatic labelled samples (combined volume of 39µl) were stored in the dark at -20ºC until ready for hybridisation.

2.3.6.4 Pre-hybridisation Blocking

Repetitive DNA sequences were blocked with COT-1 DNA prior to hybridisation using the Agilent Oligo aCGH Hybridisation Kit, according to manufacturers' instructions.

Enzymatic-Labelled Samples

Combined matched samples were transferred to pre-labelled 0.2ml thin-walled PCR tubes. A hybridisation master mix was prepared as outlined on Table 2.11. 71µl of the hybridisation master mix was added to each sample tube making the total volume up to 110µl and mixed well by pipetting up and down. They were then pulse centrifuged and transferred to a thermocycler programmed for incubation at 95ºC for 3 minutes, followed by 37ºC for 30 minutes and then held at 37ºC until ready for hybridisation assembly.

Table 2.11: Hybridisation Master Mix Components for Enzymatic-Labelled Samples

	Per tube (µl)	***For 5 tubes (µl)**
Human COT-1 DNA	5	25
10X Blocking Agent	11	55
2X RPM Buffer	55	275
Total	71	355

ULS-labelled Samples

Pre-hybridisation blocking reaction was made up by adding 61µl of a hybridisation master mix (prepared as summarised in Table 2.12) to each sample tube, making the volume up to 83µl. The reaction was mixed well by pipetting up and down and pulse centrifuged. It was then incubated on a thermocycler programmed as for enzymatic-labelled samples. The tubes were then pulse centrifuged and 27µl of CGHblock® (pre-equilibrated to room temperature) was added, to bring the total reaction volume to 110µl and mixed well by pipetting up and down.

Table 2.12: Hybridisation Master Mix Components for ULS-Labelled Samples

	Per tube (µl)	*For 5 tubes (µl)
Human COT-1 DNA	5	25
100X Blocking Agent	11	55
2X RPM Buffer	55	275
Total	61	355

2.3.6.5 Hybridisation Assembly

One clean 4X microarray gasket slide was loaded into the SureHyb® chamber base with the gasket label facing up and aligned with the rectangular section of the chamber base. Noting the position of the gaskets relative to the barcode, 100µl of each sample (Exo-Klenow or ULS-labelled) was slowly dispensed into a gasket in a 'drag and dispense' manner.

A microarray slide was then carefully placed, active "Agilent- labelled" side down, onto the gasket slide and the SureHyb® chamber cover put in place over the sandwiched slides. The clamp was then slid gently onto both pieces and tightened to complete the assembly. The assembly was rotated vertically to wet the microarray slide and ensure that trapped air bubbles are freely mobile. The chamber assembly was then placed in the rotator rack of the microarray hybridisation oven set to 65ºC and set to rotate at 20 rpm for exactly 24 hours (for fresh frozen samples) or 40 hours (for FFPE samples).

2.3.6.6 Post-Hybridisation Washing and Scanning

Two wash conditions were set up before the hybridisation assembly was removed from the incubator. The first wash setup consisted of a glass dish with slide rack filled with aCGH Wash Buffer 1 and placed on a magnetic stirrer with a rotating stir bar in place. The second wash setup involved filling a glass dish with aCGH Wash Buffer 2 pre-warmed to 37^{o}C and maintained on a heated magnetic stirrer with a rotating stir bar in place.

The SureHyb® chamber assembly was taken out of the Hybridisation Oven at the end of 24 hours and assessed to ensure that all bubbles were still mobile. It was then laid on a horizontal surface and the clamp unscrewed and gently slid off. The chamber cover was then carefully lifted off the slide sandwich, which is removed intact from the chamber and transferred in the same plane (horizontal with microarray slide on top) to a third glass dish containing Wash Buffer 1.

Keeping the sandwich submerged in Wash Buffer 1, the slides were then gently pried apart with plastic forceps and the gasket slide allowed to drop to the bottom of the glass dish. The microarray slide was then immediately transferred to the slide rack in the first wash setup and washed for 5 minutes at room temperature. Quickly, so that the slide is not allowed to dry, the slide rack was then transferred to the second wash setup and the slide washed for exactly 1 minute at 37ºC. The slide rack was then removed slowly to minimize the formation of droplets on the slide, which is then placed in a slide holder for immediate scanning.

Slides were scanned using the Agilent Surescan® microarray scanner with control software v8.3 (Agilent) configured as recommended by manufacturers. The scanned images were saved in Tagged Image File Format (.TIFF), which were examined for hybridisation artefacts or microarray damage.

2.3.6.7 Microarray Data Processing

Array Quality Assessment and Feature Extraction

Scanned images were analysed using Feature Extraction software v10.7.3 (Agilent). The software normalises the fluorescent intensity of both dyes at each probe and calculates their ratio, expressed on a logarithmic scale (probe $\log_2$-ratio). Log_2-ratios for all 180,000 probes were then exported as Feature Extraction files in Tab Delimited Text (.txt) format.

The feature extraction software generates in addition, a Quality Control (QC) report that includes statistical metrics used to assess the reproducibility and reliability of results in each microarray experiment. Average background noise and signal-to-noise ratio for each dye were determined by calculating the mean green and red signal intensities at all genomic probes compared to non-hybridising control probes. A Derivative Log Ratio Spread (DLRS), an important parameter for determining reliability of array CGH was also calculated by measuring variation in the difference between $\log_2$ ratios of consecutive oligonucleotide probes. Threshold values used for acceptance of array CGH data as valid based on these metrics are summarised in Table 2.13 below.

Table 2.13: QC metric thresholds for Array CGH Experiments by Labelling Method and DNA source

	Exo-Klenow Labelling	**ULS Labelling**	
	Fresh Tissue/Cells	*Fresh Tissue/Cells*	*FFPE Tissue*
Signal Intensity (Red or Green)	>90	>90	>90
Signal to Noise Ratio (Red or green)	>30	>20	>20
Derivative Log Ratio Spread	<0.3	<0.3	<0.4

Threshold values are based on manufacturers' recommendations

Aberration Detection

For individual arrays, genomic copy number aberrations were identified using the FASST2 (Fast Adaptive States Segmentation Technique 2) algorithm in Nexus Copy Number Software v6.1 (Biodiscovery). The algorithm uses a Hidden Markov Model (HMM) - based approach that does not aim to estimate the copy number state at each probe but uses many states to cover more possibilities, such as mosaic events. These state values are then used to make calls based on a specified $\log_2$ ratio threshold.

Log_2 ratio threshold values of +0.25 and -0.3 were used to identify single copy number gains and losses respectively, and threshold values for gains and losses of two or more copies were set at +1.14 and -1.1 respectively. Significance threshold p-value was set at a minimum of 5.0×10^{-6}, requiring at least three contiguous probes for aberration calls. All threshold values were based on analysis software manufacturers' recommendations. Aberrations were presented on graphical genomic plots called ideograms, which could be viewed at whole genome, chromosome and single gene/exon levels to permit easy visual analysis (Figure 2.1).

2.3.7 Chromosome Spread Preparation

2.3.7.1 Chromosome harvesting

Cells were cultured in flasks until they were 60 – 80% confluent. Colcemid 0.05µg/ml (around 7 drops added to 5ml of media) was then added using a 1ml syringe and the cells incubated for a further 2 - 4 hours. The media was then transferred to a 15ml universal tube (harvesting tube) and pre-warmed trypsin-EDTA at 37ºC used to dissociate the cells. The cell suspension was then added to the same harvesting tube and centrifuged at 1500rpm for 5 minutes at room temperature.

The supernatant was removed, save 0.5ml in which the pellet was gently re-suspended. Hypotonic (0.075M) potassium chloride (approximately 5ml) was then added drop-wise while gently agitating the cell suspension and the harvesting tube transferred to the 37ºC/5%CO_2 incubator for 30-45 minutes to allow the cells to swell. They were then centrifuged at 1000rpm for 10 minutes and the supernatant removed leaving approximately 0.5ml in which the pellet was gently re-suspended. Freshly prepared methanol – acetic acid fixative solution (approximately 2ml) was then added in drops while gently agitating the cells. This was followed by a repeat of the centrifugation, re-suspension and fixative steps. The harvesting tubes were then stored upright at -20ºC until ready for slide preparation.

2.3.7.2 Slide Preparation

Metaphase chromosome spreads were prepared on clean glass slides using the method originally described by Hsu and Pomerat (Hsu and Pomerat 1953). Briefly, cells that had been incubated in hypotonic solution and then fixed were dropped onto slides from a variable height and where necessary, a few drops of fixative solution added to improve chromosome spreading. Slides were then left to dry in air and stored for up to one week at room temperature or at -20ºC for longer.

2.3.8 Giemsa Banding

Giemsa banding (G-banding) was carried out on metaphase chromosome spreads prepared as described in Section 2.3.7.2 and aged by air-drying for one week. Banding was done in humid condition, on a slide rack over a sink half-filled with warm water. Slides were first covered with banding trypsin and incubated for at least 90 seconds (depending on metaphase spread quality) after which they were rinsed off with Sorenson's buffer. They were then stained with Giemsa solution for at least 30 seconds, rinsed off with Gurr's buffer, blotted with filter paper and allowed to air-dry. Slides were then analysed under an Olympus® BH-2 light microscope attached to a software-controlled Cohu® high performance Charge-Coupled Device camera (Applied Imaging®).

Duration of trypsin treatment and Giemsa staining varied depending on metaphase spread quality and experiments typically involved multiple attempts using increasing durations, until staining was optimised and clear banding patterns could be seen. Images were captured and the total number of chromosomes per metaphase was recorded from a minimum of 30 metaphase spreads. Chromosomal G-banding pattern analysis was carried out by experienced Cytogeneticist, Dr Karen Sisley.

2.3.9 Fluorescence *in situ* Hybridisation

Fluorescence *in situ* hybridisation (FISH) was used to visually confirm copy number aberrations and obtain ploidy information, using interphase nuclei and/or metaphase chromosome spreads harvested from cultured STS cells (as described in Section 2.3.7) or on FFPE tissue sections. Where available, commercial probes were used for FISH. Otherwise, target-specific DNA fragments cloned into Bacterial Artificial Chromosome (BAC clones) vectors were used to prepare customised probes.

2.3.9.1 Preparation of Bacterial Artificial Chromosome DNA

BAC clones were supplied in transformed *E.coli*, which was first cultured to amplify the cloned human DNA sequences followed by bacterial lysis and purification. *E.coli* glycerol stocks were streaked on selective LB agar plates containing chloramphenicol (12.5µg/ml) and then incubated overnight at 37°C. Single colonies were then transferred into 15ml polypropylene tubes containing 2ml selective LB broth (containing 12.5µg/ml with chloramphenicol) and incubated overnight at 37ºC on a shaker at 300rpm.

The now-turbid culture broth was then transferred to a fixed-angle centrifuge and spun at 3000rpm for 10 minutes at room temperature. The supernatants were carefully removed and each pellet re-suspended in 0.3ml of Solution P1. An equivalent volume of Solution P2 was then added to each tube, gently shaken to mix its contents and allowed to sit upright for 5 minutes at room temperature, during which the sample changed from turbid to translucent. Solution P3 (0.3ml) was then added slowly, while gently agitating the tube allowing a white precipitate of proteins and bacterial DNA to form. The tubes were again placed on ice for 5 minutes and subsequently centrifuged at 10,000rpm for 10 minutes at 4ºC on the fixed angle centrifuge.

Careful to avoid the white precipitate at the bottom, the supernatant that contained only human DNA sequences was then transferred to a 2ml microfuge tube containing 800µl of ice-cold isopropanol. The mixture was inverted a few times to mix and then incubated on ice for 5 minutes, followed by

centrifugation at 16,000 × *g* for 15 minutes at 4ºC. The supernatant was removed and 500µl of 70% ethanol was added, the tube quickly inverted a few times to wash the pellet and then centrifuged at 16,000 × *g* for 5 minutes at 4ºC. The ethanol was then removed and the tube left open to allow the pellet to air-dry completely before it was re-suspended in 1X TE buffer (40µl). DNA yield and sample purity was checked using the Nanodrop as described in Section 2.3.3.

2.3.9.2 BAC DNA Labelling (Nick Translation)

Nick translation as described by Rigby et al, was used to label BAC DNA (Rigby, Dieckmann et al. 1977). Briefly, DNAse 1 was used to form single-strand breaks (nicks) in the BAC DNA, which were used as starting points by *E. coli* DNA polymerase (holoenzyme) for degradation by 5' – 3' exonuclease activity followed by replacement of the nucleotides by its DNA polymerase activity, simultaneously incorporating fluorophore-labelled nucleotides (that were supplied in high concentration) every 20-25 nucleotides.

A nick translation reaction was prepared in a thin-walled 0.2ml PCR tube on ice. Reaction components were added in the order outlined on Table 2.14 and mixed well by pipetting up and down. The reaction tube was then transferred to a thermocycler programmed for incubation at 15°C for 8 hours and then at 70°C for 10 minutes to stop to the reaction, after which the tubes were held at 4°C.

Table 2.14: Components of Nick Translation Reaction

Component	Volume (µl)
Nuclease-free water	$17.5 - x$
1µg BAC DNA	x
0.2mM Labelled dUTP	2.5
0.1mM dTTP	5
0.1mM dNTP mix	10
10X Nick Translation Buffer	5
Nick Translation Enzyme	10
Total	**50**

Agarose gel electrophoresis (as described in Section 2.3.4) was used to visualise the labelled DNA and carried out using 5µl of the sample. The remainder was stored in 15µl aliquots at -20ºC in the dark.

2.3.9.3 Pre-treatment of FISH Slides

FFPE tissue Sections

Tissue sections (5µm thick) on glass slides were incubated on a heat block at 58ºC for 3 minutes and then de-paraffinised by two consecutive incubations in xylene for 5 minutes each under a fume hood. This was followed by dehydration in absolute ethanol for 5 minutes at room temperature, after which the slides were transferred to a coplin jar containing 0.2N HCl (50ml) and incubated for between 20 and 23 minutes. They were subsequently washed for 2 minutes in distilled de-ionised water (ddH_2O) on a shaker with gentle agitation.

The slides were then transferred to a coplin jar containing Zymed pre-treatment solution (40-50ml) held at 95ºC in a water-bath and incubated for 2 – 3 hours, after which they were washed twice in ddH_2O for 3 minutes each at room temperature with gentle agitation. Enzyme reagent was pre-warmed to room temperature and 100µl applied to 22 × 50mm coverslips, which were used to cover each tissue section. The sections were then placed in a humid chamber and incubated at room temperature for 2 – 3 hours, refreshing the enzyme after every 60 minutes. The coverslips were then removed and the slides washed thrice in ddH_2O for 2 minutes each at room temperature with gentle agitation. They were then incubated through a series of 70%, 95% and 100% ethanol for 3 minutes each to dehydrate, and allowed to dry completed in room air.

Fixed Cells and Metaphase Chromosome Spreads

Slides were treated with RNAse A (125µl) loaded onto 22 × 50mm coverslips, and incubated in a humid chamber at 37°C for 1 hour. They were then washed thrice in 2X SSC for 5 minutes each with gentle agitation. Next, the slides were immersed in a coplin jar containing pre-warmed Pepsin-HCl solution held at 37°C on a water bath and incubated for exactly 10 minutes. This was followed by two washes in PBS for 5 minutes each and a further wash in PBS-$MgCl_2$ for 5 minutes at room temperature with gentle agitation. The slides were then incubated in 37% formaldehyde solution to fix for 10 minutes at room temperature and washed again in PBS for 5 minutes with gentle agitation. They were then dehydrated by immersion in a series of 70%, 95% and 100% ethanol for 3 minutes each and air-dried.

2.3.9.4 Preparation and Application of FISH Probes

During the pre-treatment of target slides, labelled BAC probe was precipitated using the sodium acetate – ethanol precipitation method described in Section 2.3.4. COT1 DNA (1µg) and normal human genomic DNA (2µg) were added to each 100ng of BAC probe DNA as carriers to facilitate precipitation. The precipitated DNA pellet was re-suspended in LSI/WCP hybridisation buffer (7µl per slide). Where commercial probes were to be simultaneously applied to the target sample, an appropriate volume (based on probe efficiency shown on Table 2.2) of commercial probe was added to the re-suspended BAC probe. This probe master mix volume was then made up to 10µl per slide using nuclease-free water. Similarly, for samples where a combination of commercial probes were to be applied, the appropriate volumes of each probe were added to 7µl hybridisation buffer per slide, and the mixture made up to 10µl per slide with nuclease-free water. The LSI/WCP hybridisation buffer was used, except in cases where both probes were centromeric (CEP probes), in which the higher stringency CEP hybridisation buffer was used to prevent non-specific hybridisation.

Probe DNA was denatured by placing the master mix on a heat block at 80°C for 10 minutes. It was then pulse centrifuged and 10µl was loaded onto individual 22 × 22mm coverslips over which each pre-treated target slide was firmly placed and sealed with vulcanising rubber solution. The slides were then placed on the heat block at 80ºC for 2 minutes to denature target DNA, after which they were transferred to a dark, humid chamber and incubated at 37ºC for 16 -18 hours for hybridisation.

2.3.9.5 Post-Hybridisation Washes and Counterstaining

Post-hybridisation washing and subsequent steps were carried out under dim conditions. Coverslips were removed using tweezers and the slides transferred to a coplin jar filled with SSCT-1 at 73ºC on a water bath where they were incubated for exactly 2 minutes. They were then transferred to SSCT-2 for a 1-minute wash at room temperature with gentle agitation. This was followed by dehydration through a series of 70%, 95% and 100% ethanol for 2 minutes each and drying in air.

DAPI counterstain solution was then placed in two 10µl drops on a 22 × 50mm coverslip over which each slide was firmly placed facedown and sealed with nail varnish. Slides were then stored at 4ºC for at least 2 hours to allow counterstaining before analysis on a fluorescent microscope. FISH probe signals were scored from 100-200 non-overlapping intact nuclei or metaphase spreads in each case and images were captured using Cytovision® software.

2.3.10 Immunochemistry

Cultured cells fixed on glass slides were used for immunocytochemistry while FFPE sections (4µm-thickness) were used for immunohistochemistry by the modified Avidin-Biotin-Peroxidase Complex (ABC) method described by Hsu et al (Hsu, Raine et al. 1981). Briefly, samples were pre-treated as required to expose relevant epitopes and block non-specific antibody binding. They were then incubated with a specific primary antibody, washed and then a biotin-conjugated secondary antibody, which binds to the F_C portion of any bound primary antibody, is added. Samples are washed again and bound secondary antibody forms complexes with added avidin molecules linked to an enzyme reporter system that leads to a colorimetric reaction that is used to detect the presence of the target antigen in the sample. Negative controls slides in which samples were incubated with either antibody diluent or immunoglobulin fraction in place of primary antibody were included in every experiment and where available, positive control samples were also included.

2.3.10.1 Slide Preparation for Immunocytochemistry

Glass slides were cleaned by immersion in methanol for 10 minutes at room temperature. They were then left to dry under a laminar flow hood and arranged in a single layer in a sterile 20mm petri dish. Primary sarcoma cells cultured in flasks until they were 60 – 80% confluent were trypsinised as described in Section 2.3.1 and counted. Approximately 20,000 cells, re-suspended in 0.5ml of culture media were then transferred to each slide. The petri dish was placed in the 37°C/5%CO_2 incubator for 4 hours to allow the cells to attach to slides and then a further 10ml of culture media was added to the petri dish and incubated overnight.

The media was then removed and the slides washed twice with sterile PBS. The slides were then placed on a rack and incubated in an ice-cold fixative solution made up of methanol and acetone in a 1:1 mixture for 10 minutes. The slides were then dried under the laminar flow hood and used for immunocytochemistry or stored at -20ºC until needed.

2.3.10.2 Tissue Preparation and Epitope Retrieval for Immunohistochemistry

Tissue sections were de-paraffinised by two consecutive 10-minute incubations in xylene and then dehydrated by 5-minute serial incubations in 70%, 95% and 100% Ethanol. They were then incubated in 1% H_2O_2 for 30 minutes at 37ºC to quench peroxidase activity and subsequently washed in PBS for 5 minutes. Heat-induced epitope retrieval (HIER) was carried out by immersing the tissue sections in an appropriate buffer (Table 2.4) then heating in a microwave set at 900W for 8 minutes. They were allowed to cool in the same buffer for at least 15 minutes and then washed in PBS for 5 minutes.

2.3.10.3 Blocking and Primary Antibody Incubation

Relevant areas of tissue sections and fixed cells were outlined on the slides using a wax pen and washed for 5 minutes in PBS. Appropriate 10%blocking serum (Table 2.4) was then applied and incubated for one hour at room temperature. The blocking serum was then tipped off and primary antibody (diluted in 2% blocking serum) was applied to test and positive control slides. To negative control slides, either 2% blocking serum (without antibody) or relevant immunoglobulin fraction (also diluted in 2% blocking serum) was applied. Slides were then incubated for 1-2 hours at room temperature or overnight at 4ºC.

2.3.10.4 Secondary Antibody Incubation and Immunoreactivity

Slides were washed twice in PBS for 5 minutes each and the appropriate biotinylated secondary antibody (also diluted in 2% blocking serum) was applied to all slides and incubated for 30 minutes at room temperature. ABC reagent was prepared immediately and allowed to stand during secondary antibody incubation. The slides were then washed twice in PBS for 5 minutes each and ABC reagent applied and incubated for 30 minutes at room temperature. This was followed by two 5-minute washes in PBS after which freshly-prepared peroxidase substrate solution (DAB) was applied and incubated at room temperature until the desired brown stain intensity developed (up to 10 minutes).

2.3.10.5 Counterstaining and Mounting

The slides were rinsed in tap water and counterstained for 60 seconds in Gill's haematoxylin then washed in running tap water. Slides were transferred to Scott's water substitute for 10 seconds and washed further in running tap water until the water was clear.

The sections were then dehydrated by serial 3-minute incubations in 70%, 95% and 100% ethanol and de-waxed by two consecutive incubations in xylene for 3 minutes each under a fume hood. While still wet with xylene, the sections were then covered with 22 × 40mm coverslips using DePex and allowed to dry. Slides were then examined using a light microscope at the appropriate magnification.

2.3.11 Cell Proliferation Assay

The MTT ((4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay was used to assess proliferation in cultured cells treated with the Trio RhoGEF inhibitor, ITX3. It is based on the principle that only viable cells are able to metabolize tetrazolium salts to form blue formazan crystals. Briefly, equal numbers of cells were cultured in 96-well plates in the presence of ITX3 or DMSO vehicle.

MTT was then added to the cultures and any formazan crystals formed were dissolved. Compared with appropriate controls, the optical density of the resulting solution was used as a surrogate for the number of viable cells remaining in culture.

Cells were seeded at a density of 2,000 – 10,000 cells in 100µl of media in each well of a 96-well plate and incubated in a 37^{o}C/5%CO_2 incubator for 4 hours to allow cells to settle. The media in each well was then removed and replaced with culture media containing 1% DMSO (v/v) with or without incremental concentrations of ITX3 (50µM - 1mM) in triplicate. The cells were then returned to the incubator and cultured for 7-10 days. Exactly 100µl MTT (1mg/ml dissolved in PBS) was then added to each well and the culture plates were then returned to the incubator for up to 4 hours, monitoring the control wells (without ITX3) for the formation of blue formazan crystals. The media in the wells was then removed completed by pipette aspiration and replaced with exactly 100µl DMSO. They were then incubated for at least 30 minutes in the dark with gentle agitation. Optical density absorbance in each well was then measured at 570nm on a spectrophotometer. The ratio of absorbance in the test wells to control wells was calculated as relative MTT activity. Average MTT activity from five replicate wells was plotted against ITX concentration and statistical significance ($p < 0.05$) judged by two-way ANOVA with Bonferroni post hoc correction, using GraphPad Prism® software v6.0.

2.3.12 Micro-chemotaxis/Invasion Assay

A 48-well micro-chemotaxis chamber was used to assess the effect of Trio RhoGEF inhibitor treatment on invasive and chemotactic activity of primary sarcoma cells. A modification of the method described by Albini et al (Albini, Iwamoto et al. 1987) was used in which a porous polycarbonate membrane was coated with an artificial extracellular matrix composed of collagen, laminin and fibronectiin in equal concentration (Table 2.5) and foetal bovine serum used as chemo-attractant.

Briefly, cells were cultured in full culture media (RPMI 1640 supplemented with 20% serum) in 6-well plates until they were 60 – 80% confluent. The media was then replaced with serum-free media (RPMI 1640 containing 0.1% Bovine Serum Albumin) and returned to the incubator. After 4 hours, DMSO (1% v/v) with or without incremental concentrations of ITX3 was added to the media in test and control wells, respectively. An additional control well did not have any DMSO or inhibitor added. The cells were then cultured for a further 2 hours before they were trypsinised, counted and re-suspended in serum-free media containing the same DMSO/inhibitor concentration.

During inhibitor pre-treatment, polycarbonate membranes were coated on their underside with artificial extracellular matrix (ECM) solution for 45 minutes then allowed to dry in a laminar flow hood. A minimum of 3 replicates of control and test wells were set up in each experiment. Lower wells of the Boyden chamber were filled with 27-28µl full culture media and the coated membrane carefully placed over them. The silicone gasket and upper chamber were then carefully placed over the membrane and assembly screws applied and tightened. A 50µl volume of the cell suspension containing approximately 1×10^4 cells was then added to noted wells of the upper chamber and the assembly placed in the 37°C/5%CO_2 incubator for 16 – 24 hours.

The chamber was then disassembled and any cells on the top surface of the membrane gently removed by scraping over a plastic wiper. The underside of the membrane was then washed for 5 minutes in PBS and the cells fixed by washing in 100% methanol for 5 minutes. Any cells remaining on the underside of the membrane were then stained with 4% Gill's Haematoxylin for 5 minutes and excess stain removed by washing in dH_2O for another 5 minutes. The membrane was then mounted on a glass slide with DePex.

Cell present on the underside of the membrane were photographed using a light microscope at × 200 magnification. For each well, cells in 5 random fields were counted and averages taken over three replicate wells. Using GraphPad Prism® software v6.0, average cell counts were then plotted against inhibitor concentration and compared with controls, using the one-way ANOVA test with Bonferroni's post hoc correction for statistical significance (when $p < 0.05$).

CHAPTER THREE

RESULTS: OPTIMISATION OF ARRAY CGH

3.1 INTRODUCTION

Array CGH involves co-hybridising fragments of test and reference genomic DNA that have been differentially labelled with fluorescent dyes to a set of mapped and annotated DNA sequences (probes) on a microarray. By measuring the ratio of fluorescence at each probe, it is possible to detect copy number differences between the test (tumour) and reference (normal) DNA at that genomic location such that genome-wide copy number aberrations can be accurately mapped. Target probes may be cDNA sequences, Bacterial Artificial Chromosomes (BACs) or oligonucleotides and depending on the size, type and number of probes on the array, somatic copy number aberrations (SCNAs) can be detected at the level of single genes and even specific exons (Barrett, Scheffer et al. 2004). The highest resolution aCGH methods available are the oligonucleotide (60mers) arrays and with up to a million probes on an array, SCNAs may be detected with a resolution as high as 1-2 Kilobase pairs (kb). Commercially available oligonucleotide CGH arrays have an added advantage of being customisable by user-friendly *in silico* design that can focus on specific areas of the genome at no additional cost (Tan, Lambros et al. 2007).

In order to generate reproducible aCGH results, pure high molecular weight DNA (usually obtained from fresh frozen tissue, blood or cultured cells) has traditionally been used. Availability of fresh tissue is however limited, particularly in rare tumours like STS and most tumour tissue available for research is formalin-fixed and paraffin-embedded (FFPE) in order to preserve the tissue structure for histology. DNA isolated from such tissues is typically of low quality (low yield and highly fragmented) due to the degradative effects of formalin (Srinivasan, Sedmak et al. 2002) and studies comparing the aCGH performance of high and low quality DNA showed that fragment sizes < 200 base pairs (typical of FFPE DNA) produced noisy and irreproducible results (van Beers, Joosse et al. 2006, Mc Sherry, Mc Goldrick et al. 2007). Another major limitation to the use of FFPE DNA for high-resolution oligonucleotide aCGH is technical difficulty in labelling fragmented DNA. Traditional enzymatic methods for labelling DNA (Nick translation or Random priming) involve a fragmentation step with DNase or restriction digestion respectively, which in the case of FFPE DNA further fragments the DNA. An alternative to enzymatic labelling is the Universal Linkage System (ULS), which directly labels the DNA by a chemical reaction that incorporates platinum-conjugated fluorophores into DNA without the need for fragmentation, making it suitable for low quality fragmented DNA such as that from FFPE tissue (Alers, Rochat et al. 1999).

Recent efforts have been made to optimise the utility of archived FFPE tissues for array CGH after labelling by enzymatic (Hostetter, Kim et al. 2010) and ULS methods (Szponar, Yusenko et al. 2010, Braggio, McPhail et al. 2011, Chen, Liu et al. 2011, Oikawa, Yoshiura et al. 2011, Rossi, Klersy et al. 2011). Nevertheless, the use of FFPE DNA for aCGH is still regarded as technically challenging and limited to very small-scale studies, which have reported variable data quality. Previous work using the lower resolution chromosome CGH identified common aberrations that were subsequently confirmed by FISH (Ul-Hassan, Sisley et al. 2009). Optimisation of the ULS labelling protocol and subsequent high-resolution oligonucleotide array CGH analysis (Agilent® 180K platform) was thus done using archival FFPE leiomyosarcoma tissue and the results are presented in this chapter.

3.2 RESULTS

3.2.1 Tumour Samples

Twenty-two FFPE tumours diagnosed as LMS obtained from the Sheffield archive were analysed by aCGH. Their ages (archival storage time) at the time of analysis ranged from 1 - 17 years, eight cases being over 10 years old (Table 3.1). Two cases, LMS 20 and 21 had been initially processed in a different Pathology laboratory. Fresh frozen (FF) samples from three of the cases, LMS 8, 9 and 10 were also analysed. The quality of the DNA obtained from the FF samples and subsequent array CGH results were compared with those of the corresponding FFPE tumours. FISH was used to verify some of the consistent copy number changes.

Table 3.1: Characteristics of Archival FFPE LMS cases included in this study.

Case	Anatomical Site	Tumour Sampling Date	Age of Sample at Analysis (Years)
LMS 1	Bowel	2010	1
LMS 2	Lower Limb	2011	1
LMS 3	Lower Limb	2011	1
LMS 4	Bladder	2011	1
LMS 5	Stomach	2011	1
LMS 6	Stomach	1994	17
LMS 7	Lower Limb	1999	12
LMS 8	Bladder	1998	13
LMS 9[§]	Vagina	2011	1
LMS 10[§]	Retroperitoneum	2011	1
LMS 11[§]	Pelvis	2011	1
LMS 12	Stomach	2011	1
LMS 13	Uterus	2004	8
LMS 14	Bowel	1995	17
LMS 15	Uterus	1997	15
LMS 16	Uterus	1997	15
LMS 17	Nose	1998	14
LMS 18	Pelvis	2004	8
LMS 19	Retroperitoneum	2003	9
LMS 20*	Uterus	2008	4
LMS 21*	Uterus	2011	1
LMS 22	Lower Limb	2000	12

[§] - Additional fresh samples obtained and frozen before fixing in formalin
** - Initially processed in a different institution*

3.2.2 DNA Yield and Quality

The DNA yield was generally good, exceeding 10µg in most cases. In two cases however, the yield was low and vacuum centrifugation was used to bring DNA concentration to optimal values. Another FFPE DNA sample had $A_{260/230}$ < 1.50 and the DNA was re-purified by sodium acetate-isopropanol precipitation before labelling.

When visualized on agarose gels against a 1kb ladder, DNA from the FF samples and commercial pooled genomic DNA showed relatively distinct bands of high molecular weight DNA, while that from FFPE tumours showed a range of fragment sizes that varied from less than 1.0kb on average, to as high as 8.0kb. The degree of fragmentation appeared to be worse in FFPE samples that were older when compared with the more recent tumours (Figure 3.1). The tumour samples included in this part of the study were chosen to reflect a wide range of both sample age and degree to DNA degradation as visualised on agarose gels.

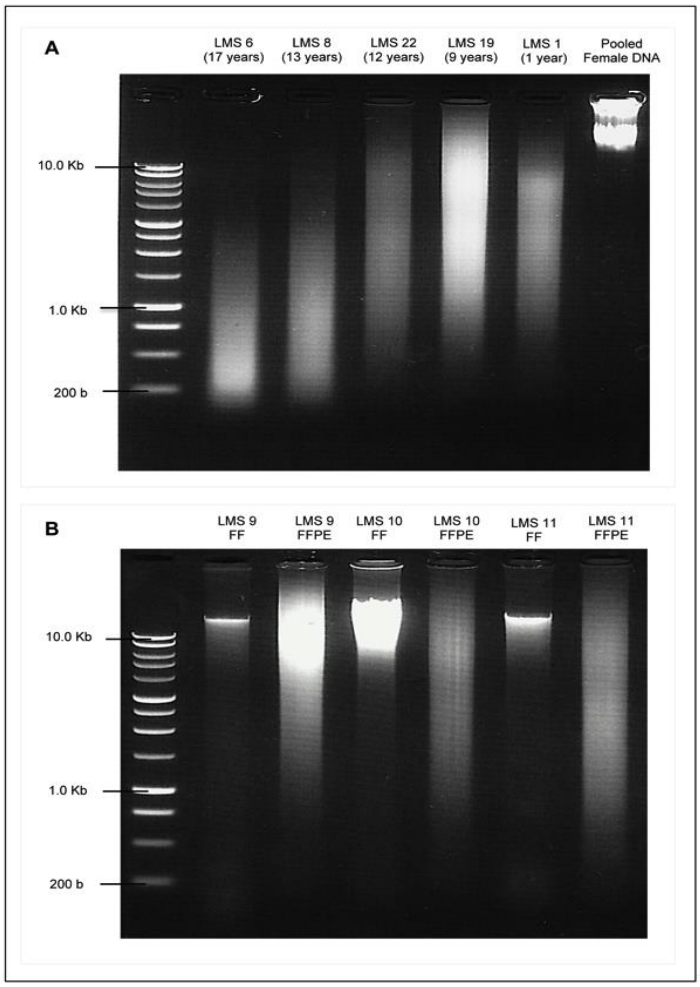

Figure 3.1: Agarose gel images illustrating the degree of fragmentation of DNA samples obtained from FFPE tissue compared with those from fresh tissue.

Panel A: DNA extracted from FFPE leiomyosarcoma samples of different ages (shown in brackets) showing degradation to varied degrees , compared with commercial pooled female genomic DNA with a clear band of large DNA fragments.
Panel B: Comparison of DNA extracted from paired FF and FFPE leiomyosarcoma tissue (LMS 9, 10 and 11). FF samples show relatively distinct bands of high molecular weight DNA, while corresponding FFPE samples show low molecular weight fragments in a wide range of sizes.

All DNA samples are compared against a 1Kb DNA ladder. DNA Electrophoresis was done on 1.0 % agarose gels were pre-stained with Ethidium Bromide and examined under UV light.

3.2.3 Optimisation of DNA Labelling

Initial attempts at DNA labelling by the ULS method produced a variable degree of labelling. The results showed that when labelling reactions were carried out on heat blocks or water baths, the degree of labelling was consistently low. In addition, for cases where post-labelling DNA quantitation showed that there was a higher amount of DNA in the reaction than initially estimated, the degree of labelling was either variable or arrays failed due to signal intensities that were below recommended thresholds.

Since the ULS system does not amplify the DNA, this suggested that the ratio of DNA to ULS Cy-dye in the reaction was high, resulting in an inefficient labelling reaction. The protocol was therefore modified to use excess ULS fluorophore relative to the amount of DNA (about 0.8 µg of DNA to 1ml of fluorophore). Reactions were also carried out in thin-walled tubes on a thermal cycler to ensure uniform optimal temperature throughout the reactions. All FFPE samples that were labelled in this way had consistently good degree of labelling. Results of experiments carried out before and after optimisation of the labelling protocol are summarised in Table 3.2.

Using this optimised labelling method, aCGH experiments were carried out using DNA from a total of 42 FFPE tumours (including the 22 LMS presented in this chapter). Quality control metrics of these array experiments were compared to those from experiments done with high quality DNA labelled using the enzymatic method (Table 3.3). In concordance with their recommended degrees of labelling (described in Section 2.3.6.3), the average signal intensities of Cy3-labelled DNA (green) were higher than those of Cy5-labelled DNA (red). The signal intensities for both fluorophores however were generally higher than the manufacturer-recommended threshold regardless of tissue sample type and labelling method. Similarly, signal-to-noise ratio was optimal in all FFPE samples labelled by the optimised protocol. Although samples ULS-labelled FFPE samples showed lower signal to noise-ratio values compared with enzymatically labelled ones, this was expected based on their recommended optimal thresholds of > 10 and > 20 respectively (Table 3.3).

The Derivative Log Ratio Spread (DLRS) is widely regarded as a robust parameter for measurement of the quality of microarray experiments. It represents the 'noisiness' of array data and a low DLRS value means that the data has small probe-to-probe variability and better ability to array to identify small aberrations and vice versa. Manufacturer-recommended thresholds for DLRS when using FFPE DNA is < 0.4, a value above which array data may be compromised. With the exception of two cases, DLRS values for all the array experiments were < 0.4 (Table 3.3) with majority of the FFPE samples having DLRS < 0.3 (Figure 3.2). Mean DLRS value was approximately of 0.24 for both FF and FFPE tissue samples types.

Table 3.2: Optimization of DNA Labelling Protocol

Sample	Tissue Type	Amount of DNA Used (µg)	ULS-Dye Used	DNA to dye ratio (µg/µl)	Post Labelling DNA Yield (µg)	Degree of Labelling (%)	Labelling Pass/Fail	Dye Signal Intensity on Array
A - Labelling reactions carried out using hot blocks and water baths								
1	FFPE	0.5	Cy5	1.0	0.49	0.4	**Fail**	*NP
2	FFPE	0.5	Cy5	1.0	0.5	0.2	**Fail**	*NP
3	FFPE	0.5	Cy5	1.0	0.5	0.23	**Fail**	*NP
4	FFPE	0.5	Cy5	1.0	0.66	0.36	**Fail**	*NP
5	Control DNA	0.5	Cy3	1.0	0.45	0.45	**Fail**	*NP
6	Control DNA	0.5	Cy3	1.0	0.45	0.45	**Fail**	*NP
7	Control DNA	0.5	Cy3	1.0	0.39	0.39	**Fail**	*NP
8	Control DNA	0.5	Cy3	1.0	0.56	0.56	**Fail**	*NP
B - Labelling reactions carried out using thermo-cycler								
9	FF	0.5	Cy5	**1.0**	0.89	2.01	**Pass**	99
10	FFPE	0.5	Cy5	**1.0**	0.72	2.39	**Pass**	94
11	FFPE	0.5	Cy5	**1.0**	0.59	3.27	**Pass**	79
12	Control DNA	0.5	Cy3	**1.0**	0.73	1.08	**Fail**	*NP
13	Control DNA	0.5	Cy3	**1.0**	0.68	1.15	**Fail**	*NP
14	Control DNA	0.5	Cy3	**1.0**	0.88	0.97	**Fail**	*NP
15	Control DNA	0.5	Cy3	**1.0**	0.83	2.15	**Pass**	153
16	Control DNA	0.5	Cy3	**1.0**	0.78	2.19	**Pass**	148
17	Control DNA	0.5	Cy3	**1.0**	0.82	2.19	**Pass**	103
C - Labelling reactions carried out using thermo-cycler and excess dye								
18	FFPE	0.8	Cy5	**0.8**	0.8	2.24	**Pass**	307
19	FFPE	0.8	Cy5	**0.8**	0.6	2.32	**Pass**	245
20	FFPE	0.8	Cy5	**0.8**	0.64	1.56	**Pass**	275
21	FFPE	0.8	Cy3	**0.8**	0.91	2.04	**Pass**	383
22	Control DNA	0.8	Cy3	**0.8**	0.95	3.08	**Pass**	1503
23	Control DNA	0.8	Cy3	**0.8**	0.81	3.07	**Pass**	1251
24	Control DNA	0.8	Cy3	**0.8**	0.76	3.1	**Pass**	1228
25	Control DNA	0.8	Cy3	**0.8**	0.87	3.11	**Pass**	1403

NP – Array CGH **not performed on sample (if sample failed on spectrophotometry). Spectrophotometry for DNA concentration and Degree of Labelling (DoL) done using Nanodrop® ND-2000 and optimal DoL = 0.75 – 2.5% (for ULS- Cy5); or 1.75 – 3.5% (for ULS Cy3). Dye Signal Intensity calculated as part of Quality Control Metrics by Agilent Feature Extraction Software (v10.7.3).*

Table 3.3: Comparison of array CGH quality control metrics for ULS-labelled and Enzymatic-labelled samples

	Enzymatic labelling (n=32)			ULS labelling (n=42)		
	Threshold	**Mean**	**Range**	**Threshold**	**Mean**	**Range**
Cy3 Signal Intensity (Green)	*>90*	**758**	(170 – 4204)	*>90*	**1334**	(749 – 2145)
Signal-to-noise ratio (Green)	*>20*	**63.4**	(14.2 – 118.7)	*>10*	**43.7**	(25.1 – 79.7)
Cy5 signal Intensity (Red)	*>90*	**603**	(73 – 2167)	*>90*	**517**	(117 – 1266)
Signal-to-noise ratio (Red)	*>20*	**54.8**	(24.9 - 90.3)	*>10*	**28.2**	(10.5 – 51.0)
Derivative Log Ratio Spread (DLRS)	*<0.3*	**0.24**	(0.12 – 0.51)	*<0.4*	**0.24**	(0.17 – 0.40)

Array CGH quality control metrics were calculated using Agilent® Feature Extraction software and thresholds are recommended by microarray platform manufacturer.

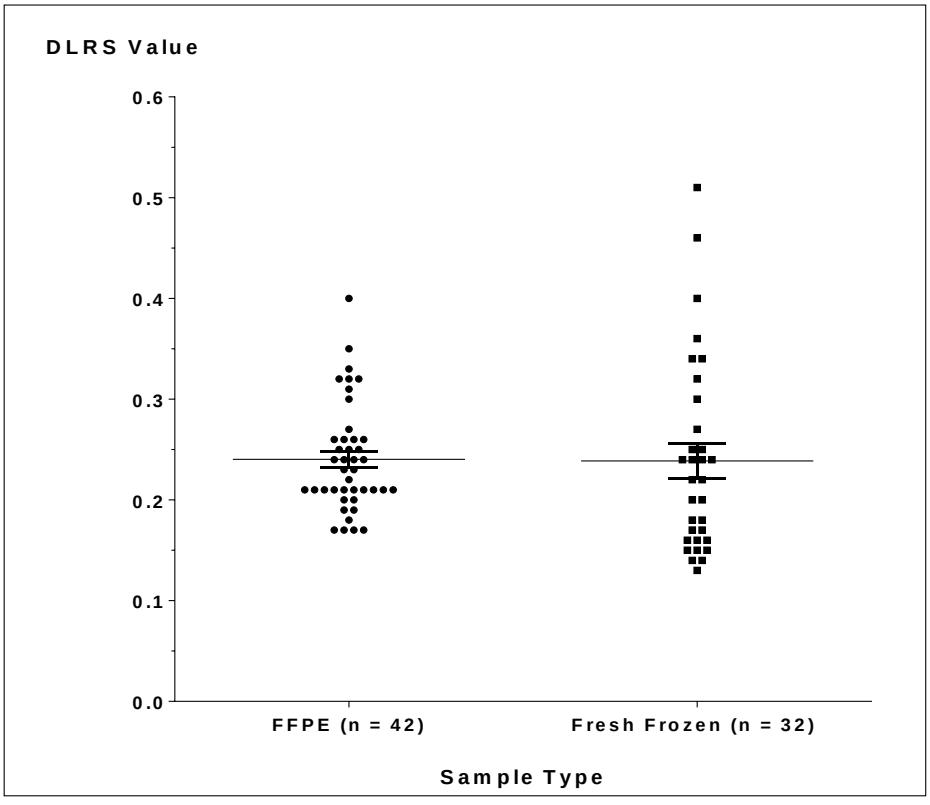

Figure 3.2: Derivative Log Ratio Spread (DLRS) Values of Array CGH Experiments

Tumour sample types are shown on the X-axis. DLRS values for individual microarray experiments are plotted as dots while means and standard errors for each sample type are shown as long and short horizontal lines, respectively plotted against the y-axis.

DLRS values were calculated using Agilent feature Extraction software.

3.2.4 Comparison of Paired Fresh Frozen *vs.* FFPE Tumour DNA

In three LMS cases (LMS 9, 10 and 11), paired samples of macroscopically sampled fresh tumour and macro-dissected FFPE tumour were obtained. High molecular weight DNA from the fresh frozen (FF) tissue was labelled by the enzymatic method, while fragmented FFPE DNA from the same tumours was labelled using the one-step ULS method. There was good correlation between paired FF and FFPE samples in two out of the three cases. The Pearson correlation coefficients (r) of overall probe $\log_2$-ratios were 0.58 ($p < 0.0001$) and 0.54 ($p < 0.0001$) for LMS 9 and LMS 10 respectively (Table 3.4).

Table 3.4: Correlation of Probe log2 ratios of paired FF and FFPE LMS samples

Correlation	LMS 9	LMS 10	LMS 11
Number of Probes	180,880	180,880	180,880
Pearson Coefficient, *r*	0.59	0.54	−0.02
95% Confidence Interval	0.5830 to 0.5891	0.5365 to 0.5431	−0.03 to −0.021

Pearson's Correlation, r of $\log_2$ ratio values of all probes on tumour DNA samples was calculated using GraphPad® Prism software and statistically significant ($p < 0.0001$).

Genomic profiles of detected SCNAs in both sample types were also very similar in both LMS cases with most chromosomes showing near identical loss and gain patterns (Figures 3.3 and 3.4). One of the most significant differences in SCNAs was seen in LMS 9, where a low level amplification detected on chromosome 4 in the macro-dissected FFPE sample was not seen in FF sample (Figure 3.3C). Similar moving average patterns of higher amplitude were however retained at the telomeric ends of 4q in both sample types. In both LMS 9 and 10, on most chromosomes where the aberrations detected by the calling algorithm were dissimilar, closer examination showed that the moving average pattern of probe $\log_2$ ratios remained similar with amplitude close to the threshold set in the algorithm for SCNA detection. An example is shown in Figure 3.4C.

The third case (LMS 11) compared in this way however, showed poor overall probe $\log_2$ ratio correlation ($r = -0.02$). Although a few small aberrations were seen in both sample types, most of the SCNAs detected in the macro-dissected FFPE tissue were not detected in FF tissue (Figure 3.5). Histological examination showed that unlike the other two cases, the LMS 11 sample was composed of less than 50% tumour cells with a significant ad-mixture of normal cells.

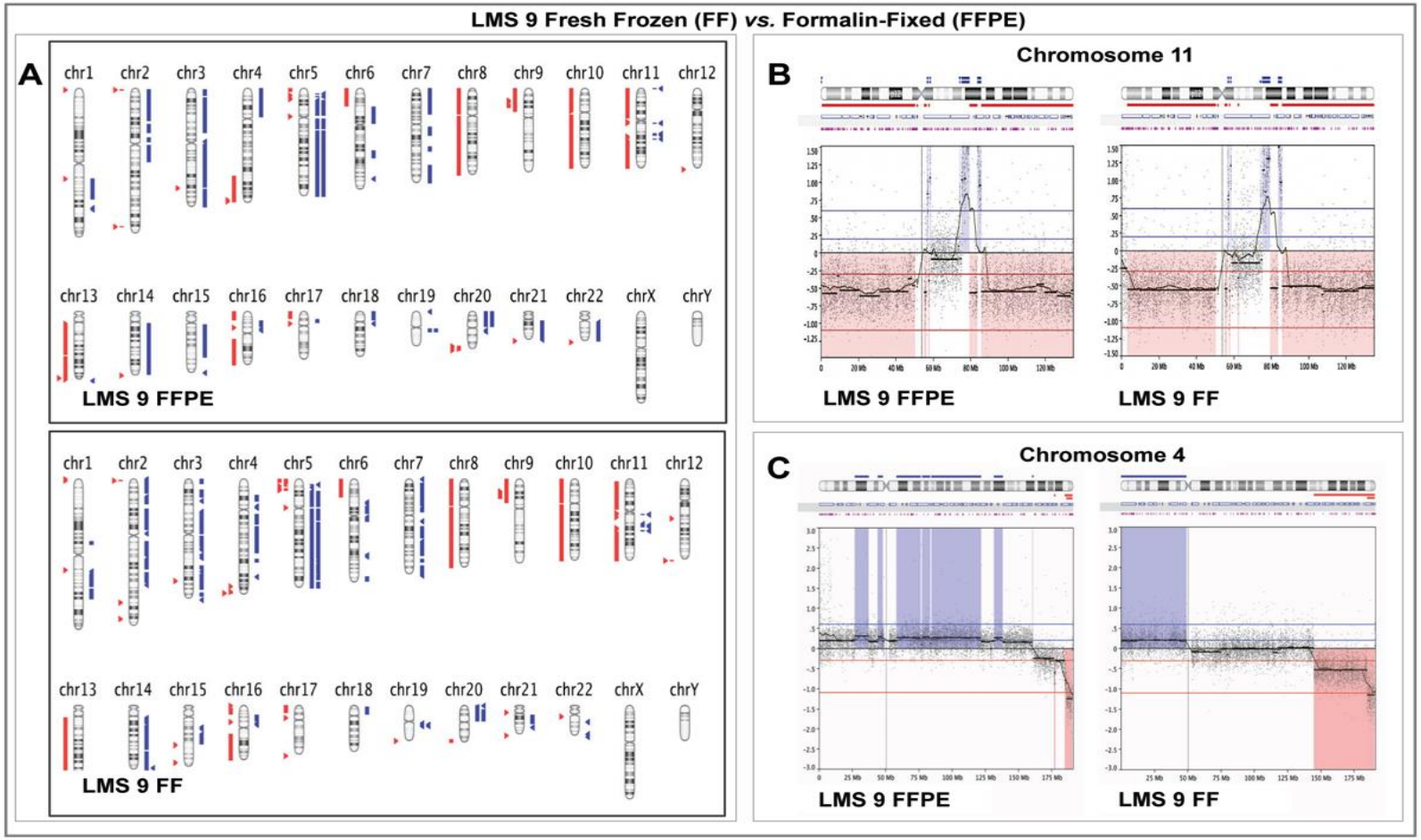

Figure 3.3: Comparison of Array CGH results in paired Fresh Frozen and Formalin-fixed Paraffin -Embedded samples from LMS 9

Panel A: Graphical whole-genome views of copy number aberrations (SCNAs) identified in both sample types showing close similarities on most chromosomes. **Panel B**: Higher resolution graphical views of Chromosome 11 showing the close similarity in gain and loss patterns detected in both sample types. **Panel C**: High-resolution views showing the most dissimilar SCNA pattern detected between both sample types on chromosome 4.
On Panel A, aberrations called by FASST2 algorithm are represented by blue triangles to the right (amplifications) and red triangles to the left (deletions) of corresponding chromosomes. Double blue and red triangles/lines represent high-level amplifications and two-copy deletions, respectively. On Panels B and C, dots represent individual probe log_2 ratios plotted as a function of their chromosomal position with a moving average of probe log_2 ratios (wavy dark blue line). Horizontal blue lines above zero line represent log ratio detection thresholds for single copy and high level amplifications and horizontal red lines below zero line represent log ratio detection thresholds for single and two-copy deletions. Aberration calls are represented by thick black lines with corresponding shaded blue areas above (amplifications) and red areas below (deletions) the zero line.

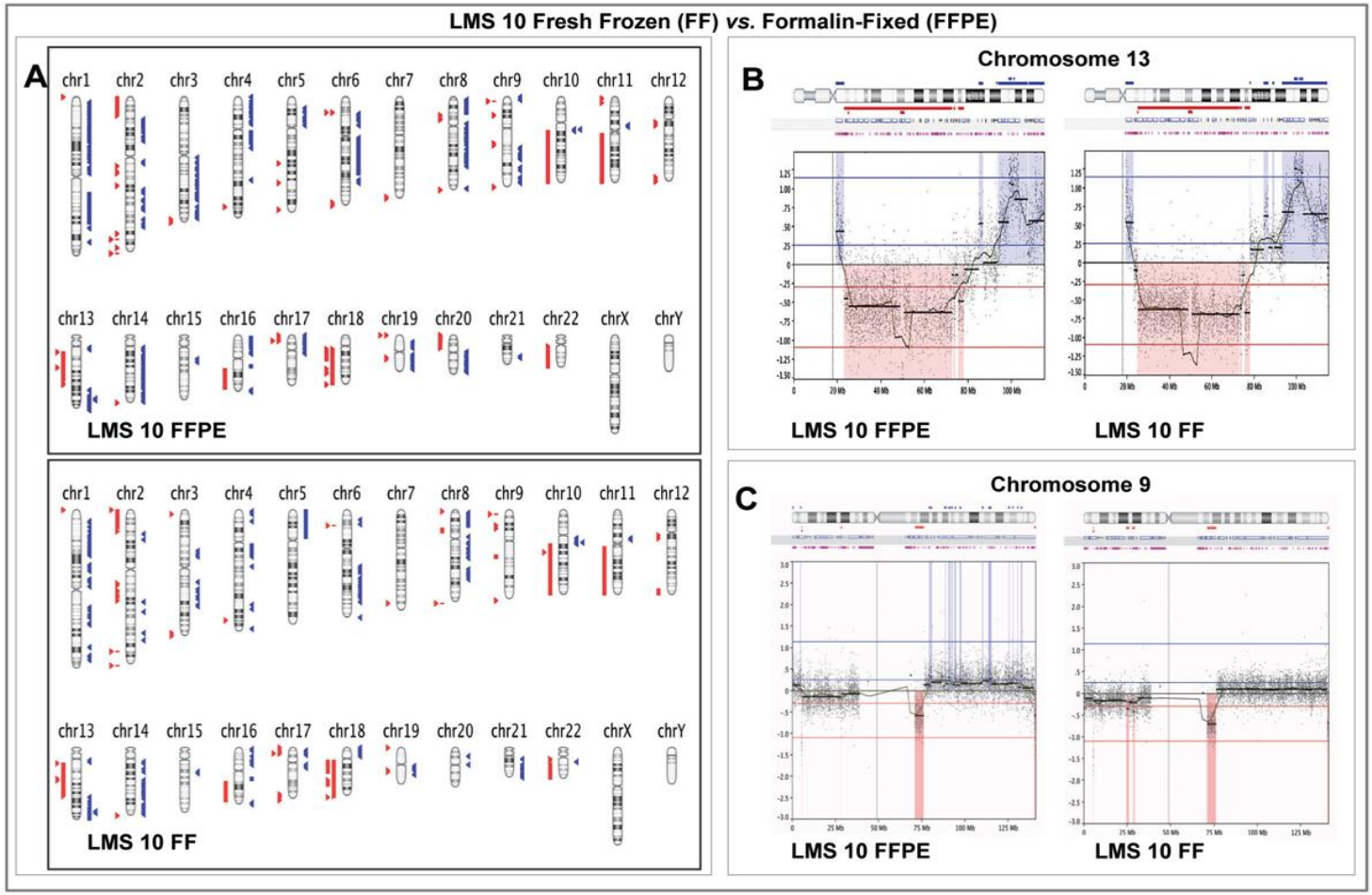

Figure 3.4: Comparison of Array CGH results in paired Fresh Frozen and Formalin-fixed Paraffin -Embedded samples from LMS 10

Panel A: Graphical whole-genome views of copy number aberrations (SCNAs) identified in both sample types showing close similarities on most chromosomes. **Panel B:** Higher resolution graphical views of Chromosome 13 showing the close similarity in gain and loss patterns detected in both sample types. **Panel C:** High-resolution views showing that even though the SCNAs identified by the calling algorithm on chromosome 9 are not identical, the moving averages of probe $\log_2$ ratios in both sample types remain similar.

On Panel A, aberrations called by FASST2 algorithm are represented by blue triangles to the right (amplifications) and red triangles to the left (deletions) of the chromosomes. Double blue and red triangles/lines represent high-level amplifications and two-copy deletion, respectively. On Panels B and C, dots represent individual probe $\log_2$ ratios plotted as a function of their chromosomal position with a moving average of probe $\log_2$ ratios (wavy dark blue line). Horizontal blue lines above zero line represent log ratio detection thresholds for single copy and high level amplifications and horizontal red lines below zero line represent log ratio detection thresholds for single and two-copy deletions. Aberration calls are represented by thick black lines with corresponding shaded blue areas above (amplifications) and red areas below (deletions) the zero line.

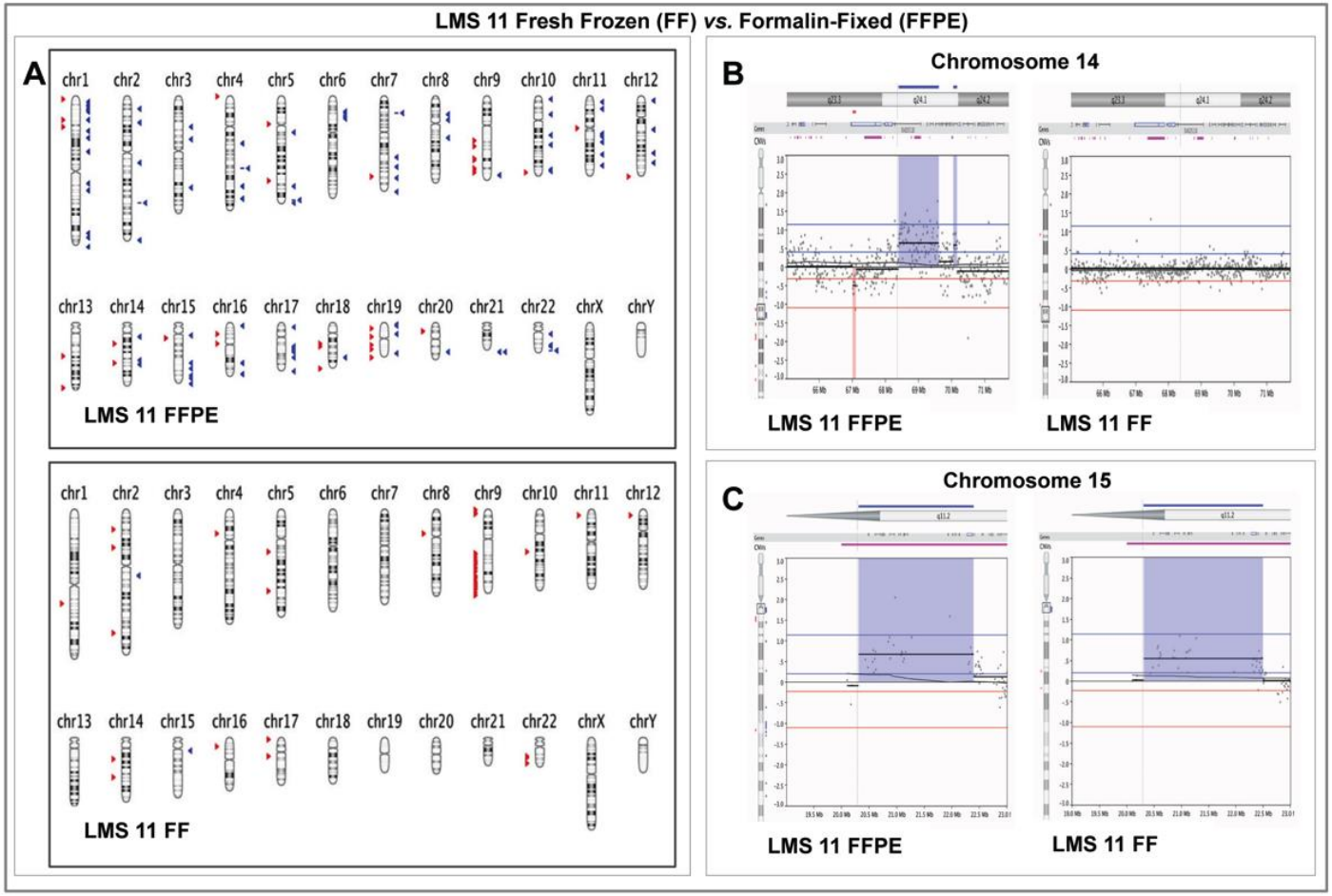

Figure 3.5: Comparison of Array CGH results in paired Fresh Frozen and Formalin-fixed Paraffin -Embedded samples from LMS 11

Panel A: Graphical whole-genome views of both sample types showing that majority of the copy number aberrations (SCNAs) identified in the macro-dissected FFPE sample were not detected in the FF sample. Deletions on the long arms of chromosomes 9, 14 and 15 as well as the short arm of chromosome 16 were the called on both sample types. **Panel B**: High resolution graphical views of a 6Mb region along on Chromosome 14 (14q24.1) showing a group of probes with an average $\log_2$ ratio of approximately 0.6 and the corresponding single copy amplification detected in the FFPE sample but no aberrations detected in the FF sample. **Panel C**: High-resolution graphical views showing a closely similar copy number aberration detected on Chromosome 15 (15q11.2) in both sample types with similar probe $\log_2$ ratios.

On Panel A, aberrations called by FASST2 algorithm are represented by blue triangles to the right (amplifications) and red triangles to the left (deletions) of the chromosomes. Double blue and red triangles/lines represent high-level amplifications and two-copy deletion, respectively. On Panels B and C, dots represent individual probe log2 ratios plotted as a function of their chromosomal position with a moving average of probe log2 ratios (wavy dark blue line). Horizontal blue lines above zero line represent log ratio detection thresholds for single copy and high level amplifications and horizontal red lines below zero line represent log ratio detection thresholds for single and two-copy deletions. Aberration calls are represented by thick black lines with corresponding shaded blue areas above (amplifications) and red areas below (deletions) the zero line.

3.2.5 Common Aberrations

Common aberration analysis of the 22 FFPE LMS cases was carried out to identify candidate driver SCNA. Overall, deletions were more common than amplifications. Statistical significance of common aberrations was determined using the GISTIC algorithm with recommended thresholds of G-score > 1.0 and q-bounds < 0.05 (discussed in detail in Chapter 4).

Five significant amplifications and five deletions were detected; involving around 500 genes in total were identified. Significant deletions involved important tumour suppressor genes like *PTEN*, *RB1* and *TP53*. Among important genes affected by the significant amplifications was *MYOCD*, a gene that is involved in smooth muscle differentiation. These significant SCNA confirm results of multiple previous array CGH studies in LMS and are summarised in Figure 3.6. Detailed results of other SCNA detected and candidate genes are presented in chapter 4.

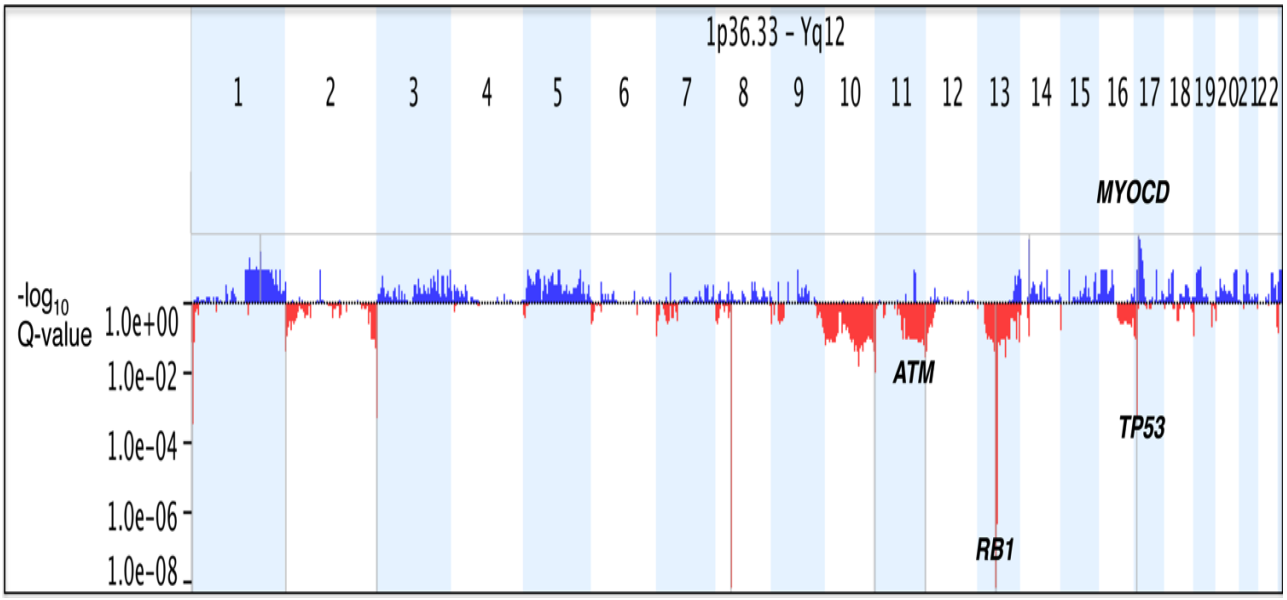

Figure 3.6: Statistically-Significant Common Genomic Copy Number Aberrations among 22 FFPE Leiomyosarcomas

Statistically significance of common aberrations was determined using the GISTIC algorithm. Commonly aberrant regions are plotted along the x-axis as a function of their chromosomal position and their q-values are plotted on the y-axis on a negative $\log_{10}$ scale so that the highest bars represent most significant genomic regions. Blue bars represent commonly amplified regions and red bars represent commonly deleted regions. Genomic regions with G-score > 10 and q-values < 0.05 are considered significant (shaded grey) and important candidate genes in these regions e.g. *RB1, MYOCD* are shown in black.

Aberrations in individual samples were called using FASST2 Algorithm.

3.3 DISCUSSION

Traditional large-scale projects for the evaluation of genome wide copy number aberrations in cancer were designed as prospective studies, which exclusively utilised fresh frozen tumour tissue. This was due to a perceived requirement of high quality DNA for the microarray-based methods that are typically utilised. Prospective study design is a problem when studying rare cancers such as LMS because it would take many years to accumulate large enough numbers for meaningful analysis. Tissue fixation in formalin is the standard procedure in most institutions and over many years, large FFPE tissue archives have been accumulated. Such archives are an essential source of tumour tissue for research and they come complete with associated clinical data such as disease progression, outcome and therapeutic responses, which can readily be correlated with molecular genetic data.

Formalin fixation, which is aimed primarily at preserving tissue protein structure for histopathological studies results in the formation of inter- and intra-strand cross-links between DNA molecules that produces low yields of highly fragmented nucleic acids (Srinivasan, Sedmak et al. 2002). Other effects, such as strand cleavage and base modifications make PCR amplification of whole genome DNA prone to bias and errors. As expected, DNA from the FFPE LMS tumours in this study showed varying degrees of fragmentation compared to that from FF tissues and the degree of degradation appeared to be worse with older samples (Figure 3.1). However, factors such as pre-fixation and intra-fixation durations and tissue penetration are known to influence the degree of formalin effects on tissue DNA (Srinivasan, Sedmak et al. 2002). Some of the more degraded samples in this study were fixed before standardized protocols for tissue fixation, to preserve their suitability for molecular studies, were widely-established. Two other LMS cases (among five such cases in the entire study) were initially processed in a different Pathology laboratory.

Initial attempts to label FFPE DNA using standard ULS protocols were not consistently successful. The use of a thermocycler with a heated lid gave higher degrees of labelling than heat blocks or circulating water baths for incubations during labelling, presumably because the temperature is more uniformly maintained throughout the labelling reaction. Insufficient dye amounts relative to DNA were also found to lead to variable or poor ULS labelling, and even small errors (from user or equipment) that caused underestimation of DNA concentration gave poor labelling results. Using an excess of dye relative to DNA in labelling reactions gave consistent good degree of labelling and successful arrays (Table 3.2). This is in keeping with results published in a recent study that showed that estimation of DNA concentration is critical to sample assessment for labelling (de Jong, Verbeke et al. 2011).

Regardless of sample age or fixation protocol, all aCGH experiments utilising FFPE tissue in this study showed good DLRS values when labelling was done using the modified protocols (Figure 3.2). A review of relevant literature did not reveal any studies that reported such consistent good DLRS values from ULS-labelled archival FFPE tumour DNA. A number of recent studies using FFPE tissues of similar age have reported variable DLRS values, the majority of cases reported having DLRS values between 0.3 and 0.5, with some as high as 1.0 (Braggio, McPhail et al. 2011, Oikawa,

Yoshiura et al. 2011, Hirsch, Camps et al. 2012). The reliability of results were further strengthened by the observation that when FFPE samples from female patients were hybridised against sex-mismatched DNA, the expected X-chromosome gain or loss detected had probe $\log_2$ ratios near the expected values, even in cases where there were few other SCNA detected (details are shown in Section 4.3.6).

Two out of three paired FF and FFPE samples from identical tumours that were compared showed good overall probe $\log_2$-ratio correlation with Pearson's coefficients similar to those reported from a similar study (de Jong, Verbeke et al. 2011). In these two cases, shown in figures 3.3 and 3.4, the SCNAs detected across the entire genome in the compared FF and corresponding FFPE samples were also similar. For a few chromosomes where the SCNAs detected by the calling algorithm were dissimilar, visual examination at a higher resolution showed that in most cases a similar moving average pattern was retained and the amplitude of average probe $\log_2$ ratios was close to the set thresholds for low-level (single copy) aberration detection, thus explaining why aberrations were differentially called in the two sample types. The calling algorithm thresholds set for analysis of array CGH data in this study were chosen based on previous studies in literature and regarded as valid as they enabled the detection of common aberrations among the LMS cases that concur with previous studies (discussed later in this section).

The third case (LMS 11) showed poor statistical correlation of the overall probe $\log_2$-ratios from both sample types and the results showed that although a few common SCNAs were seen, the majority of SCNAs detected in the FFPE sample were not detected in the FF one. A minority of the genomic regions in LMS 9 and 10 also showed significant difference in the probe moving average pattern in addition to SCNA detected. This prompted a revisit of the histology of all three tumours, which showed that LMS 11 was very heterogenous and contained large areas of haematoma and normal cells (not shown), while the former two were composed relatively homogenously of tumour cells. In any whole genome nucleic acid isolation, the presence of germ-line DNA from normal cells 'contaminating' a tumour sample can potentially mask genomic aberrations. The ability to macro-dissect tumour cells from the FFPE tissues apparently helps to reduce the masking of genomic copy number aberrations (Hostetter, Kim et al. 2010, Yau, Mouradov et al. 2010). One oligonucleotide array CGH study modelled mosaicism using leucocytes from patients with trisomy 21 diluted with those from normal patients. It was demonstrated that CNA would only be readily detected if present in around 30% of cells in a mosaic population with an average log ratio of around 0.21, and that this log ratio would reduce to 0.08 if only 10% of cell carried the aberration (Neill, Torchia et al. 2010). Another study utilising SNP arrays estimated the minimum percentage of aberrant cells at 25% (Gondek, Tiu et al. 2008). Heterogeneity of tumour cell populations between the two areas of the whole tumour that were sampled independently and represented by the FF and FFPE samples could potentially account for some of the low level differential aberrations detected in these paired samples in the cases that looked more homogenous on histological examination.

Common aberration analysis was carried out among the 22 FFPE LMS cases in this study. Frequent common aberrations detected were in concordance with those reported in previous studies. Deletions on 10q and 13q have been reported as frequent among LMS in numerous studies (Derre,

Lagace et al. 2001, Hu, Rao et al. 2005, Larramendy, Kaur et al. 2006, Ylipaa, Hunt et al. 2011). Previous work in the Rare Tumour Research Group (RTRG) demonstrated a frequent loss on 11q that involves the locus of the ATM gene (Ul-Hassan, Sisley et al. 2009) that is mirrored in the current study and subsequently confirmed by FISH. Focal genomic regions with statistically high frequency of copy number aberrations over the "background" aberration were identified. The results are very similar to those reported by Barretina et al, who showed deletions on 10p, 10q, 13q, 17p and an amplification on 17p as the most statistically significant common aberrations from data obtained from fresh frozen tumour samples (Barretina, Taylor et al. 2010). The regions identified contain loci for well-established tumour suppressor and cell cycle regulatory genes such as *PTEN*, *RB1* and *TP53*. A frequent focal amplification detected on 17p among the LMS cases in this study specifically covered the *MYOCD* gene locus. The same region was recently shown to be frequently amplified and over-expressed in at least one subset of LMS and is now widely considered to be a likely driver SCNA for well-differentiated LMS (Perot, Derre et al. 2009, Taylor, Barretina et al. 2011). Over a quarter of the cases presented in this study showed SCNAs on genomic regions that involve loci for at least three of the five genes mentioned above, and may well represent a subset of LMS, although the small number of cases does not allow any correlation with clinical data to have sufficient statistical power. The potential for expanding such a retrospective study to improve its statistical power cannot however, be overemphasized.

At present, the cost per sample of labelling DNA by the ULS method is less than that of the enzymatic method and labelling and clean-up is complete within one hour, compared with the enzymatic methods that require at least five times that duration. Most importantly however, in a 24-month period that has seen only three operable LMS cases treated in this centre with the possibility of obtaining fresh tissue, access to the FFPE archives has permitted the analysis of more than twenty cases for which progression and survival data is available.

In summary, this optimised protocol makes ULS-labelling a very reliable, cheap, fast method that provides access to long-term archival samples prepared using even non-standard protocols, and has proven to be applicable to samples processed in other laboratories. The results generated from this archival tissue are in close concordance with those from multiple previous studies that utilized fresh tumour tissue, adding to the data validity and it is now possible to carry out retrospective studies in LMS and other rare STS subtypes utilising high-resolution genomic copy number mapping, maximising its potential to unravel pathogenetic mechanisms.

CHAPTER FOUR

RESULTS: ARRAY CGH

4.1 INTRODUCTION

Elucidation of their molecular genetic characteristics has played an important role in the current understanding of the pathology of many STS subtypes and led to advances in their classification, diagnosis and even treatment (Taylor, Barretina et al. 2011, Fletcher, Bridge et al. 2013). However, the majority of STS subtypes are characterised by genetic instability as evidenced by pervasive genomic abnormalities, which appear to be random and as yet have not been useful in tumour characterisation (Barretina, Taylor et al. 2010). There is a strong association between exposure to ionising radiation and the development STS as well as a higher incidence of these tumours in individuals who carry inherited, inactivating mutations of DNA damage response genes. It is therefore likely that genomic copy number aberrations that lead to inactivation of pathways involved in cellular response of DNA damage are not only an important mechanism leading to genomic instability, they are strongly linked to the cancer phenotype (Davoli and de Lange 2011, Lord and Ashworth 2012).

Among cancers in general, certain aberrations have been shown to be recurrent and preserved even as the tumours evolve and accumulating evidence has led to the current view that genetic instability is an enabling characteristic that leads to the cancer phenotype and that the resulting recurrent somatic copy number aberrations (SCNA) are an important clue to pathogenetic mechanisms (Hanahan and Weinberg 2011, Taylor, Barretina et al. 2011). Analysis of recurrent SCNA has led to the identification of genes with roles in tumour induction and/or progression and even suggested novel therapeutic approaches in various cancers including lung cancer (Weir, Woo et al. 2007, Chitale, Gong et al. 2009), ovarian cancer (Eder, Sui et al. 2005, The Cancer Genome Atlas Research Network 2011), acute lymphoblastic leukaemia (Lahortiga, De Keersmaecker et al. 2007, Mullighan, Goorha et al. 2007), hepatocellular cancer (Zender, Spector et al. 2006), glioblastoma (The Cancer Genome Atlas Research Network 2008, Wiedemeyer, Brennan et al. 2008) and even STS (Taylor, Barretina et al. 2008, Barretina, Taylor et al. 2010).

In an integrative study of over 3000 cancer genomes, Beroukhim and colleagues found that in almost all cancer types the most frequent SCNA involved either short genomic regions (focal SCNA) or whole chromosome arms (gains or losses). Around 10% of the cancer genome on average is affected by focal SCNA, the frequency of which showed an inverse relationship with genomic region size. Focal SCNA were also far more likely to occur with high amplitude (many more copies or homozygous deletion), compared with whole-arm events (Beroukhim, Mermel et al. 2010). Focal SCNA are therefore statistically more likely to target specific genes and from a research point of view, the smaller genomic regions affected make target gene identification more feasible. On a background of overall genetic instability as seen in STS, the focal SCNA are likely to be numerous with the majority being random so-called ‘passenger’ aberrations that have no functional role. The challenge is to differentiate these from ‘driver’ SCNA that have consequences that contribute to the cancer phenotype (Beroukhim, Mermel et al. 2010).

High-resolution oligonucleotide array CGH is elegantly designed to address the problem of identification of recurrent focal SCNA, particularly those of such small size that they might have been missed by previous studies that utilised lower resolution methods such as spectral karyotyping, chromosomal CGH or even BAC arrays. It is also particularly suitable for the analysis of tumours such as STS where the majority of subtypes have characteristically complex karyotypes. This chapter details the results of copy number analysis and the shortlisting process for candidate 'driver' SCNA and affected genes in three STS subtypes comprising leiomyosarcoma (LMS), Gastrointestinal Stromal tumours (GIST) and Undifferentiated Pleomorphic Sarcoma (UPS).

A combination of approaches was used to assess statistical likelihood that SCNA were non-random events. Genes affected by these non-random SCNA represented a shortlist of candidate 'driver' genes that were examined individually in terms of the biological consequences of their aberrations. Evidence that deletion or amplification of potential candidate genes could contribute to the acquisition of any of the hallmarks of cancer (reviewed by Hanahan and Weinberg in 2011) meant that they were added to a final list of candidate genes. In addition, specific attention was paid to deletion of individual genes and pathways that play a role in DNA damage responses and maintenance of genomic integrity, as these could provide insights into the mechanisms for genomic instability of STS. Aberrations involving genes relevant to differentiation in cells of mesenchymal origin was an additional criterion that was used in selection of candidate genes, as these may provide clues to the biological origin of STS subtypes.

The overall strategy used for candidate gene identification in this study is summarised in Figure 4.1 and discussed in detail in subsequent sections. Array CGH analysis and aberration calling have previously been discussed in Chapter 3.

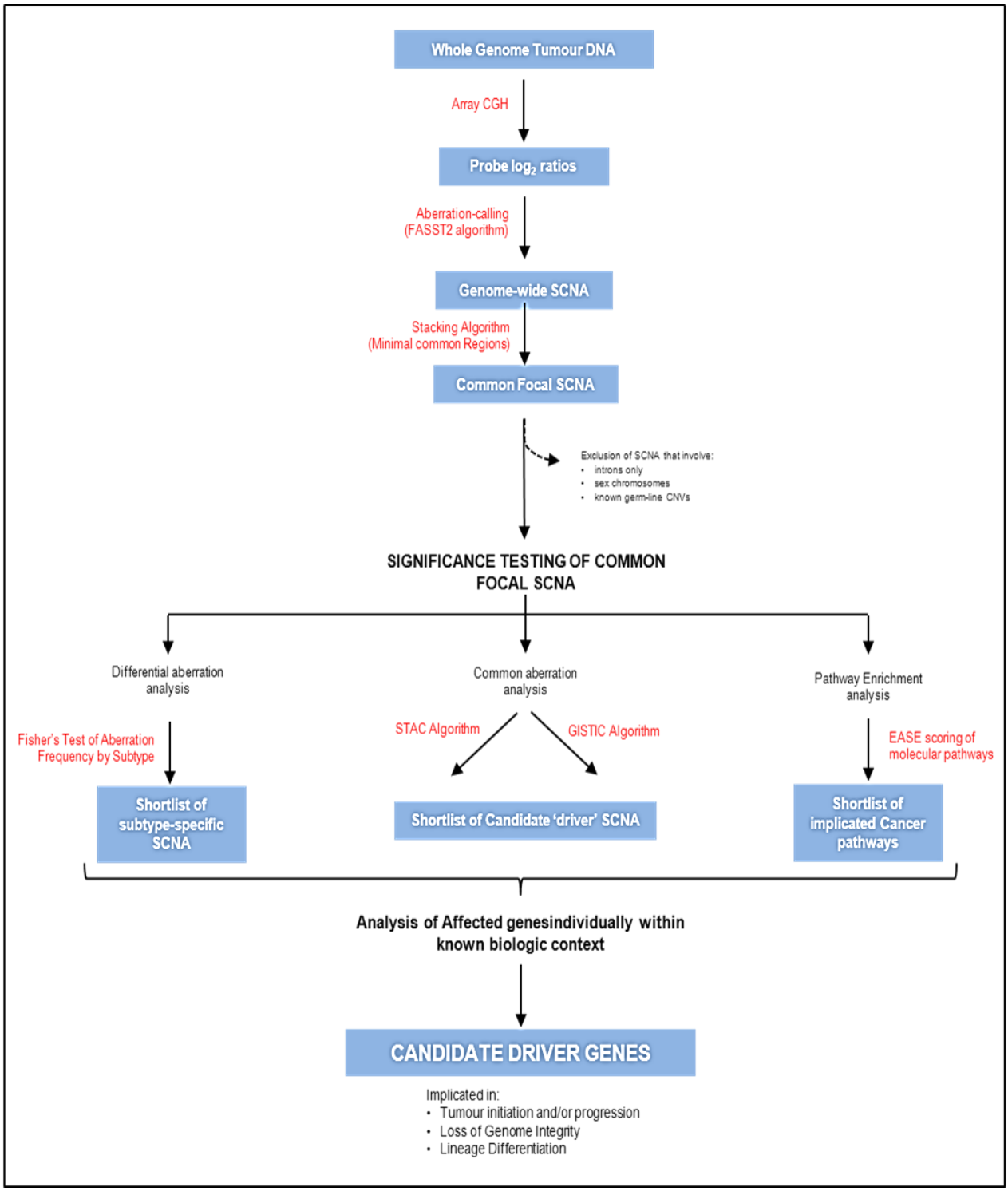

Figure 4.1: Overview of Work-flow for Array CGH Data Analysis and Candidate Gene identification

4.2 ARRAY CGH DATA ANALYSIS WORK-FLOW

4.2.1 Common Focal SCNA identification

For ease of reference, the strategy of Beroukhim and colleagues was adopted such that SCNA of a size larger than 5Mb (e.g. whole-arm events) were generally described as gains or losses to differentiate them from smaller focal SCNAs which were referred to as amplification or deletions (Beroukhim, Mermel et al. 2010). Using a stacking algorithm, all SCNA identified within specific genomic regions in a relevant set of STS cases were 'stacked' over one another and a frequency plot generated, as shown in Figure 4.2.

Common focal SCNA were identified as the minimal common region (MCR) of overlap among the SCNAs covering that locus (Figure 4.2). This region is generally regarded as being most likely statistically, to contain a targeted gene (Beroukhim, Mermel et al. 2010). The threshold frequency for common focal SCNA was set at 20% in order to increase the sensitivity of the data analysis.

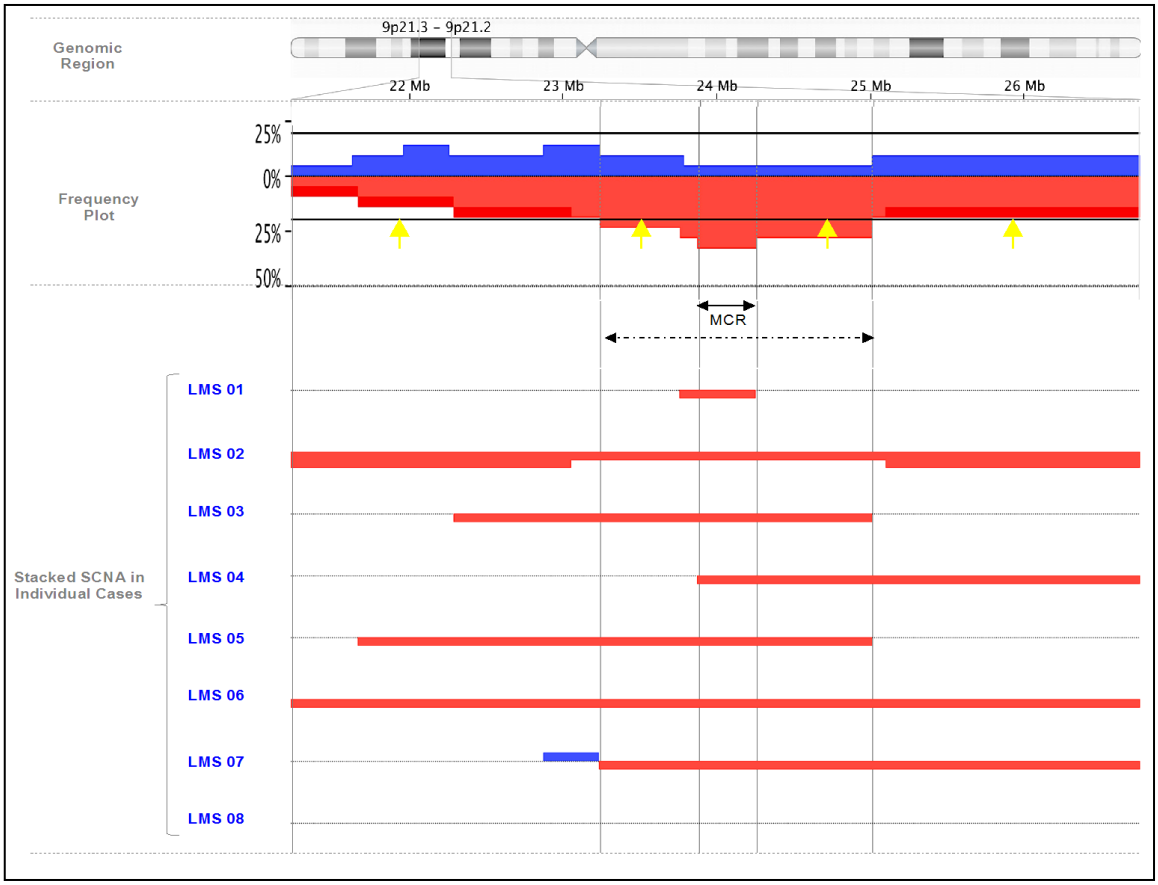

Figure 4.2: Frequency Plot and Stacked SCNA from Individual STS Cases showing an example of a Minimal Common Region (MCR)

Top panel shows the chromosomal region (9p21) and its approximate size. Middle panel shows the corresponding frequency plot of aberrations for all 8 LMS cases (bottom panel) plotted as percentages along the y-axis. Blue shading above zero line represent amplification frequency and red shading below zero line represents deletion frequency. Horizontal lines in bottom panel represent individual samples as shown on the left. Grey segments have normal copy number; blue segments above represent amplifications; while red segments below represent deletions.
The MCR (solid double arrow) is the smallest region that is aberrant in all the affected samples. It differs from the whole region of common aberration (indicated by the dotted double arrow) that is aberrant with a frequency higher than the threshold of 20% (indicated by the yellow arrows).
All SCNAs are detected using the FASST2 algorithm

4.2.2 Exclusion Criteria for Focal SCNA

4.2.2.1 Germ-line Copy Number Variations

Copy number variations (CNV) refer to DNA sequences that are found at different copy numbers in the germ-line DNA of two different individuals. They are generally defined as a form of structural variation in the human genome with a size > 0.5 to 1 kilobase (Valsesia, Mace et al. 2013). Although they are believed to have no direct pathological consequence, associations with complex genetic traits (multifactorial disorders) have been described (Conrad, Pinto et al. 2010). To differentiate then from common focal SCNAs therefore, comparisons were made between all SCNA identified and known CNVs from the Toronto Database of Genomic Variants (Iafrate, Feuk et al. 2004).

All common focal SCNAs that had 100% overlap with known CNVs were further examined at high resolution for the type and frequency of aberration, as well as SCNA breakpoints. Among this subset with 100% CNV overlap, SCNAs were concluded to be likely CNVs and excluded from subsequent analysis only if:

- they contained no known genes (unlikely to have functional consequences), or
- they showed within an identical pair of breakpoints, amplification in some samples and deletion in others (normal population variation), and/or
- both their 5' and 3' breakpoints coincided exactly with those of the known CNVs

Figure 4.3 shows an example of a SCNA identified at 1p31.1 among 16 pleomorphic sarcoma cases that was chosen for exclusion.

4.2.2.2 Sex Chromosomes

Reference DNA used for array CGH experiments on archival FFPE tissue was not matched for gender because the samples were initially anonymised. Overall, samples from seven female and six male patients were found to have been sex-mismatched.

In all the sex-mismatched cases where male reference DNA was used, the expected results of a single copy gain of the X chromosome and homozygous loss of the Y chromosome was confirmed by array CGH. An example is shown in Figure 4.4A, where apart from the sex chromosomes, no other genomic copy number aberrations were detected. The reverse (i.e. single copy loss of chromosome X and high level amplification of chromosome Y) was seen when female reference DNA was used (Figure 4.4B). These cases therefore served as positive control experiments. However, copy number data from sex chromosomes had to be excluded from subsequent common aberration analyses.

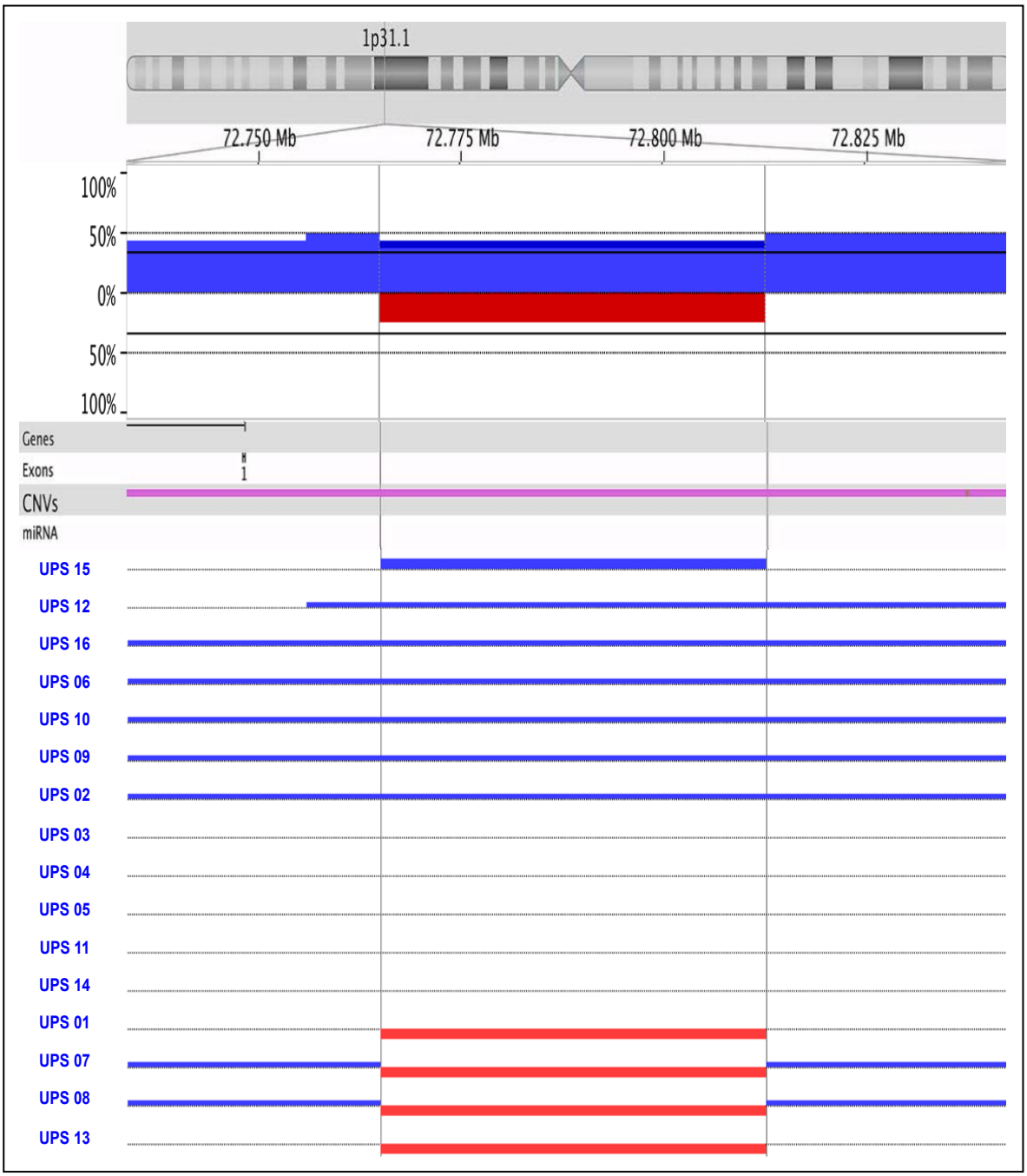

Figure 4.3: Frequency Plot and Individual Samples showing an example of an Aberrant Region that represents a Copy Number Variation (CNV)

The region below the vertical lines lies completely with in a known CNV region (purple track), shows variable aberration in different cases with identical breakpoints, and contains no gene loci. It was therefore concluded to be a likely CNV and excluded from further analysis. Top shows the chromosomal region, its approximate size and corresponding frequency plot of aberrations for all 16 UPS cases, plotted as percentages along the y-axis. Blue shading above zero line represent amplification frequency and red shading below zero line represents deletion frequency. Horizontal lines below represent individual samples arranged in decreasing order of SCNA log ratio with their identities shown on the left. Grey segments have normal copy number, blue segments represent gains and red segments represent losses. Vertical lines show exact alignment of SCNA breakpoints.

All SCNAs are detected using the FASST2 algorithm; known CNV region is from the Toronto Database of Genomic Variants

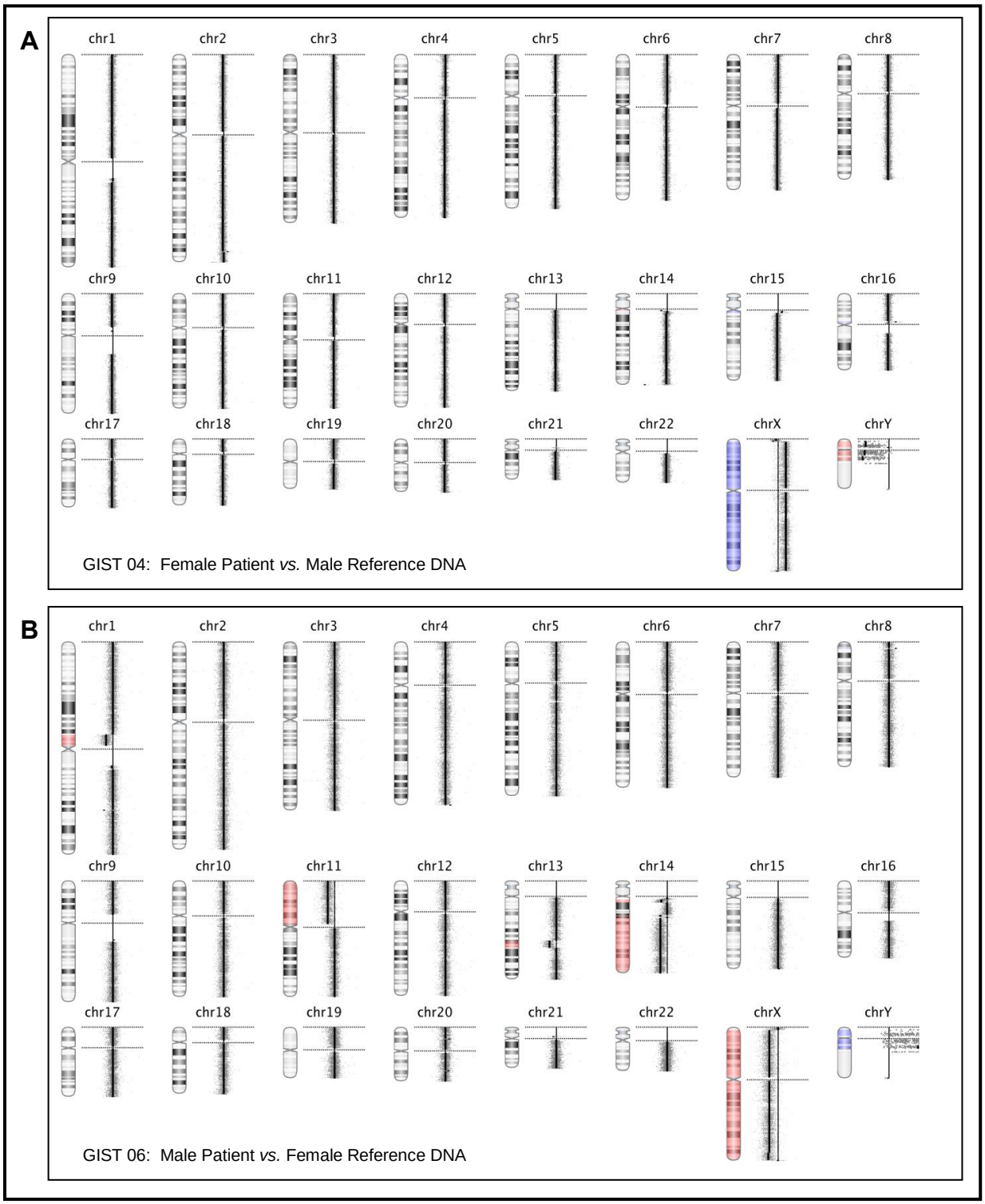

Figure 4.4: Genome View Ideograms of aCGH experiments using sex-mismatched reference DNA

Panel A – whole genome view of SCNA detected in tumour DNA from a female that was matched with male reference DNA showing the expected single copy gain of the X chromosome and homozygous loss of the Y chromosome. **Panel B** – shows a sex mismatched scenario opposite to that in Panel A and the expected SCNAs. Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the amplitude of log ratio.
All SCNAs are detected using the FASST2 algorithm

4.2.3 Significance Testing of Common Focal SCNA

4.2.3.1 Common Aberration Analyses

Two validated methods, STAC and GISTIC (see below) were used for statistical analysis of common focal SCNA to identify potential driver aberrations based primarily on their frequency of occurrence. Both tools are built into Nexus Software (Biodiscovery®) and utilise SCNAs that have already been identified using the FASST2 calling algorithm, as described in Chapter 3. They however apply different statistical approaches to significance testing (as summarised in Table 4.1), an overall strategy that makes the data analysis more robust (Rueda and Diaz-Uriarte 2010).

Significance Testing for Aberrant Copy Number (STAC)

Introduced by Diskin et al in 2006, STAC uses a global frequency statistic approach to attempt to identify a set of aberrations that are stacked on top of one another such that it would not occur randomly. It uses permutations of the SCNA in each arm of the chromosome to see how likely it is for a SCNA to occur at any location with a particular frequency. Using a p-value cut-off of 0.05, it then highlights the regions of common aberration that have a frequency that is higher than would be accounted for by random chance (Diskin, Eck et al. 2006).

Genomic Identification of Significant Targets in Cancer (GISTIC)

GISTIC incorporates in addition to aberration frequency, the amplitude of SCNA (log ratio values) for assignment of G-scores to each region. It then determines the probability of a score occurring by chance against a genome-wide random distribution of aberrations (Beroukhim, Getz et al. 2007) and applies a False Detection Rate correction for multiple sample testing with which it determines a q-value for that region. The program also identifies within these regions, 'peak' regions that have the highest statistical likelihood of containing affected genes (maximal G-score and minimal q-value).

Table 4.1: Differences between the STAC and GISTIC algorithms

	STAC	GISTIC
Criteria or Region Selection	Frequency of Aberrations Only	Both Frequency and Amplitude of aberration
Null model (for statistical significance)	Permutation of regions within a chromosome arm	Permutation of probes over the entire genome
Correction for Multiple Sample Testing	Does not require correction	Requires correction (FDR)
Peak Region identification	No	Yes

FDR – False Discovery Rate Correction

4.2.3.2 Enrichment Analysis of Affected Genes

Enrichment is a statistical approach that is used for the functional analysis of the long lists of genes, such as those affected by common focal SCNA (Huang da, Sherman et al. 2009). Based on their biological function, the affected genes were systematically mapped to their associated biological/molecular pathways. Statistical analysis was then carried out to identify pathways that were significantly over-represented (enriched) among the aberrant genes. In this study, the publicly available Database for Annotation, Visualisation, and Integrated Discovery (DAVID) Bioinformatics Tool v6.7 (http://david.abcc.ncifcrf.gov, accessed March 2013) was used for enrichment analysis.

The tool is able to handle gene lists with as many as 3000 annotated genes and determines the proportion of genes on the list that are involved in a particular molecular pathway (pathway mapping). It then compares this to the proportion of genes in the entire genome (population background) that are involved in the same pathway and uses a modified Fisher's Exact test (EASE score) to determine significance.

A simple ratio of these proportions known as Fold Enrichment is also calculated as a measure of the magnitude of pathway enrichment. This provides an idea of the gene distribution of the enriched pathways and fold enrichment values above 1.5 are generally considered to be 'interesting'. This is not a strict cut-off, however and fold enrichment is typically taken together with EASE score for a robust assessment of the statistical relevance of these pathways. For example, if 10% of genes on a list are involved in apoptosis versus only 1% of genes in the genome, the EASE score (p-value) is < 0.05 with a ten-fold enrichment and therefore the apoptotic pathway is considered to be very significant. (Huang da, Sherman et al. 2009).

Genes involved in amplified and deleted focal SCNA were analysed separately to facilitate biological interpretation of results. The affected genes were mapped to gene lists from curated pathways in the well-known and validated Kyoto Encyclopaedia of Genes and Genomes (KEGG) and Biocarta® databases, which include a wide range of molecular signalling pathways, including those with specific relevance to cancer. In addition to EASE score and fold enrichment, output from the DAVID tool also includes graphical representations of the significant pathways, highlighting all affected genes on the list submitted (as shown in Figures 4.7 - 9 and, 4.13 - 15).

4.2.3.3 Differential Aberration Analysis

Differential aberration analysis was used in an attempt to identify 'driver' genes that are specific for certain STS subtypes. Hypothetically, subtype-specific 'drivers' would be selected for in that STS subtype and the aberrations would occur at a higher frequency in that subtype relative to others. All the common focal SCNA identified in each STS subtype examined were therefore compared with those identified in a set of other STS subtypes.

Subtypes with few or no SCNA were excluded from the control group in order to add an additional level of stringency to the differential aberration analysis. Control groups for differential aberration analysis in this study therefore included STS that are known to have complex karyotypes as defined by Guillou and Aurias in their review (Guillou and Aurias 2009) and included leiomyosarcoma, undifferentiated pleomorphic sarcoma, myxofibrosarcoma, dedifferentiated and pleomorphic

liposarcoma, malignant peripheral nerve sheath tumours, pleomorphic rhabdomyosarcoma and angiosarcoma. Comparison of genomic regions with SCNA that were detected with a differential frequency (test subtype minus control group frequency) above a 40% threshold was carried out using a tool built into the Nexus® Copy Number Software. Fisher's exact test was used to assess the statistical significance of the difference in aberration frequency with $p < 0.05$ set as a cut-off.

4.2.4 Assessment of Shortlisted Candidate Genes

The statistical approaches above described generated more manageable shortlists of potential candidate genes (those involved in the statistically significant common SCNA, differential SCNA and enriched pathways). These genes were then individually examined in terms of biological function and reported involvement in cancer using the Atlas of Genetics and Cytogenetics in Oncology and Haematology (http://AtlasGeneticsOncology.org) as a starting point. A free, online database, the atlas contains detailed and curated information on genes, cytogenetics, and clinical entities in cancer, and cancer-prone diseases with PubMed links to relevant peer-reviewed articles and other resources (Huret, Ahmad et al. 2013). A final list of candidate genes was then compiled for each STS subtype based on the potential functional implications of the amplification or deletion in which they are involved, as well as reported abnormalities in various cancers with particular attention paid to STS. As previously discussed, functional assessment of the shortlisted candidate genes was based on known role in acquisition of cancer hallmarks as well as genome maintenance and mesenchymal cell lineage differentiation (Figure 4.1).

4.3 STS Subtypes and General Array CGH Profile Features

Array CGH was initially adapted to fresh tissue samples regardless of subtype and DNA from fresh frozen tissue samples from 32 cases was analysed. Following the optimisation of ULS labelling method, STS subtypes known to have complex karyotypes but no specific markers, including LMS and UPS were chosen for more extensive analysis. GISTs were chosen as a third subtype for analysis in order to follow up on previous work in the research group that suggested the presence of a shared genetic abnormality with LMS (Ul-Hassan, Sisley et al. 2009). DNA from archival FFPE samples for 45 cases were chosen for analysis. Three cases had both fresh and FFPE samples analysed and used for comparison and optimisation as described in Section 3.2. In total therefore, 74 separate STS cases, comprising 15 different clinico-biologic subtypes were analysed using array CGH as summarised on Table 4.2.

After exclusion of CNVs and sex chromosomes, featureless (flat) arrays were seen in one GIST (shown in Figure 4.4A), one WDLPS and 9 other cases including all three angiosarcomas, both ASPS, the synovial sarcoma, the EMCS, the SFT and Ewings' sarcoma. With the exception of the angiosarcomas and SFT, all the other STS subtypes where all cases had featureless arrays are known to have characteristic chromosomal translocations (Table 1.2). All of the 63 other cases had aberrations detected on at least one autosome. Eight out of the twelve liposarcoma cases (all 3 subtypes) showed the characteristic amplification on 12q13-15 involving the *MDM2* and *CDK4* genes (data not shown) and in some of these cases, this was the only aberration detected across the entire genome. Complex karyotypes with aberrations involving up to half of the autosomes were seen in

the MPNST, the PRMS and all 3 MFS cases. Detailed features of the GIST, LMS and UPS (48 cases in total) are discussed in subsequent sections of this chapter.

Table 4.2: Summary of Soft Tissue Sarcoma subtypes analysed by array CGH

STS Subtype	No. of FFPE Samples	No. of Fresh Samples	Number of Cases
Undifferentiated Pleomorphic Sarcoma (UPS)	12	4	16
Leiomyosarcoma (LMS)	21	3[§]	21
Gastrointestinal Stromal Tumour (GIST)	11	0	11
Pleomorphic Liposarcoma (PLPS)	1	3	4
De-differentiated Liposarcoma (DDLPS)	0	4	4
Well-differentiated Liposarcoma (WDLPS)	0	4	4
Myxofibrosarcoma (MFS)	0	3	3
Angiosarcoma	0	3	3
Alveolar Soft Part Sarcoma (ASPS)	0	2	2
Malignant Peripheral Nerve Sheath Tumour (MPNST)	0	1	1
Extraskeletal Myxoid Chondrosarcoma (EMCS)	0	1	1
Pleomorphic Rhabdomyosarcoma (PRMS)	0	1	1
Synovial Sarcoma	0	1	1
Ewing's Sarcoma	0	1	1
Solitary Fibrous Tumour (SFT)	0	1	1
Total	45	32	**74**

[§] *- Cases analysed as both Fresh and FFPE samples*

4.4 Array CGH Data Analysis by Subtype

Two out of fifteen STS subtypes (LMS and UPS) comprised half of the total 74 cases analyses by array CGH (Table 4.2), reflecting an intentional bias in sampling that favoured STS subtypes with complex and unbalanced karyotypes. Not only are these STS subtypes the best suited for analysis by array CGH, a method that only detects unbalanced genomic abnormalities, they constitute the majority of STS that lack defining genomic abnormalities.

Not surprisingly, the bias was also reflected in the results of common aberration analysis of all 74 cases. Using the GISTIC and STAC tools, the majority of significant common focal SCNA detected among the entire set were the same as those detected in the over-represented subtypes. The GISTIC algorithm detected in addition, aberrations that recurred in relatively few cases but with very large amplitude such as the high-level 12q14 amplifications that are characteristic of liposarcomas. The results (common aberration analysis only) were therefore unlikely to be useful for identification of novel common aberrations among STS in general and are not presented within this chapter, but available in Appendix 3.

Detailed analyses of SCNA detected among 3 STS subtypes including 21 LMS, 16 UPS and 11 GIST cases were carried out using the approach described above and summarised in Figure 4.1.

4.4.1 Copy Number Aberrations among Leiomyosarcomas

4.4.1.1 Leiomyosarcoma Cases

Although 22 cases, diagnosed as LMS were analysed by array CGH, only 21 of these were included for identification of common focal SCNA. The remaining case was excluded because it was re-classified as a GIST based on its genomic profile and confirmation by immunohistochemistry (discussed in detail in Section 4.5). In three of the cases LMS 08, 09 and 10 both fresh and FFPE samples were analysed by array CGH. However, only data from FFPE samples (for all 21 cases) was used for common aberration and subsequent analyses.

The tumour samples were obtained from patients aged between 39 and 82 (mean = 59) years at the time of diagnosis. There was a female predominance with 16 out of the 21 patients being female but only six tumours arose in the female genital tract even though most tumours occurred within in the abdominal or pelvic cavities. The majority (18 out of 22) were high grade (Trojani grade 3) with no obvious relationship to site or size. The features of all the LMS tumours analysed are summarised in Table 4.3.

Table 4.3: Summary of LMS Cases used for Identification of Common Focal SCNA

Case	Sampling Date	Gender	Age (years) at Diagnosis	Site	Size	Grade
LMS 01	2010	Female	63	Bowel	210	3
LMS 02	2011	Female	38	Lower Limb	60	3
LMS 03	2011	Female	47	Lower Limb	50	3
LMS 04	2011	Male	67	Bladder	85	3
LMS 05	2011	Female	63	Stomach	50	2
LMS 06	1999	Female	82	Lower Limb	60	3
LMS 07	1998	Male	37	Bladder	50	3
LMS 08	2011	Female	63	Vagina	50	3
LMS 09	2011	Male	39	Retroperitoneum	95	3
LMS 10	2011	Female	49	Pelvis	85	1
LMS 11	2011	Female	80	Stomach	60	3
LMS 12	2004	Female	58	Uterus	85	3
LMS 13	1995	Female	76	Bowel	40	3
LMS 14	1997	Male	77	Uterus	130	3
LMS 15	1997	Male	54	Uterus	160	3
LMS 16	1998	Female	69	Nose	50	2
LMS 17	2004	Female	54	Pelvis	15	3
LMS 18	2003	Female	49	Retroperitoneum	180	3
LMS 19	2008	Female	52	Uterus		3
LMS 20	2011	Female	69	Uterus	82	3
LMS 21	2000	Female	49	Lower Limb	35	1

4.4.1.2 Recurrent Gains and Losses

Only general observations were made with respect to gains and losses, because the focus of the study was on focal SCNA. Most LMS cases showed complex genomic profiles (with SCNAs detected on ten or more chromosomes). Only three cases (LMS 07, 12 and 20) had relatively simple genomic profiles. For all the LMS cases, copy number losses were more common than gains across the entire genome. The most frequent deletions (seen in > 40% of cases) include whole or near-whole arm deletion in 10p, 10q, 13q, 16q, and 11q as well as Amplifications on 1q and 17p. A penetrance plot summarising the frequency of chromosomal gains and losses is shown in Figure 4.5 below.

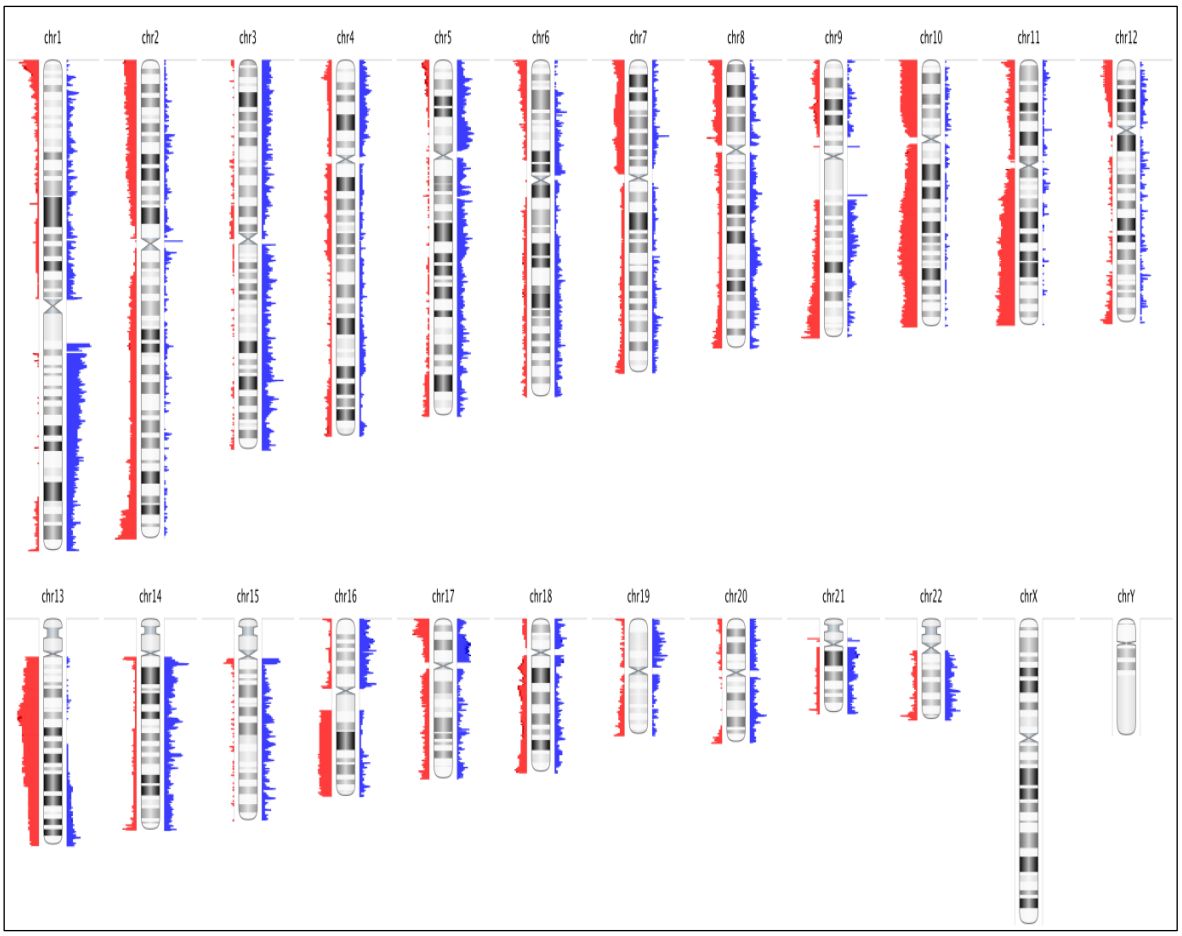

Figure 4.5: Frequency Plot of Common Genomic Copy Number Aberrations among 21 FFPE Leiomyosarcomas

Commonly aberrant regions are plotted as a function of their chromosomal position. Red bars to the left of the chromosome represent frequency of deletions and blue bars to the right of the chromosome represent amplifications. The heights of the bars correspond to the frequency of aberrations among the cases.

All SCNAs are detected using the FASST2 algorithm

4.4.1.3 Common Focal SCNA

A total of 1253 common focal SCNAs were detected among LMS with a minimum frequency of 20%. Covering approximately 20% of the genome, they ranged in size from 2kb to 7.3 Mb (mean = 317kB, median= 206kB). Of these, 96 regions contained no genes while another 18 represented CNVs and so were excluded from further analysis. Of the remaining, 720 regions had copy number amplifications (involving a total of 2031 gene loci) while 419 regions (involving 1842 gene loci) had copy number deletions. Deletion in 13q14.2 – q14.3 was one of the most frequent focal aberrations detected, present in 14 of the 21 LMS cases (67%). The 11q22.3 region covering the locus of the *ATM* gene was also frequently deleted (in approximately 43% of cases). A candidate gene identified in a previous study in our laboratory (Ul-Hassan, Sisley et al. 2009), the *ATM* gene locus deletion was confirmed using Fluorescence *in situ* Hybridisation (FISH) on two LMS cases, FFPE sections of LMS 13 (not shown) and in cells cultured from the primary tumour in LMS 08 (Figure 4.6).

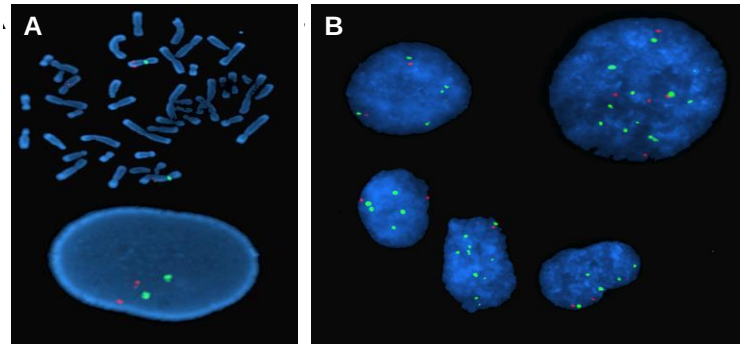

Figure 4.6: Two-colour Interphase Fluorescence in situ Hybridisation (FISH) images of nuclei of cultured leiomyosarcoma cells.

A – Metaphase and interphase nucleus derived from normal lymphocytes showing expected diploid complement with 2 copies each of the chromosome 11 centromere (green) and *ATM* locus (red). B - Cells were derived from culture of fresh tissue from LMS 08. Most cells have polysomy 11 (five or more green chromosome 11 centromere signals), but relatively few copies of the red *ATM* gene locus signal at 11q22 representing an overall copy number deletion. Nuclei are counter-stained with DAPI (blue).
Images were captured using x63 magnification objective

4.4.1.4 Significance Testing of Common Focal SCNA

STAC

Using the STAC approach, 38 common focal amplifications in LMS were found to be statistically significant (frequency ≥ 20%, $p < 0.05$). All but one genomic region involved at least one gene and the amplifications covered loci for a total of 165 genes. Table A1 in Appendix 4 lists all the genes affected by significant amplifications identified by STAC among LMS. Nine of these genes have been implicated with gain in copy number and/or function abnormalities in various cancers (highlighted in Table 4.4). Only one of these *ARHGEF2*, a putative oncogene which encodes a rho-guanine nucleotide GTPase exchange factor (rhoGEF) has previoulsy been implicated in STS (Frolov, Chahwan et al. 2003, Brecht, Steenvoorden et al. 2005). Another, *PIK3R1* that encodes the p85a regulatory subunit of phophatidyl inotisol-3-kinase (PIK3) has oncogenic properties but has also been reported as down-regulated, in a number of cancers (Taniguchi, Winnay et al. 2010). More recent findings in endometrial carcinomas however clearly demonstrate gain of function mutations in this gene (Cheung, Hennessy et al. 2011) that correlate with its amplification being a driver aberration.

Table 4.4: Significant Common Focal Amplifications among LMS (STAC)

	Cytoband	Region	Length (Mb)	Frequency (%)	P-Value	Genes	Cancer Genes
Chr 1	p12 - p11.2	120,563,106-120,756,223	0.19	48	0.000	1	*NOTCH2*
	q21.1	144,009,907-144,995,110	0.99	71	0.000	8	*PDE4DIP*
	q21.1	145,112,508-145,278,200	0.17	76	0.000	2	-
	q21.2	149,311,087-149,596,705	0.29	62	0.016	6	-
	q21.2	149,818,426-149,873,111	0.05	62	0.016	11	-
	q21.3	154,899,255-154,958,577	0.06	62	0.016	6	*CKS1B SHC1*
	q22	155,177,125-155,253,539	0.08	62	0.016	8	-
	q22	155,894,065-156,045,700	0.15	62	0.016	8	***ARHGEF2***
	q24.2	168,235,779-168,258,696	0.02	62	0.016	1	-
Chr 2	q22.3	145,162,871-145,304,149	0.14	29	0.016	1	-
	p22.2	37,823,758-38,245,167	0.42	33	0.009	2	-
	q11.2	97,527,264-97,623,299	0.01	38	0.000	2	-
	q13.13	48,766,079-49,005,661	0.24	52	0.000	3	-
Chr 3	q26.31	71,958,976-172,051,964	0.09	52	0.003	1	-
Chr 5	q32	148,717,819-148,929,994	0.21	43	0.043	5	-
	p13.2	37,012,156-37,067,386	0.06	52	0.015	1	-
	q11.2	54,728,700-54,928,310	0.20	43	0.043	3	-
	q12.3	64,264,133-64,412,850	0.15	43	0.043	1	-
	q13.1	67,067,899-67,548,574	0.48	43	0.043	1	*PIK3R1*
	q13.2	71,430,376-71,606,831	0.18	48	0.001	2	-
	q14.3	90,402,321-90,686,262	0.28	48	0.001	2	-
Chr 6	p22.2	26,124,423-26,259,251	0.13	38	0.007	17	-
Chr 9	p22.1	19,095,812-19,317,927	0.22	29	0.033	3	-
	q22.31	94,034,499-94,186,403	0.15	38	0.012	2	-
Chr 11	p13	35,110,040-35,316,017	0.20	29	0.006	2	*CD44*
	q12.3	62,314,338-62,641,063	0.33	24	0.003	35	-
Chr 12	q24.11	109,123,797-109,147,883	0.02	33	0.004	1	-
Chr 13	q34	110,918,305-111,046,505	0.13	43	0.000	2	-
Chr 14	q22.1	51,964,027-52,127,715	0.16	57	0.002	1	-
	q24.1	69,144,273-69,519,867	0.38	57	0.002	3	-
Chr 16	q22.1	69,228,790-69,753,047	0.52	29	0.043	10	*NQ01*
	q22.3	72,964,471-73,143,118	0.18	33	0.001	2	-
	q24.1	84,963,962-85,068,069	0.10	33	0.001	1	-
Chr 17	q22	55,948,928-56,056,018	0.11	38	0.023	2	-
	q23.1	57,816,707-57,935,524	0.12	38	0.023	1	-
Chr 19	p13.11	18,429,433-18,583,639	0.15	48	0.003	7	*ELL GDF15*
Chr 22	q12.3	36,665,486-36,786,805	0.12	48	0.013	1	*MYH9*
						165	11

Genomic regions shown have amplification frequency ≥ 20% and p < 0.05 as judged by the STAC algorithm. Only regions that involve at least one known gene locus are listed and only the affected genes that have been implicated in cancer are shown. Genes with reported gain of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red while those that have been described with the opposite abnormalities in cancer are shown in grey (likely passengers). Genes that have been implicated in STS are highlighted in bold red text.
STAC analysis was carried out using Nexus® Copy Number Software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

There were 31 significant common focal deletions identified among LMS (Frequency ≥ 20%, p < 0.05). Thirty of these involve one or more gene locus with 113 genes affected in total (listed on Table A2 in Appendix 4). Only four of the affected genes have been implicated in cancer and none of them have specifically associated identified in STS (Table 4.5). With the exception of *PRDM16*, whose knockdown has been shown to be involved in myogenic differentiation (Seale, Bjork et al. 2008), the remaining three genes have reported reduction in copy number or expression in various cancers.

Table 4.5: Significant Common Focal Deletions among LMS (STAC)

	Cytoband	Region	Length (Mb)	Frequency (%)	P-Value	Genes	Cancer Genes
Chr 1	p36.33	1,075,323-1,151,667	0.08	62	0.000	3	-
	p36.33	1,870,020-2,195,283	0.03	38	0.009	3	-
	p36.32	2,472,466-2,539,213	0.07	52	0.000	3	*TNFRSF14*
	q44	248,983,071-249,250,621	0.27	33	0.000	4	-
	p36.32	3,184,391-3,448,830	0.26	48	0.000	3	*PRDM16*‡
	p36.32	4,793,782-5,354,520	0.56	38	0.009	1	-
Chr 2	q36.3	229,013,827-229,907,621	0.89	48	0.024	2	-
	q36.3	230,795,681-230,895,774	0.10	48	0.024	1	-
	q37.1	233,213,865-233,427,159	0.21	52	0.001	10	-
	q37.1	233,903,041-234,336,019	0.43	48	0.024	6	-
	q37.1	234,507,893-235,056,307	0.55	48	0.024	14	-
	q37.2	235,880,185-236,368,103	0.49	52	0.001	1	-
	q37.3	241,538,072-241,782,200	0.24	57	0.000	5	-
	q37.3	242,503,450-242,825,417	0.32	57	0.000	10	*BOK*
Chr 4	p16.1	8,771,782-8,886,137	0.11	33	0.002	1	-
Chr 6	p25.3	1,696,679-1,919,998	0.22	38	0.019	1	-
Chr 8	q24.3	143,656,045-144,082,427	0.43	33	0.019	12	-
Chr 9	q34.11 - q34.12	133,145,616-133,553,257	0.41	43	0.038	3	-
	q34.3	137,870,941-138,031,761	0.16	48	0.002	1	-
	q34.3	140,630,365-140,833,019	0.20	57	0.000	2	-
Chr 11	p15.5	975,029-1,093,963	0.12	43	0.003	3	*MUC2*
Chr 12	p13.31	7,796,205-7,921,641	0.13	38	0.026	5	-
Chr 14	q32.33	104,561,057-104,763,502	0.20	33	0.002	2	-
Chr 19	p13.3	0-307,937	0.31	29	0.002	6	--
Chr 20	q13.33	60,072,327-60,546,948	0.47	29	0.000	1	-
	q13.33	61,064,386-61,378,893	0.31	24	0.023	2	-
	q13.33	61,638,941-61,818,375	0.18	24	0.023	2	-
	q13.33	62,844,161-63,025,520	0.18	33	0.000	2	-
Chr 21	q22.3	45,818,895-45,951,075	0.13	33	0.000	3	-
Chr 22	q13.32	49,017,773-49,105,738	0.09	52	0.000	1	-
						113	4

Genomic regions shown have deletion frequency ≥ 20% and p < 0.05 as judged by the STAC algorithm. Only regions that involve at least one known gene locus are listed and only the affected genes that have been implicated in cancer are shown. Genes with reported loss of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red. Genes that have been implicated in STS are highlighted in bold red text.

‡ *- gene with role in mesenchymal lineage differentiation*

STAC analysis was carried out using Nexus® Copy Number Software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

It was notable that so few candidate genes were identified among the deleted focal SCNA in LMS compared to the amplifications. The STAC algorithm assesses significance of focal SCNA against the background aberration in chromosome arm on which it is located; and whole arm losses were more common than gains among LMS (Figure 4.5). It is therefore presumable that the high frequency of losses (copy number loss > 5Mb in size) made some of the deletions less significant based on the algorithm's criteria. This adds further justification to the use of more than one statistical approach to determination of focal SCNA significance.

GISTIC

The GISTIC algorithm determined a total of 5 amplifications and 5 deletions to be significant in LMS (G-score >1.0, q-bound <0.05). The significant focal SCNA ranged in size from 0.5 to 18.6Mb and had peak regions of between 53 and 498kb in size (summarised on Tables 4.6 and 4.7). The significantly amplified focal SCNAs occurred with a higher frequency in general (median frequency = 67%) compared with the significant deletions (median frequency = 52%) but had lower G-scores (mean amplification score 7.1 *vs*. mean deletion score 12.1).

The significant amplifications involved loci for a total of 135 genes (listed in Table B1 of Appendix 4). Five of these genes have been implicated in cancer, but *PDE4DIP* and *MYOCD* have been reported with the appropriate abnormalities to qualify them as candidate drivers in LMS. *MYOCD* is particularly notable because it is involved in smooth muscle differentiation and was shown to be amplified and over-expressed in a subset of LMS (Perot, Derre et al. 2009). It was also the only gene located in the peak region of its amplicon (Table 4.6).

Table 4.6: Significant Common Focal Amplifications among LMS (GISTIC)

	Cytoband	Region	Length (Mb)	Frequency (%)	G-Score	Q-Bound	No. of Genes	Genes in Peak	Cancer Genes
Chr 1	1q21.1	144,009,907-145,492,482	1.5	81	6.85	0.03	15	*SEC22B NOTCH2NL*	*PDE4DIP*
	1q42	228,255,947-231,657,562	3.4	24	6.49	0.05	36	*GALNT2*	-
Chr 3	3q25.1-26.1	150,855,624-165,822,232	15	67	7.41	0.02	55	-	*GMPS MLF1*
Chr 14	14q11.2	22,394,168-23,535,702	1.1	76	6.64	0.04	16	-	-
Chr 17	17p12	11,235,749-14,746,194	3.5	43	8.08	0.02	13	*MYOCD*	*MAP2K4* ***MYOCD***
							135	4	5

Genomic regions shown have amplification G-score ≥ 1.0 and q < 0.05 as judged by the GISTIC algorithm. Only affected genes that have been implicated in cancer are shown in the last column. Genes with reported gain of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red while those that have been described with the opposite abnormalities in cancer are shown in grey (likely passengers). Cancer-implicated genes with as yet undetermined or inconsistent functional status are shown in black. Genes that have been implicated in STS are highlighted in bold red text.
GISTIC analysis was carried out using Nexus® Copy Number software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

Like the large genomic region losses, the significant focal deletions identified by GISTIC covered larger regions of the genome than amplifications and involved nearly 350 genes (summarised in Table B2 of Appendix 4). Twenty-five of these genes have reported functions that led to their selection as candidate drivers. Two of these - *PRDM16* and *CEP55* were selected because of their involvement in mesenchymal cell differentiation and maintenance of genome integrity, respectively (Table 4.7)

Fourteen of the deleted genes were present within peak regions of the significant SCNA. Five of these were identified candidate genes by functional assessment. Notable genes located in the peak regions include the well-known tumour suppressor genes *PTEN* and *TP53* (both deleted in 52% of cases). Other notable tumour suppressor genes identified as candidates (but not located in peak regions) include *RB1* and *FAS* (Table 4.7).

Table 4.7: Significant Common Focal Deletions among LMS (GISTIC)

	Cytoband	Region	Length (Mb)	Frequency (%)	G-Score	Q-Bound	No. of Genes	Genes in Peak	Genes in Atlas
Chr 1	1p36	2,285,830-4,637,196	2.4	48	9.69	0.00	23	*PRDM16* *ARHGEF16*	*TNFRSF14* *DFFB* *PRDM16*‡
Chr 2	2q37	233,732,597-243,199,373	9.4	67	13.59	0.00	88	-	*BOK* *ING5* *INPP5D*
Chr 10	10q22.3-24.2	81,212,702-99,855,478	18.6	52	7.97	0.03	133	*ATAD1* *CFL1P1* *KLLN* *PTEN* *RNLS*	*BMPR1A* *LGI1* *RBP4* *TNKS2* *KLLN* ***PTEN*** *FAS* *CEP55**
Chr 13	13q14	46,723,507-53,957,341	7.2	67	18.78	0.00	53	*CDADC1* *CAB39L* *SETDB2* *PHF11*	***RB1*** *ARL11* *INTS6* *PHF11* *RCBTB1* *RCBTB2*
Chr 17	17p13.1	7,111,674-7,650,119	0.5	52	11.18	0.00	47	*TP53* *WRAP53* *EFNB3*	*TP53* *CLDN7* *DLG4* *KCTD11* *TNFSF12*
							344	14	25

Genomic regions shown have deletion G-score ≥ 1.0 and q < 0.05 as judged by the GISTIC algorithm. Only affected genes that have been implicated in cancer are shown in the last column. Genes with reported loss of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red. Genes that have been implicated in STS are highlighted in bold red text.

** - gene with potential role in maintenance of genome integrity*

‡ - gene with role in mesenchymal lineage differentiation

GISTIC analysis was carried out using Nexus® Copy Number software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

4.4.1.5 *Enrichment Analysis of Affected Genes*

Amplified Genes

Enrichment analysis of the 2383 genes involved in common focal amplifications in LMS (Section 4.4.1.3) identified a total of 29 molecular pathways with genes that were significantly over-represented (EASE p-value < 0.05) as summarised in Table C1 of Appendix 4. Detailed examination of the over-represented pathways showed that nine of them have functions that are directly related to the cancer phenotype (presented in Table 4.8). Among these were three pathways relevant to specific cancers, including chronic myeloid leukaemia, small cell and non-small cell lung cancer. Signalling pathways relevant to these and various other cancers including the Epidermal Growth Factor and Mitogen-activated protein kinase pathways (shown in Figure 4.7) were also over-represented. The Focal adhesion and Rho signalling pathways may be important for cell motility and cell-matrix interaction involved in tumour progression.

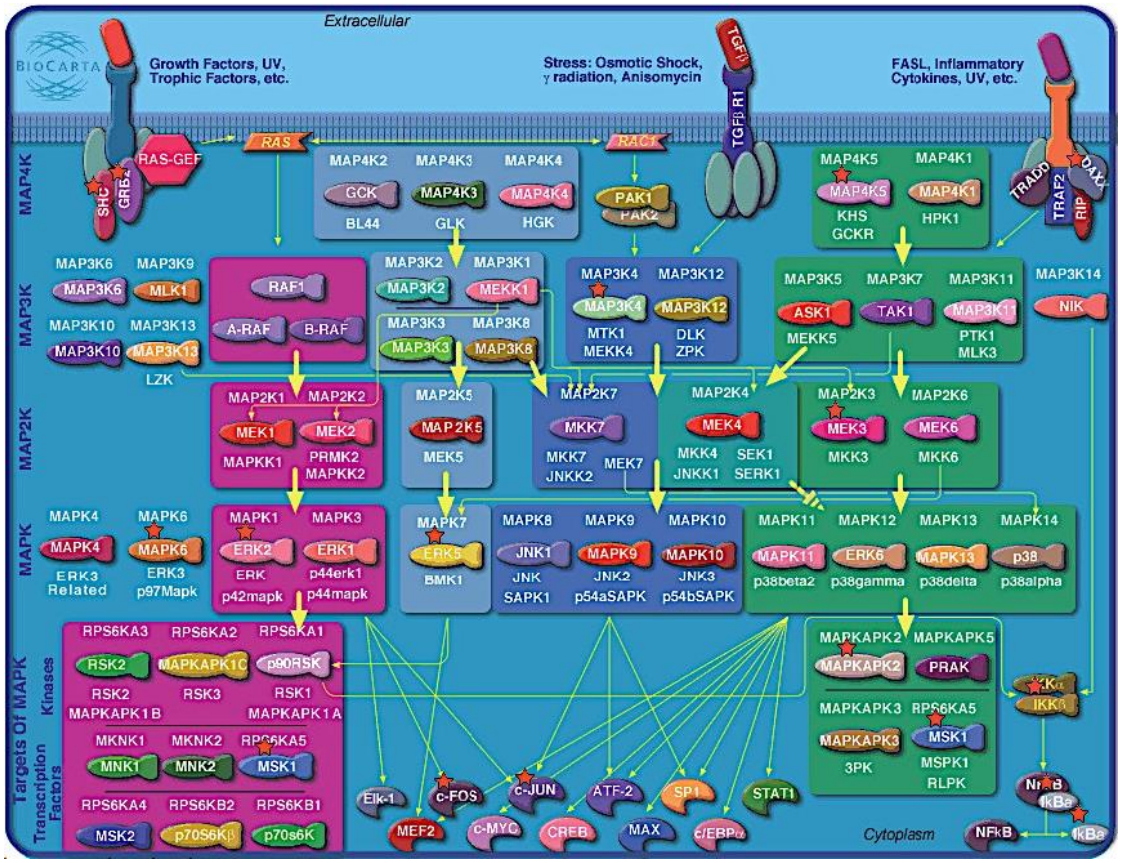

Figure 4.7: Graphic output showing enrichment of members of the Biocarta® Mitogen-activated Protein Kinase (MAPK) Signalling pathway among genes frequently amplified in LMS

All genes involved in amplified focal SCNA with a frequency ≥ 20% among LMS were submitted and the MAPK Signalling was significantly enriched (EASE p-value <0.05). Yellow arrows represent activation. Pathway members coded for by genes deleted in LMS are highlighted by red stars.

Pathway and highlighted genes are graphical results automatically generated following enrichment analysis by the DAVID analysis tool (http://david.abcc.ncifcrf.gov, accessed March 2013). Pathway information is based on Biocarta® database (http://www.biocarta.com/pathfiles/h_mapkpathway.asp, accessed March 2013).

Table 4.8: Cancer-related Molecular Pathways significantly over-represented in commonly amplified focal SCNA among LMS

Pathway ID	Term	Gene Count	Fisher p-value	EASE P-Value	Genes Involved	Fold Enrichment
BIOCARTA h_ptenPathway	PTEN-dependent Cell Cycle Arrest and Apoptosis	7	4.6E-3	0.0181	*MAPK1* *BCAR1* *SHC1* *PIK3CA* *FOXO3B* *GRB2* *PIK3R1*	3.13
BIOCARTA h_rhoPathway	Rho cell motility signalling pathway	7	1.2E-2	0.0276	*DIAPH1* *SRC* *ARHGEF11* *PIP5K1B* *ROCK1* *ARHGAP5* *MYLK*	2.87
BIOCARTA h_egfPathway	EGF signaling pathway	7	1.3E-2	0.0399	*JUN* *SHC1* *PIK3CA* *GRB2* *FOS* *SRF* *PIK3R1*	2.65
BIOCARTA h_integrinPathway	Integrin Signaling Pathway	9	5.4E-3	0.0171	*MAPK1* *ROCK1* *GRB2* *FYN* *JUN* *BCAR1* *RAP1A* *SHC1* *SRC*	2.60
KEGG: hsa05222	Small cell Lung Cancer	21	9.9E-5	0.0003	*BCL2* *LAMC2* *TRAF4* *MYC* *LAMC1* *E2F2* *CCNE1* *IKBKB* *FHIT* *PIK3R1* *PIK3CA* *LAMA3* *LAMA4* *NFKBIA* *PTGS2* *CKS1B* *BCL2L1* *RARB* *RXRB* *COL4A2* *COL4A1*	2.39
KEGG: hsa05220	Chronic Myeloid Leukaemia	16	3.9E-3	0.0093	*SHC4* *MYC* *E2F2* *IKBKB* *GRB2* *PIK3R1* *MAPK1* *SMAD3* *SHC1* *PIK3CA* *SOS2* *SHC3* *NFKBIA* *TGFBR2* *BCL2L1* *CBLB*	2.04
KEGG hsa05223	Non-small cell lung cancer	11	2.2E-2	0.0495	*MAPK1* *SOS2* *RASSF1* *PIK3CA* *FOXO3B* *E2F2* *GRB2* *RARB* *FHIT* *RXRB* *PIK3R1*	1.95
KEGG hsa04510:	Focal adhesion	40	3.4E-5	0.0001	*CAV2* *DIAPH1* *GRB2* *BCAR1* *COL3A1* *ITGA11* *ITGB5* *SRC* *ARHGAP5* *TNR* *BCL2* *SOS2* *COL6A2* *PIK3CA* *COL6A1* *SHC1* *SHC3* *THBS2* *THBS3* *PIK3R1* *SHC4* *THBS4* *COL4A2* *COL4A1* *VAV3* *ROCK1* *MYLPF* *ACTN1* *COL5A2* *MAPK1* *ITGA9* *LAMA4* *LAMA3* *FYN* *ITGA5* *JUN* *RAP1A* *LAMC2* *LAMC1* *MYLK*	1.90
BIOCARTA h_mapkPathway	MAP kinase signalling pathway	15	7.9E-3	0.0249	*JUN* *MAP2K3* *MAP3K4* *MAPK6* *IKBKB* *GRB2* *MAPKAPK2* *MAPK1* *SHC1* *DAXX* *NFKBIA* *MAP4K5* *MAPK7* *FOS* *RPS6KA5*	1.85

All genes involved in amplified focal SCNA with a frequency ≥ 20% among LMS were submitted for enrichment analysis. Only cancer-related pathways with EASE p-value < 0.05 are shown (in decreasing order of fold enrichment). All affected genes are listed with genes with gain of copy number and/or function abnormalities in cancer (candidate driver genes) shown red. Genes that have been described with deletion or loss of function in cancer are shown in grey (likely passengers).

Enrichment analysis was carried out using DAVID Bioinformatics tool (http://david.abcc.ncifcrf.gov, accessed March 2013) based on information from Biocarta® (http://www.biocarta.com/genes/index.asp, accessed March 2013) and KEGG® (http://www.genome.jp/kegg/pathway.html, accessed March 2013) databases.

Surprisingly, the PTEN dependent cell cycle arrest and apoptosis pathway was identified as being over-represented among the amplified genes in LMS with the highest enrichment score. Amplifications leading to increased overall activity of this pathway (i.e. cell cycle arrest and apoptosis) would not be in keeping with a malignant phenotype. However, functional examination of the 7 genes on the pathway that were actually amplified in LMS showed that they play a negative regulatory role to - or are inhibited by the *PTEN* tumour–suppressor (Figure 4.8). Most notable among these are *PIK3CA* and *PIK3R1* that encode the 110α and 85α- subunits respectively, of phosphatidyl inositol–3-kinase (PIK3), which functions in signal transduction following activation by growth factor receptor tyrosine kinases and is antagonised by the *PTEN* product. PIK3 activity is frequently deregulated in cancer it is considered an important druggable target (Willems, Tamburini et al. 2012).

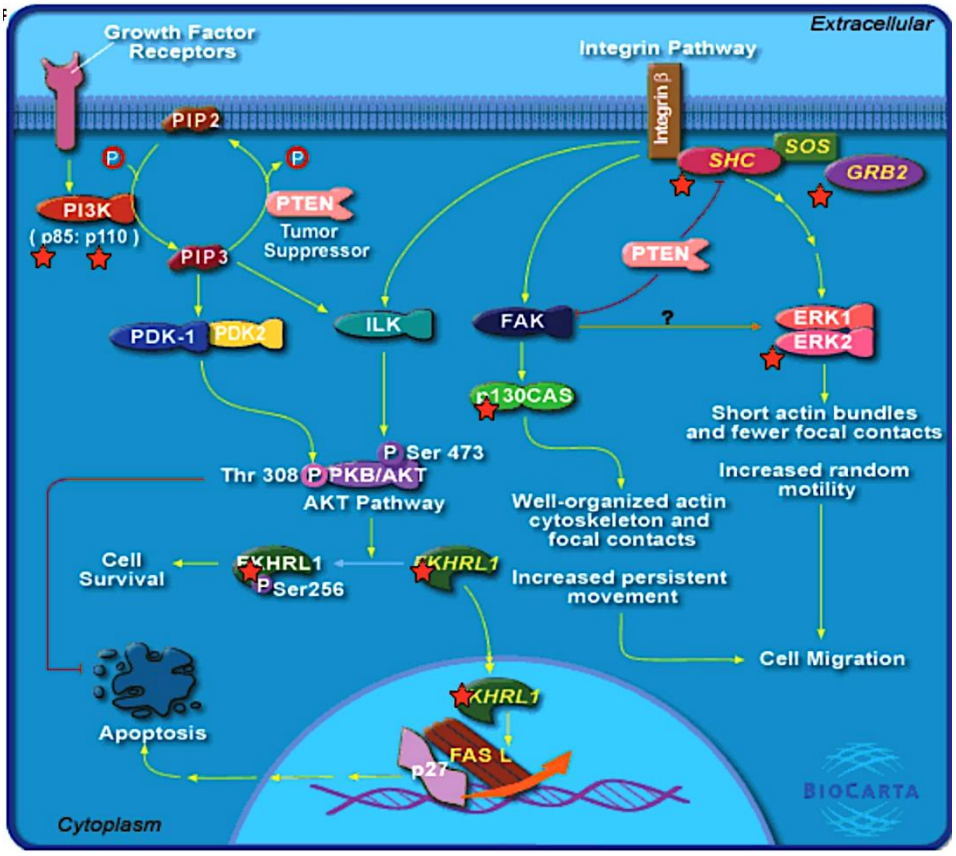

Figure 4.8: Graphic output showing enrichment of members of the Biocarta® PTEN-dependent cell cycle arrest and apoptosis pathway among genes frequently amplified in LMS.

All genes involved in amplified focal SCNA with a frequency ≥ 20% among LMS were submitted for enrichment analysis and the pathway was significantly enriched (EASE p-value <0.05). Yellow arrows represent activation while red lines represent inhibition. Pathway members coded for by genes amplified in LMS are highlighted by red stars. Although the pathway is named for the role of the PTEN tumour suppressor, the members of the pathway affected by amplifications in LMS (with the exception of FKHRL1) function to increase cell survival, cause cell migration and inhibit apoptosis. Within the pathway, they are directly or indirectly negatively regulated by the PTEN tumour suppressor. The affected genes are listed in Table 4.8. PI3K subunits p85 and p110 are coded for by *PIK3R1* and *PIK3CA* respectively. SHC, GRB2, P130Cas and ERK2 are encoded by *SHC1*, *GRB2*, *BCAR1* and *MAPK1* respectively. *FOXO3B* is a pseudogene of *FOXO3A*, a known tumour suppressor that encodes FKHRL1.

Pathway and highlighted genes are graphical results automatically generated following enrichment analysis by the DAVID analysis tool (http://david.abcc.ncifcrf.gov, accessed March 2013). Pathway information is based on Biocarta® database (http://www.biocarta.com/pathfiles/h_ptenpathway.asp, accessed March 2013).

Deleted Genes

A total of eight pathways were found to be significant (EASE p-value < 0.05) among 2219 commonly-deleted genes in LMS (summarised in Table C2 in Appendix 4). Only one, the regulation of transcriptional activity by PML is relevant to cancer pathogenesis. The key member of the PML pathway (shown graphically in Figure 4.9) is encoded by the pro-myelocytic leukaemia gene, *PML* and forms large protein complex known as the PML nuclear bodies (PML–NB). One of the frequently deleted genes among LMS, *SUMO1* encodes small ubiquitin-like modifier (SUMO) proteins, which are involved in post-translational modification of the PML protein that is required for interaction with other nuclear proteins such as p53, RB and DAXX. De-acetylation of p53 mediated by SIRT1 (whose encoding gene *SIRT1* was also commonly deleted gene in LMS) facilitates its interaction with the PML-NB (Appella and Anderson 2001, Lallemand-Breitenbach and de Thé 2010).

The interactions of the PML-NB with p53 and RB1 have a positive regulatory effect on their transcriptional activation and both encoding tumour-suppressor genes, were frequently deleted in LMS. DAXX on the other hand causes transcriptional repression, which is abrogated by interaction with PML-NB. This PML-DAXX interaction is one of the mechanisms of Fas-mediated apoptosis (Lallemand-Breitenbach and de Thé 2010) and Fas was found to be frequently deleted among LMS. Overall, the biologic function of the PML pathway members involved in common deletions in LMS thus suggest that disruption of this pathway may play an important role in tumour pathogenesis (Figure 4.9).

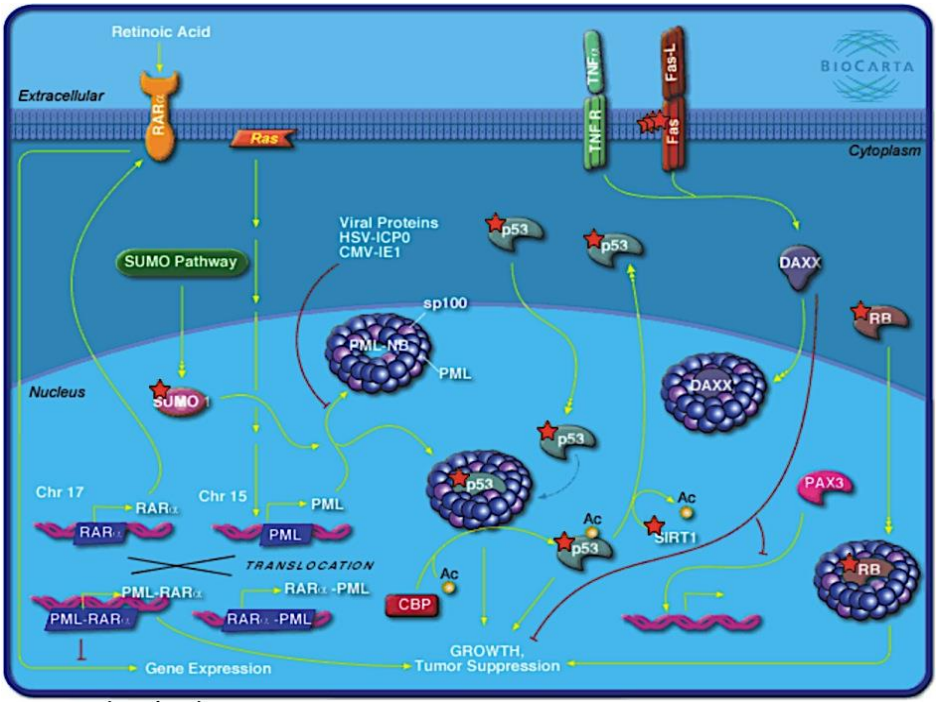

Figure 4.9: Graphic output showing enrichment of members of the Biocarta® Regulation of Transcriptional Activity by PML pathway among genes frequently deleted in LMS

All genes involved in deleted focal SCNA with a frequency ≥ 20% among LMS were submitted and the PML pathway was significantly enriched (EASE p-value <0.05). Yellow arrows represent activation while red lines represent inhibition. Pathway members coded for by genes deleted in LMS are highlighted by red stars.

Pathway and highlighted genes are graphical results automatically generated following enrichment analysis by the DAVID analysis tool (http://david.abcc.ncifcrf.gov, accessed March 2013). Pathway information is based on Biocarta® database (http://www.biocarta.com/pathfiles/h_ptenpathway.asp, accessed March 2013).

4.4.1.6 Differential Aberration Analysis

Differential aberration analysis was carried out to compare common focal SCNA in LMS to identical regions on other STS. The control group was comprised of 32 STS cases of various subtypes known to have complex genomic profiles (discussed in Section 4.2.3.3). Using a threshold of at least 40% differential frequency, four common focal amplifications and five focal deletions were found to be significant (p-value < 0.05). A complete list of genes affected by the significant differential amplifications and deletions among LMS is available in Appendix 4 (Tables D1and D2).

The significant differential amplifications involved only around 0.45Mb in total and covered loci for six genes, among which two were chosen as candidate drivers (summarised in Table 4.9). One of the genes *MYH11*, encodes a definitive marker of smooth muscle differentiation (Tamama, Sen et al. 2008, Kurpinski, Lam et al. 2010) and was chosen based on a potential role in mesenchymal lineage differentiation of these tumours. The other *ACTN1* encodes α-actinin, an actin filament that has been found to be over-expressed in various cancers with functional implications (Quick and Skalli 2010). It was also identified by enrichment analysis as a candidate amplified driver gene based on its function in the KEGG® focal adhesion pathway (Section 4.5.1.5 and Table 4.8). Neither had previously been implicated in STS.

Table 4.9: Differential Amplifications in LMS

	Region	Cytoband	Count	Region Length (Mb)	Freq. in LMS (%)	Freq. in Others (%)	p-value	q-bound	Genes in Atlas
Chr 1	203,369,354-203,514,881	q32.1	2	0.15	52.4	9.4	0.0010	0.21	-
	183,311,689-183,353,653	q25.3	1	0.04	57.1	15.6	0.0025	0.25	-
Chr 14	69,333,317-69,519,867	q24.1	2	0.19	57.1	9.4	0.0003	0.18	*ACTN1*
Chr 16	15,891,532-15,957,512	p13.11	1	0.07	52.4	6.3	0.0002	0.15	***MYH11****
			6						2

Genomic regions shown have a differential frequency of amplification > 40% in LMS (compared with average of other STS subtypes with complex karyotypes) and p < 0.05 as judged by the Fisher's exact test. Only affected genes that have been implicated in cancer are shown in the last column. Genes with reported gain of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red.
** - gene possibly involved in differentiation*

Differential aberration analysis was carried out using Nexus® Copy Number software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

The five regions with significant differential copy number deletions in LMS involved around 0.77Mb in total and covered the loci for seven genes. Only two of the seven genes have been implicated in cancer (Table 4.10). One of these was the *PRDM16* gene, which was involved in a focal deletion at 1p36.32 that was identified as significant using both the STAC and GISTIC algorithms and whose deletion would prevent differentiation into an adipocytic lineage in favour of smooth muscle (Section 4.4.1.4).

The other gene *BTRC* encodes a member of the F-box proteins called βTrCP1 that plays positive and negative regulatory roles in the NF-κB and Wnt signalling pathways, respectively (Koch, Waha et al. 2005, Müerköster, Arlt et al. 2005). It has been found to be over-expressed in breast (Bhatia, Herter et al. 2002), pancreatic (Müerköster, Arlt et al. 2005), colorectal (Ougolkov, Zhang et al. 2004) and hepatocellular carcinoma (Koch, Waha et al. 2005) . Its reported functions therefore did not support a driver role in LMS.

Table 4.10: Differential Deletions in LMS

	Region	Cytoband	Count	Region Length (Mb)	Freq. in LMS (%)	Freq. in Others (%)	p-value	q-bound	Genes In Atlas
Chr 1	1,151,667-1,161,955	p36.33	1	0.01	57.1	3.1	0.0000	0.08	-
	3,184,391-3,448,830	p36.32	3	0.26	47.6	3.1	0.0002	0.11	*PRDM16*‡
Chr 10	103,015,529-103,301,856	q24.32	1	0.29	61.9	21.9	0.0045	0.19	*BTRC*
Chr 11	78,557,943-78,714,348	q14.1	1	0.16	47.6	3.1	0.0002	0.11	-
	92,589,653-92,644,197	q14.3	1	0.05	57.1	15.6	0.0025	0.17	-
			7						*2*

Genomic regions shown have a differential frequency of deletion > 40% in LMS (compared with average of other STS subtypes with complex karyotype) and p < 0.05 as judged by the Fisher's exact test. Only affected genes that have been implicated in cancer are shown in the last column. One gene with reported gain of copy number and/or function abnormalities in cancer (likely passenger) is shown in grey.
‡ - gene with possible role in mesenchymal cell lineage differentiation.
Differential aberration analysis was carried out using Nexus® Copy Number software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

4.4.1.7 Summary of Candidate Genes in LMS

Using the different statistical tools, a total of 95 candidate genes were identified in LMS and summarised in Table 4.11. There was a predominance of amplified genes identified as candidates (68 out of 95) and most of these were involved in enriched pathways. Two candidate genes including *MYH11* and *PRDM16* (amplified and deleted, respectively) have reported roles in mesenchymal lineage differentiation, while one deleted gene, *CEP55* is involved in genome maintenance.

Six amplified and five deleted candidate genes were identified using more than one analysis tool and comprise the strongest of the candidate genes (summarised on Table 4.12). They include important putative oncogenes such as *PIK3R1* and tumour suppressor genes such as *TP53*, *RB1* and *FAS.*

Table 4.11: Summary of Candidate genes identified in LMS

Analysis Tool	Amplified candidate Genes			Deleted Candidate Genes	
Common Aberration		*ARHGEF2* *CKS1B* *ELL* *GDF15* *GMPS* *MAP2K4* *MLF1* *MYOCD* *NOTCH2* *NQ01* *PDE4DIP* *PIK3R1* *SHC1* *(n = 13)*		*ARL11* *BMPR1A* *BOK* *CEP55** *CLDN7* *DFFB* *DLG4* *FAS* *ING5* *INPP5D* *INTS6* *KCTD11* *KLLN* *(n = 25)*	*LGI1* *MUC2* *PHF11* *PRDM16* ‡ *PTEN* *RB1* *RBP4* *RCBTB1* *RCBTB2* *TNFRSF14* *TNFSF12* *TNKS2* *TP53*
Pathway Enrichment	*ACTN1* *ARHGAP5* *ARHGEF2* *BCAR1* *BCL2* *BCL2L1* *CAV2* *CBLB* *CCNE1* *CKS1B* *COL3A1* *COL5A2* *COL6A1* *COL6A2* *DAXX* *DIAPH1* *E2F2* *FOS* *FYN* *GRB2*	*IKBKB* *ITGA11* *ITGA5* *ITGA9* *ITGB5* *JUN* *LAMA3* *LAMA4* *LAMC1* *LAMC2* *MAP2K3* *MAP3K4* *MAP4K5* *MAPK1* *MAPK6* *MAPK7* *MAPKAP* *MYC* *MYLK* *(n = 58)*	*MYLPF* *PIK3CA* *PIK3R1* *PIP5K1B* *PTGS2* *RAP1A* *ROCK1* *RPS6KA* *SHC1* *SHC3* *SHC4* *SMAD3* *SOS2* *SRC* *SRF* *TGFBR2* *TNR* *TRAF4* *VAV3*	*FAS* *RB1* *SIRT1* *SUMO1* *TP53* *(n = 5)*	
Differential aberration		*ACTN1* *MYH11* ‡ *(n = 2)*		*PRDM16* ‡ *(n = 1)*	
Total Individual Genes		**69**		**27**	

Genes shown in red have reported gain of copy number and/or function abnormalities in cancer, while those shown in green have reported loss of copy number and/or function abnormalities. Genes that were identified using more than one analysis tool are highlighted in grey.

‡ *- gene with possible role in mesenchymal cell lineage differentiation.*
* *- gene with potential role in maintenance of genome integrity*

Table 4.12: Strong Candidate genes identified in LMS

	Gene	Putative Biologic Function	Frequency (%)	Likely Pathogenetic Role	Previously Identified in STS?
Amplified	*ARHGEF2*	Guanine nucleotide exchange factor for Rho GTPase – cell migration/invasion	62	Tumour Progression	Yes
	PIK3R1	Positive regulation of PIK3/Akt signalling pathway	43	Oncogenesis Tumour progression	No
	ACTN1	Cell motility	57	Tumour Progression Mesenchymal differentiation	No
	SHC1	Ras regulation, growth factor signalling and survival	62	Oncogenesis Tumour progression	No
Deleted	*PRDM16*	Negative regulation of myogenic differentiation	48	Mesenchymal differentiation	No
	FAS	Induction of apoptosis	52	Oncogenesis Tumour progression	No
	RB1	Tumour suppressor	67	Oncogenesis Cell Cycle Regulation	Yes
	TP53	Cell cycle regulation	52	Oncogenesis	Yes

All genes shown were identified as significantly aberrant among LMS by two or more of the statistical approaches (Common Aberration, Pathway Enrichment and Differential Aberration) used in this study.

4.4.2 Copy Number Aberrations in Gastrointestinal Stromal Tumours

4.4.2.1 GIST Cases

Eleven GIST cases were analysed by array CGH, including one case (GIST 11) that was re-classified as a GIST based on its array CGH profile as discussed in Section 4.5. Characteristics of all the GIST cases are summarised in Table 4.13 below. With one exception (GIST 07), the patients were in the 6th - 8th decade of life at the time of surgery with fairly even distribution (5:6) of male and female gender, respectively.

The majority of samples (10 out of 11) were primary tumours that originated in the stomach, which is the commonest site for GIST (Fletcher, Bridge et al. 2013). Only one case (GIST 08) was sampled from a liver metastasis of a gastric primary tumour. The eleven tumours ranged in size from 27 to 100mm in maximum dimension, with a fairly even distribution of tumour grade (Table 4.13). FFPE tumour material was used for array CGH in all cases with half of the samples being five years or more in age at the time of analysis in 2012.

Table 4.13: Summary of GIST cases used for common focal SCNA Identification

Case	Sampling Date	Gender	Age (years) at diagnosis	Site	Size	Grade
GIST 01	2009	Female	75	Stomach	62	High
GIST 02	2010	Male	70	Small bowel	100	Intermediate
GIST 03	2011	Female	71	Small bowel	35	High
GIST 04	2010	Female	59	Stomach	39	Intermediate
GIST 05	2004	Female	76	Stomach	70	Low
GIST 06	2004	Male	61	Stomach	27	Low
GIST 07	2005	Male	27	Small Bowel	95	Intermediate
GIST 08	2006	Male	58	Liver Met	67	High
GIST 09	2010	Male	74	Stomach	50	Intermediate
GIST 10	2010	Female	73	Stomach	54	Low
GIST 11	1994	Male	79	Stomach	30	-

4.4.2.2 Recurrent Gains and Losses

In general, GISTs had simple genomic copy number profiles when compared with LMS. Profiles were characterised mostly by large whole- or near whole-arm chromosomal aberrations. A majority of large aberrations detected were losses while the focal SCNA were mostly amplifications. The common gains and losses detected among GIST are summarised as a penetrance plot in Figure 4.10.

Loss of the long arm of chromosome 14 was the most common aberration, seen in all but two cases (82%). Other frequent losses were seen on 22q (55%), 1p (45%) and 15q (36%). These findings were in concordance with reports from previous studies (Barretina, Taylor et al. 2010, Ylipaa, Hunt et al. 2011).

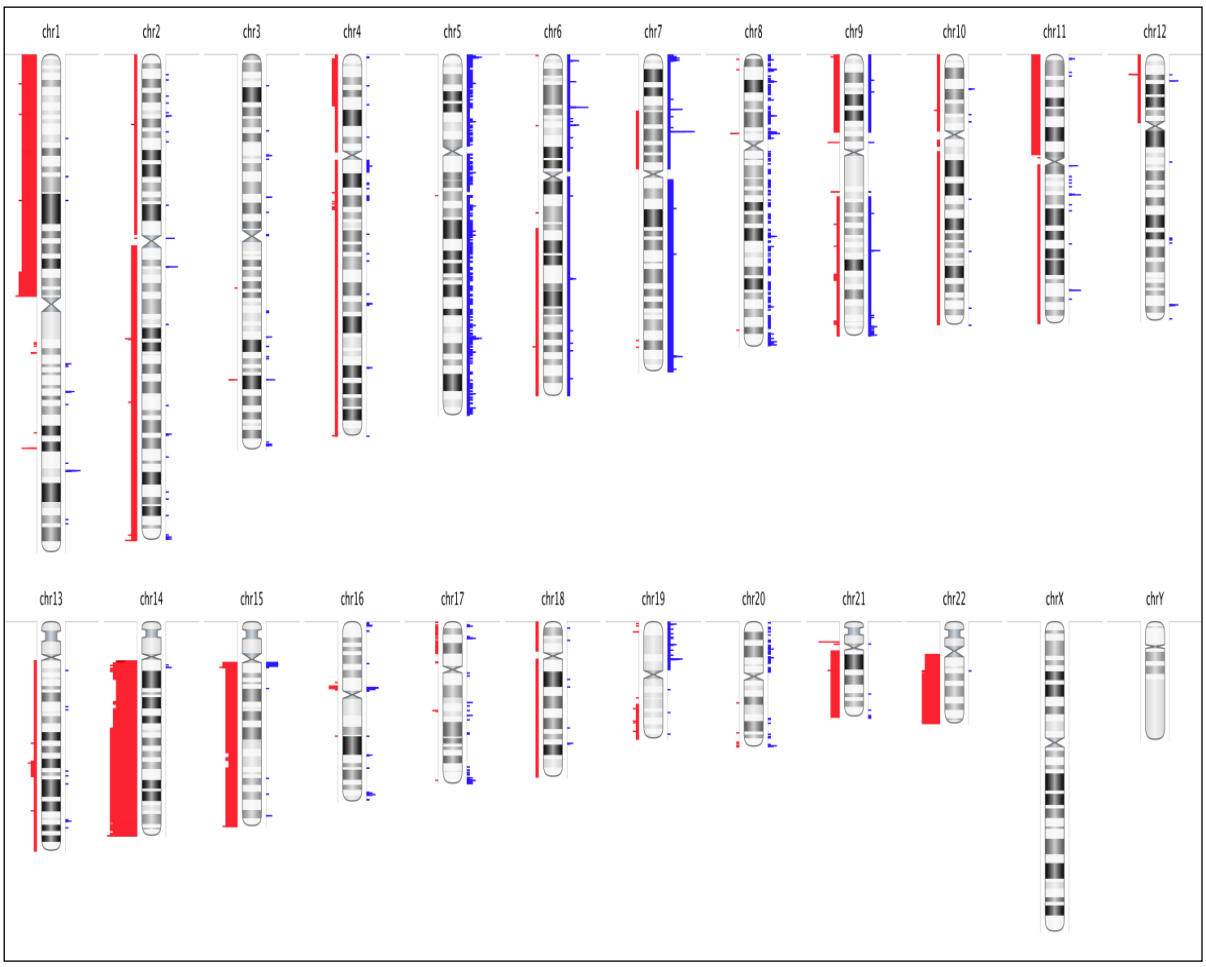

Figure 4.10: Frequency Plot of Common Genomic Copy Number Aberrations among 11 Gastrointestinal Stromal Tumours

Commonly aberrant regions are plotted as a function of their chromosomal position. Red bars to the left of the chromosome represent frequency of deletions and blue bars to the right of the chromosome represent amplifications. The heights of the bars correspond to the frequency of aberrations among the cases.
All SCNAs are detected using the FASST2 algorithm

4.4.2.3 Common Focal SCNA

Using the minimal common region (MCR) method, only 93 common aberrations affecting approximately 7% of the genome were identified with a prevalence of at least 20% among GIST. Eight of these which were in non-coding regions of the genome and a further 11 SCNAs (9 amplifications and 2 deletions) representing CNVs were excluded from further analysis.

Of the remaining 74 focal SCNA, 47 were amplifications that ranged in size from 0.032 to 0.318Mb (mean = 0.282Mb, median = 0.237Mb) and involved loci for around 7 genes on average. The most frequent amplification (seen in 64% of cases) was a 0.13Mb region located at 6p22.2 and involving a cluster of histone genes. Focal deletions were fewer (27 out of the 74 aberrations), but larger in size overall (mean = 5.2Mb, median 0.615Mb). Three of these deletions were very large (over 5Mb in size, so technically losses), including a near whole-arm loss on chromosome 14 that was 50Mb in size and covered as many as 510 gene loci.

4.4.2.4 Significance Testing of Common Focal SCNA

STAC

About a third of the 47 common focal amplifications in GIST were determined to be significant ($p < 0.05$) using the STAC algorithm (summarised in Table 4.14). However, none of the common focal deletions was found to be statistically significant. This was not surprising because the STAC algorithm tests the significance of SCNA using the average copy number state of the whole chromosome arm as a background and unlike the amplifications, many of the common deletions in GIST either occurred within or were large common aberrations that involved near whole chromosome arms. The significant SCNA identified by STAC are summarised in Table 4.9 and comprehensive list of affected genes is provided in Tables A1 and A2 of Appendix 5.

The significant amplifications identified by STAC covered loci for 75 genes, among which eight have previously been implicated in cancer (Table 4.14). For 3 of these genes however, missense mutation, deletion and/or reduced expression are the previously-reported abnormalities, making their amplification unlikely to be driver abnormalities. Four of the remaining five genes have consistently been reported to have gain of function abnormalities in relation to cancer including *JUND*, which encodes a member of the well-known family of oncogenes. They were thus readily included among candidate drivers for GIST.

The last gene *FHL2* was initially described as down-regulated among rhabdomyosarcoma cells compared to normal myoblasts (Genini, Schwalbe et al. 1997). It has however subsequently been described as frequently up-regulated in a primary tissue, *in vitro* cell cultures and even tumour xenografts of a number of other cancers, including ovarian (Gabriel, Mildenberger et al. 2004), gastrointestinal (Wang, Yang et al. 2007), breast (Martin, Kleiber et al. 2007) and prostate cancers (Heemers, Regan et al. 2007) as well as high grade gliomas (Li, Wang et al. 2008). Further, FHL2 up-regulation has been associated with adverse outcomes in breast and prostate cancer (Gabriel, Fischer et al. 2006, Kahl, Gullotti et al. 2006). These latter findings are more in keeping with the putative gene functions, which include roles in tumour cell survival, adhesion and motility (Kleiber, Strebhardt et al. 2007).

Table 4.14: Significant Common Focal Amplifications among GIST (STAC)

	Cytoband	Region	Length (Mb)	Frequency (%)	P-Value	Genes	Genes in Atlas
Chr 1	q24.2	168,378,978-168,685,455	0.31	27	0.004	3	-
	q32.2	207,984,254-208,137,993	0.15	45	0.000	1	-
Chr 2	q12.1 - q12.2	105,949,598-106,147,840	0.20	36	0.000	1	*FHL2*
Chr 5	p15.33	1,035,059-1,353,444	0.32	45	0.000	6	-
	q31.3	141,696,250-142,040,249	0.34	45	0.000	2	-
Chr 6	p25.2	2,836,816-3,001,306	0.16	36	0.020	6	-
	p22.2	26,124,423-26,250,955	0.13	64	0.000	16	-
	p21.31	34,958,314-35,188,392	0.23	36	0.020	3	-
Chr 7	q36.1	150,913,717-150,954,753	0.04	45	0.000	3	-
Chr 9	q22.32	97,847,752-98,088,612	0.24	36	0.007	1	*FANCC*
Chr 11	q23.3	117,706,762-117,781,447	0.07	36	0.000	3	-
	q13.3	69,906,662-70,052,101	0.15	36	0.000	2	-
Chr 12	p13.1	12,863,830-12,896,612	0.03	27	0.000	2	*CDKN1B*
	q24.31	124,988,479-125,313,180	0.32	27	0.001	2	-
Chr 17	p13.1	8,017,959-8,127,962	0.11	27	0.002	6	*PER1*
Chr 19	p13.11	18,380,369-18,583,639	0.20	45	0.000	9	*ELL* *GDF15* *JUND*
Chr 20	q13.33	61,818,375-62,138,166	0.32	27	0.010	9	*EEFIA2*
						75	8

Minimum frequency of SCNA = 20%, $p < 0.05$. Only affected genes found to be implicated in cancer are shown. Genes with gain of copy number and/or function abnormalities in cancer are in red and those reported in sarcoma are highlighted in bold. Any genes that show the opposite abnormalities in cancer are shown in grey.

GISTIC

In contrast to the STAC algorithm, only focal deletions were identified as significant among the GIST cases. Four deletions were identified in total, all of which were large genomic regions (losses) ranging in size from 29 to 121Mb (mean = 62.5Mb, median 49.6Mb) and represented most of 1p, 14q, 15q and 22q (summarised in Table 4.15). The extended regions involved over 2000 gene loci in total, while the peak regions contained only five genes.

Given the large number of affected genes in the extended regions, only those genes in the peak regions were examined individually in terms of biologic function. The remaining genes in the extended regions were examined by enrichment analysis (Section 4.4.2.5) and a full list of genes is presented in Appendix 5 (Tables B). None of the five peak region genes has any known abnormality that has been associated with cancer, and three of their loci corresponded exactly with a known CNV on chromosome 22 that was confirmed in other STS subgroups in this study (data not shown).

Table 4.15: Significant Common Focal Deletions among GIST (GISTIC)

	Cytoband	Region	Length (Mb)	G-Score	Q-Bound	No. of Genes	Genes in Peak	Genes In Atlas
Chr 1	p36.33 - p11.2	chr1:0-121,322,377	121.3	4.91	0.004	1098	-	-
Chr 14	q21.1 - q32.33	43,345,398-107,349,540	64.0	7.04	0.003	433	-	-
Chr 15	q24.1 - 32.33	72,998,432-102,531,392	29.5	3.15	0.046	225	*TM2D3* *TARSL2*	-
Chr 22	q11.1-13.33	16,133,474-51,304,566	35.2	5.56	0.004	459	*GSTTP1* *GSTT1* *GSTTP2*	-
						2215	*5*	

Minimum G-score = 1.0, q < 0.05. None of the genes in the peak regions was found to be implicated in cancer. The remaining genes in the affected regions were functionally examined using enrichment analysis (see below).

4.4.2.5 Enrichment Analysis of Affected Genes

Amplified Genes

A total of 350 genes whose loci were within the common amplified SCNA regions in GIST were submitted to DAVID for enrichment analysis. The only molecular pathway that was significantly over-represented among these genes was the KEGG Systemic Lupus Erythematosus (SLE) pathway (EASE p-value = 0.003). The pathway is not cancer-related and all 7 affected genes belong to a cluster of histone genes located at 6p22 (described in section 4.4.2.3).

Deleted Genes

Enrichment analysis of the 1557 genes affected by deleted regions showed significant over-representation of genes in 6 molecular pathways, which are summarised on Table 4.16 below. The involved genes are shown on Table C of Appendix 5. None of the pathways or genes involved have any biologic relevance to tumour pathogenesis in GIST. Enrichment analysis of the 2215 genes affected by significant deletions identified using the GISTIC algorithm (Section 4.4.2.4) identified the exact same pathways.

Table 4.16: Molecular Pathways significantly over-represented among commonly deleted focal SCNA among GIST

Pathway ID	Term	Gene Count	Fisher p-value	EASE p-value	Fold Enrichment
BIOCARTA h_cpsfPathway	Polyadenylation of mRNA	3	2.3E-02	2.33E-02	11.34
BIOCARTA h_circadianPathway	Circadian Rhythms	4	2.4E-04	4.69E-03	10.08
KEGG hsa04740	Olfactory transduction	92	5.3E-28	2.16E-27	3.35
KEGG hsa03050	Proteasome	10	1.6E-03	5.73E-03	2.93
KEGG hsa00600	Sphingolipid metabolism	8	6.0E-03	2.01E-02	2.83

All genes involved in amplified focal SCNA with a frequency ≥ 20% among GIST were submitted for enrichment analysis. Only cancer-related pathways with EASE p-value < 0.05 are shown in order of their fold enrichment. None of the affected genes are shown because none of them have been described with loss of copy number and/or function abnormalities in cancer.
Enrichment analysis was carried out using DAVID Bioinformatics tool (http://david.abcc.ncifcrf.gov, accessed March 2013) based on information from Biocarta® (http://www.biocarta.com/genes/index.asp, accessed March 2013) and KEGG® (http://www.genome.jp/kegg/pathway.html, accessed March 2013) databases.

4.4.2.6 Differential Aberration Analysis

Differential aberration analysis of the common SCNA among GIST compared with those in all other STS subtypes in this study showed that whole arm deletions of 1p, 14q and 22q, were not only the most frequent aberrations, but also the most significant (Fisher's p-value < 0.05). As the significant differential aberrations were mostly large losses and not focal SCNA, they are presented graphically in Figure 4.11. One small region of significant aberration highlighted at 13q14 was deleted with a significantly higher frequency among other STS subtypes (particularly LMS and UPS) and not in GIST.

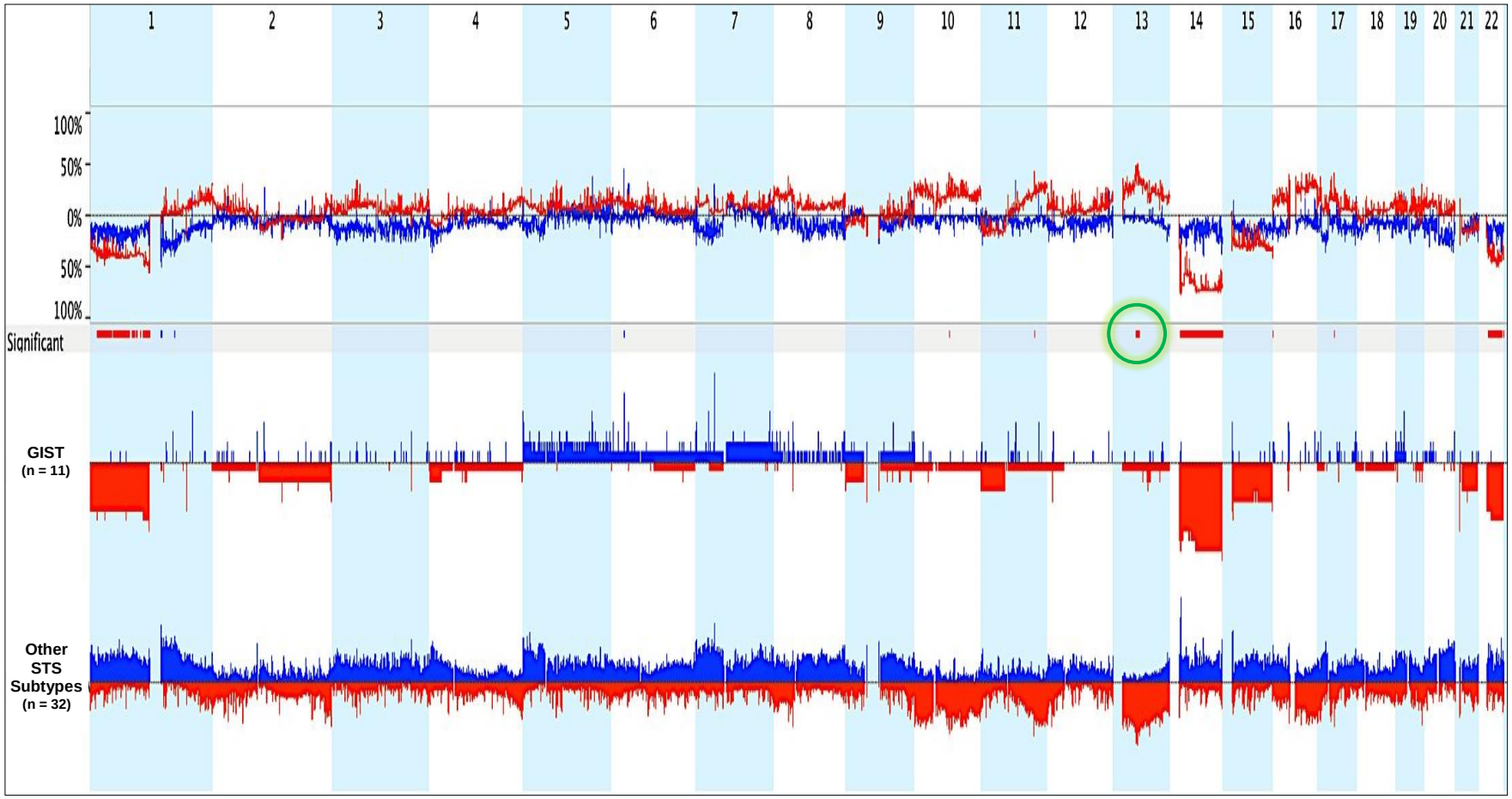

Figure 4.11: Frequency Plots of Common Genomic Copy Number Aberrations among Gastrointestinal Stromal Tumours compared with other STS Subtypes

Significant differentially aberrant regions (highlighted by the red (deletion) and blue (amplification) bars in the 'significant' track) were calculated based on an absolute difference in frequency ≥ 40% and Fisher p-value < 0.05.
In the bottom panel, commonly aberrant regions in GIST (above) and other STS subtypes (below) are plotted as a function of their chromosomal position. Red downward bars represent frequency of deletions and blue upward bars represent amplifications. The heights of the bars correspond to the frequency of aberrations among the cases.
Unlike the other significant differential aberrations shown, the differentially deleted region on chromosome 13 (circled in green) is commonly deleted among other STS subtypes (mainly LMS and UPS) but not among GIST.

All SCNAs are detected using the FASST2 algorithm

4.4.2.7 Summary of Candidate Aberrations in GIST

The candidate genomic aberrations identified among GIST are summarised in Table 4.17 below. Only five final individual candidate gene aberrations were identified by statistical analysis of common focal SCNA among GIST, none of which was a deletion. Three near-whole chromosome arm losses were however identified by both common and differential aberration analyses. All three losses are described in literature as being common in GIST and this pattern of losses led to the reclassification of one of the STS cases in this study (discussed in detail in Section 4.5)

Table 4.17: Summary of Candidates identified in GIST

Analysis Tool	Amplified Candidates	Deleted Candidates
Common Aberration	*FHL2* *ELL* *GDF15* *JUND* *EEFIA2* *(n = 5)*	1p 14q 22q *(n = 3)*
Pathway Enrichment	-	-
Differential aberration	-	1p 14q 22q *(n = 3)*
Total	**5**	**3**

Genes shown in red have reported gain of copy number and/or function abnormalities in cancer, while genomic regions shown in green have reported loss of copy number abnormalities. Aberrations that were identified using more than one analysis tool are highlighted in grey.

4.4.3 Copy Number Aberrations in Undifferentiated Pleomorphic Sarcoma

4.4.3.1 Undifferentiated Pleomorphic Sarcoma Cases

Sixteen (16) UPS cases were included for common aberration analysis and their clinical characteristics are summarised in Table 4.18. Among these, four cases (UPS 04, 05, 06 and 07) were analysed as fresh frozen samples while the others were FFPE samples. The tumours ranged in size between 55 and 270mm in maximum dimension, with the largest located in the abdominal cavity, although the majority were from the lower limb. All but two tumours were high grade (Trojani Grade 3), as expected in this STS subgroup (Fletcher, Bridge et al. 2013). There was an almost equal mix of patient gender and ages ranged between 34 and 82 years.

Table 4.18: Summary of LMS Cases used for Identification of Common Focal SCNA

Case	Gender	Age	Site	Size (mm)	Grade	Status
UPS 01	Male	65	Lower Limb	120	3	Dead
UPS 02	Female	54	Lower Limb	230	3	Alive
UPS 03	Male	44	Trunk	108	3	Dead
UPS 04	Female	36	Lower Limb	72	2	Alive
UPS 05	Male	76	Lower Limb	175	3	Dead
UPS 06	Female	67	Lower Limb	115	3	Dead
UPS 07	Female	74	Upper Limb	90	3	Dead
UPS 08	Female	75	Lower Limb	102	3	Dead
UPS 09	Female	34	Intraabdominal	270	3	Dead
UPS 10	Male	82	Trunk	93	3	Alive
UPS 11	Male	64	Intraabdominal	100	3	Alive
UPS 12	Male	65	Lower Limb	113	3	Alive
UPS 13	Male	82	Trunk	87	3	Dead
UPS 14	Male	66	Lower Limb	48	3	Alive
UPS 15	Female	67	Lower Limb	72	2	Alive
UPS 16	Male	80	Lower Limb	55	3	Alive

4.4.3.2 *Recurrent Gains and Losses*

All but one of the 16 UPS cases showed highly complex genomic copy number profiles with multiple SCNA present on almost every chromosome. A frequency plot showing gains and losses detected in at least 20% of cases is shown below in Figure 4.12. There was a fairly even distribution of frequent gains and losses.

The most frequent aberration was the near whole-arm loss of 13q, seen in around 80% of cases. Other frequent aberrations (seen in > 40% of cases) included near-whole losses of both arms of chromosome 10 and losses at telomeric ends of 1q, 4q, and 11q. Frequent gains involved whole or near-whole arms of 1p, 3p, 5p, 7p, 9q, 12p, 20q, as well as centromeric ends of 1q and 3q.

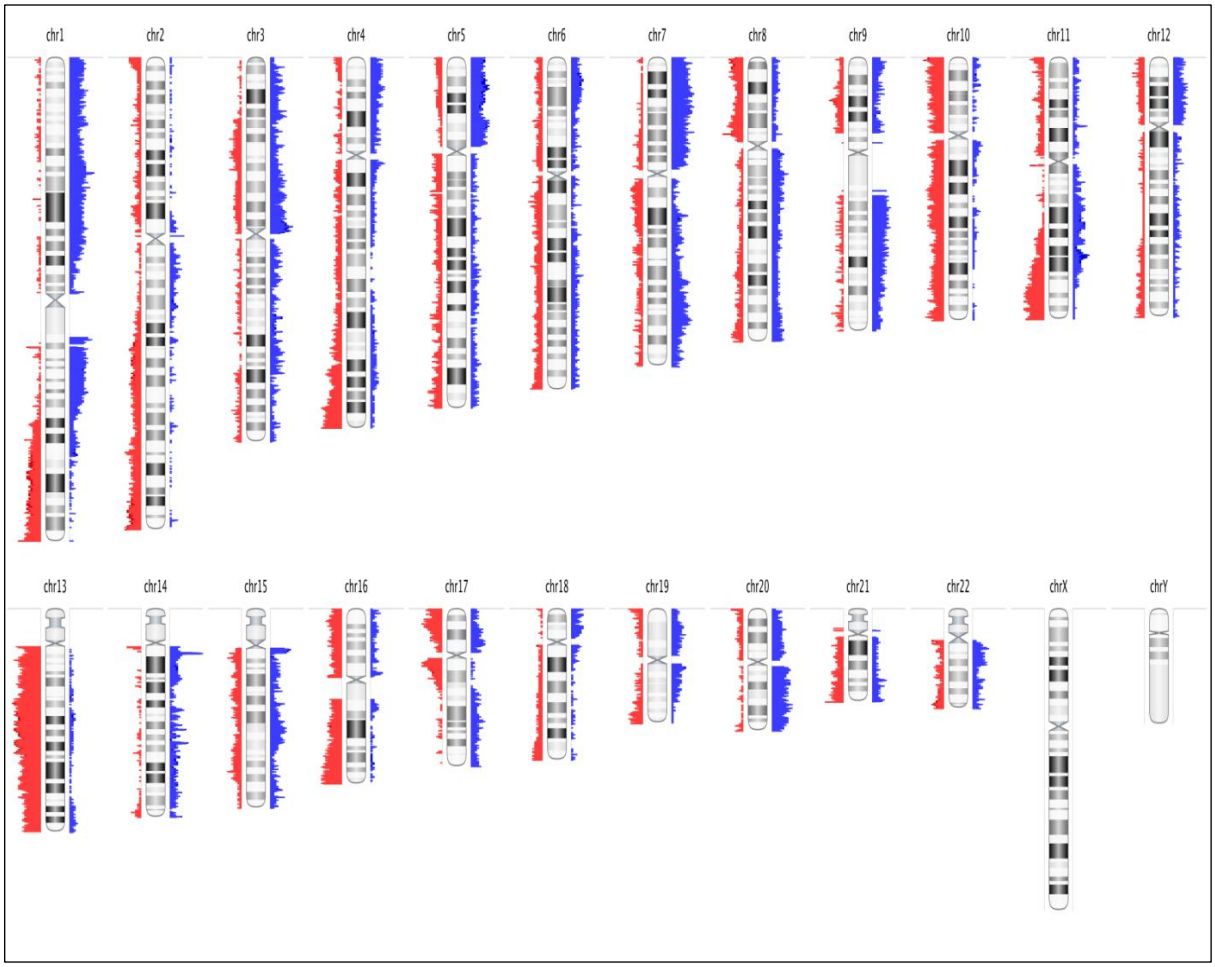

Figure 4.12: Frequency Plot of Common Genomic Copy Number Aberrations among 16 Undifferentiated Pleomorphic Sarcomas

Commonly aberrant regions are plotted as a function of their chromosomal position. Red bars to the left of the chromosome represent frequency of deletions and blue bars to the right of the chromosome represent amplifications. The heights of the bars correspond to the frequency of aberrations among the cases.
All SCNAs are detected using the FASST2 algorithm

4.4.3.3 Common Focal SCNA

Reflecting the complexity of their copy number profiles, over 2000 common focal SCNA spread over all autosomes and involving about one-third of the genome were detected with a minimum frequency of 20% among UPS. Approximately 200 of these SCNA were excluded from further analysis because they lay in non-coding genomic regions and/or represented germ-line CNVs, including a small (0.2Mb) region of copy number amplification at 14q11.2 that was detected in all sixteen UPS samples. The remaining focal SCNA comprised 956 amplifications and 888 deletions with a size range of SCNA of between 4.0kB and 4.3Mb (mean = 0.36Mb, median = 0.24Mb).

Around 7000 genes in total were affected by common focal SCNA in UPS (3815 amplified and 3359 deleted). Both the common focal deletions and amplifications contained around 4 gene loci on average. The most common aberration was a deletion at 13q14.2 detected in 14 out of the 16 cases (87.5%) that contains the locus for a single gene, *SUCLA2*. The region lies in very close proximity to the locus for the *RB1* tumour suppressor gene, which was involved in a focal deletion seen in 13 out of 16 cases (81.25%). The most common amplification involved a 0.5Mb region on 1p22 that was seen in 75% of cases (12 out of 16) and contained the locus for the well-known *JUN* oncogene.

4.4.3.4 Significance Testing of Common Focal SCNA

STAC

The STAC algorithm found nearly twice as many common focal deletions as amplifications in UPS to be statistically significant ($p < 0.05$). There were 10 significant amplifications involving loci for 14 genes as summarised in Table A1 of Appendix 6. Half of the involved genes have previously been implicated in cancer, including STS (for 3 genes) and all have reported abnormalities that would support a driver role for their amplification in UPS (Table 4.19). One of the genes, *RAD51B* may have implications for genomic instability because reported evidence in leukaemia suggests that its enhanced activity via the BCR/ABL kinase pathway is associated with defective homologous recombination and disease progression (Slupianek, Jozwiakowski et al. 2009).

Nineteen significant common focal deletions in UPS involved loci for over 130 genes spread over eight chromosomes. All the deleted SCNA identified using the STAC algorithms are summarised on Table 4.20 and the full list of affected genes is presented in Appendix 6 (Table A2). Only five of the affected genes have been implicated in cancer and all have been described as having reduced copy number or expression in relation to various cancers, although increased expression of *FAT1* has also been associated with migration and invasion of breast cancer (Kwaepila, Burns et al. 2006).

Table 4.19: Significant Common Focal Amplifications among UPS (STAC)

	Cytoband	Region	Length (Mb)	Frequency (%)	P-Value	No. of Genes	Genes in Atlas
Chr 1	p32.1	59,228,562-59,722,944	0.49	75	0.000	2	***JUN***
	q21.1	145,070,378-145,278,200	0.21	69	0.000	3	*PDE4DIP*
Chr 2	q12.3	108,867,623-108,889,968	0.02	38	0.024	1	-
Chr 3	p12.1	86,880,017-87,076,959	0.20	69	0.012	1	***VGLL3***
Chr 4	q12	54,400,737-54,592,199	0.19	44	0.013	1	***PDGFRA***
Chr 7	q33	135,033,284-135,102,970	0.07	63	0.022	1	-
Chr 8	q12.1	59,724,608-59,965,524	0.24	50	0.042	1	-
Chr 11	p13	35,080,110-35,282,073	0.20	44	0.000	2	*CD44*
Chr 14	q24.1	68,824,970-69,144,273	0.32	56	0.027	1	*RAD51B*
Chr 21	q22.12	36,617,496-36,764,360	0.15	44	0.001	1	***RUNX1***
						14	

Genomic regions shown have amplification frequency ≥ 20% and p < 0.05 as judged by the STAC algorithm. Only regions that involve at least one known gene locus are listed and only the affected genes that have been implicated in cancer are shown. Genes with reported gain of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red while those that have been described with the opposite abnormalities in cancer are shown in grey (likely passengers). Genes that have been implicated in STS are highlighted in bold red text.
STAC analysis was carried out using Nexus® Copy Number Software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

Table 4.20: Significant Common Focal Deletions among UPS (STAC)

	Cytoband	Region	Length (Mb)	Frequency (%)	P-Value	No. of Genes	Genes in Atlas
Chr 4	q35.1	184,425,375-185,143,402	0.72	56	0.032	5	***ING2***
	q35.1 - q35.2	186,797,948-187,792,442	1.00	56	0.032	9	*FAT1* ***SORBS2***
	q35.2	188,398,868-189,739,168	1.30	56	0.032	3	-
	q35.2	190,624,209-191,154,276	0.50	63	0.001	37	***DUX4***
Chr 11	q24.1	121,300,330-121,397,853	0.10	63	0.018	1	-
	q24.1	122,766,398-123,471,184	0.70	63	0.018	4	-
	q24.2	125,355,927-126,358,069	1.00	63	0.018	22	***FEZ1***
	q24.3	129,023,372-130,446,982	1.40	63	0.018	10	-
Chr 14	q32.33	104,566,967-104,715,087	0.10	31	0.042	2	-
Chr 17	q11.2	26,726,585-27,042,613	0.32	50	0.006	13	-
	q11.2	27,203,586-27,211,735	0.01	56	0.000	1	
	q11.2	27,623,194-27,880,882	0.26	63	0.000	1	-
	q11.2	28,424,381-28,525,378	0.10	56	0.000	3	-
	q11.2	29,041,847-29,371,754	0.33	63	0.000	7	-
Chr 19	p13.3	0-350,965	0.40	44	0.016	7	--
Chr 20	p13	0-210,288	0.20	38	0.043	5	-
	p13	398,763-434,066	0.04	38	0.043	2	--
Chr 21	q22.3	47,760,929-47,790,354	0.03	56	0.000	1	-
						135	5

Genomic regions shown have deletion frequency ≥ 20% and p < 0.05 as judged by the STAC algorithm. Only regions that involve at least one known gene locus are listed and only the affected genes that have been implicated in cancer are shown. Genes with reported loss of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red. Genes that have been implicated in STS are highlighted in bold red text.

‡ - gene with role in mesenchymal lineage differentiation

STAC analysis was carried out using Nexus® Copy Number Software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

GISTIC

Of the nearly 2000 SCNA detected among UPS, the GISTIC algorithm found only seven (four amplifications and 3 deletions) to be significant with G-score > 1.0 and q-bounds < 0.05. They ranged between 1.7 and 16.6Mb in size, covering loci for between 6 and 57 genes. A total of 194 genes were affected with a fairly even ratio of amplified to deleted genes (summarised in Tables B1 and B2 of Appendix 6).

With a mean G-score of 9.7, the significantly amplified SCNA occurred with an average frequency of 50.2% among UPS. Their extended regions covered a total of 95 genes, but peak regions involved only 2 gene loci. Fifteen of the genes affected by significant deletions have been implicated in cancer, among which six genes were described as being amplified and/or overexpressed in STS of various subtypes (Table 4.21). Two of these genes, *YAP1* and *VGLL3* were also identified using the STAC approach (Section 4.6.3.1). One of the significant amplicons located at 5p15.2 has been highlighted in a number of previous studies in sarcoma and three of the genes involved, *TRIO, CTNND2* and *DNAH5* have been suggested as candidate drivers (Ylipaa, Hunt et al. 2011).

Table 4.21: Significant Common Focal Amplifications among UPS (GISTIC)

	Cytoband	Region	Length (Mb)	Frequency (%)	G-Score	Q-Bound	No. of Genes	Genes in Peak	Genes in Atlas
Chr 3	3p12.1-11.1	86,816,036-88,468,010	1.7	63	8.19	0.03	6	*CHMP2B*	***VGLL3***
Chr 5	5p15	10,606,546-16,879,792	6.3	56	7.96	0.03	14	-	***TRIO*** ***CTNND2*** ***DNAH5***
Chr 7	7q11	77,437,091-94,074,985	16.6	38	8.06	0.03	57	*SEMA3E*	*AKAP9* *CDK6* ***HGF*** *DMTF1* *ABCB1* *STEAP1* *STEAP2* *COL1A2*
Chr 11	11q22.1 - q22.3	101,263,543-103,276,617	2.0	50	14.43	0.00	18	-	***YAP1*** *BIRC2* *BIRC3*
							95		**15**

Genomic regions shown have amplification G-score ≥ 1.0 and q < 0.05 as judged by the GISTIC algorithm. Only affected genes that have been implicated in cancer are shown in the last column. Genes with reported gain of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red while those that have been described with the opposite abnormalities in cancer are shown in grey (likely passengers). Cancer-implicated genes with as yet undetermined or inconsistent functional status are shown in black. Genes that have been implicated in STS are highlighted in bold red text.
GISTIC analysis was carried out using Nexus® Copy Number software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

The most frequent deleted SCNA at 13q14 was highly significant (q-bound = 0.00001) and its peak region specifically contained the locus for the *RB1* gene. The three significant deleted regions identified by GISTIC covered a total of 99 genes, among which only four have previously been implicated in cancer (Table 4.22). Apart from *RB1*, the deleted regions involved the locus of one other well-known tumour suppressor gene, *NF1* that is located at 17q11.2 and has previously been identified as frequently deleted in pleomorphic sarcomas (Taylor, Barretina et al. 2008). The other two cancer-related genes involved, *OPCML* and *INTS6* have also been described either deleted or inactivated in various cancers.

Table 4.22: Significant Common Focal Deletions among UPS (GISTIC)

	Cytoband	Region	Length (Mb)	Frequency (%)	G-Score	Q-Bound	No. of Genes	Genes in Peak	Genes in Atlas
Chr 11	11q24.3 - q25	128,837,921-135,006,516	6.2	56	6.13	0.03	23	-	*OPCML*
Chr 13	13q14	47,927,728-52,108,681	4.2	81	10.80	0.00	31	*RB1* *LPAR6*	*INTS6* ***RB1*** *
Chr 17	17q11.2	27,203,586-29,571,405	2.4	63	6.91	0.01	34	*TAOK1*	***NF1*** *TAOK1*
							99		**2**

Genomic regions shown have deletion G-score ≥ 1.0 and q < 0.05 as judged by the GISTIC algorithm. Only affected genes that have been implicated in cancer are shown in the last column. Genes with reported loss of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red. Genes that have been implicated in STS are highlighted in bold red text.
** - gene with role in maintenance of genome integrity*
GISTIC analysis was carried out using Nexus® Copy Number software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

4.4.3.5 Enrichment Analysis of Affected Genes

Amplified Genes

Over 3000 amplified genes identified by common aberration analysis in UPS were submitted for analysis. Sixteen pathways in total were significantly enriched (EASE p-value < 0.05) as shown in Appendix 6 (Table C1). Among these, only 6 pathways were cancer-related (summarised in Table 4.23). One of these, the Mitogen-Activated Protein Kinase (MAPK) signalling pathway was enriched based on data from both KEGG® and Biocarta® databases. The more detailed schematic of the MAPK signalling pathway from the KEGG® database is shown in Figure 4.13. Interestingly this pathway was also enriched among LMS, as was the functionally related focal adhesion pathway (shown in Figure 4.14).

Table 4.23: Cancer-related Molecular Pathways significantly over-represented in commonly amplified focal SCNA among UPS

Pathway ID	Term	Gene Count	Fisher's p-value	EASE p-Value	Genes Involved				Fold Enrichment
KEGG hsa04510	Focal adhesion	52	0.0003	0.0005	JUN COL11A1 RAC3 COL6A1 PTENP1 COL5A1 COL1A2 CTNNB1 COL6A2 PIK3CG TLN2 VEGFB CRK	ITGA9 VAV2 PAK6 MYL12B COL3A1 PAK1 LAMB4 ITGA1 PRKCA HGF COL6A3 LAMA5 MYL12A RELN	LAMA1 COL4A1 LAMB1 ITGB5 MET COL4A2 FYN TNC LAMA3 ARHGAP5 BAD ITGA11 RAPGEF1	COL5A2 CAV3 SHC3 PDGFB EGFR PARVA RAC2 BRAF TNR PAK4 ITGA2 LAMA4	1.59
KEGG hsa04010	MAPK signalling pathway	62	0.0016	0.0026	NTRK2 TGFBR2 NTF3 TRAF2 CACNG1 IL1R1 FGF7 ACVR1B MAP3K7 FGF10 JUN GNA12 RAC3 NFATC4 MECOM	MOS MAP3K2 NFATC2 MAP3K14 GADD45A CRK MAP2K4 CACNA1D MAP2K7 TGFB1 PAK1 PRKCA CACNG4 MAPK7 MAP2K6 RASGRP1	FGF20 CACNA1C MAP3K11 IL1A NR4A1 MKNK1 MAX CACNG5 CACNA1E PLA2G4E TAB2 MAP2K3 JUND CACNA2D3 MAP4K2 ARRB1 PDGFB	PPP3R2 CACNA1I EGFR TGFBR1 MAP3K6 RELA GNG12 DUSP7 PLA2G2C RAC2 BRAF TGFB3 STK3 ECSIT	1.42
KEGG hsa05200	Pathways in cancer	72	0.0035	0.0052	PTCH1 TGFBR2 APC2 TRAF2 RALB FGF7 ACVR1B FGF10 JUN RAC3 WNT16 GLI3 PTENP1 CTNNB1 MECOM CDKN2A SHH DVL1	CDKN2B KIT RXRA TGFB1 WNT5A LAMB4 PRKCA HGF SMAD3 RET CDK6 PIK3CG VEGFB RALBP1 CRK EGLN3 RUNX1T1 CEBPA CDKN1B	BAD MAX ARNT2 MMP1 LAMA5 LAMA1 COL4A1 LAMB1 FGF20 NFKBIA TFG MET MITF COL4A2 EGLN2 LAMA3 PDGFB RARB	EGFR BCL2L1 TGFBR1 RELA AXIN2 DAPK1 RAC2 BRAF TPR TGFB3 RALGDS WNT2 RXRG IL6 ITGA2 MLH1 LAMA4	1.35
KEGG hsa05222	Small cell lung cancer	23	0.0066	0.0132	TRAF2 PTENP1 PIK3CG CDKN2B FHIT RXRA	LAMB4 CDK6 LAMA5 LAMA1 COL4A1 LAMB1	NFKBIA COL4A2 LAMA3 CDKN1B MAX RARB	BCL2L1 RELA RXRG ITGA2 LAMA4	1.68
KEGG hsa05212	Pancreatic cancer	20	9.30E-03	0.0189	TGFBR2 RALB ACVR1B RAC3 CDKN2A	PIK3CG VEGFB RALBP1 TGFB1 SMAD3	CDK6 BAD EGFR BCL2L1 TGFBR1	RELA RAC2 BRAF TGFB3 RALGDS	1.70
KEGG hsa04520	Adherens junction	20	0.0200	0.0368	TGFBR2 ACVR1B MAP3K7 RAC3 CTNNB1	SNAI2 WASL SSX2IP SMAD3 MLLT4	MET FYN INSR BAIAP2 EGFR	TGFBR1 WASF2 PTPN1 RAC2 PTPRF	1.59

All genes involved in amplified focal SCNA with a frequency ≥ 20% among UPS were submitted for enrichment analysis. Only cancer-related pathways with EASE p-value < 0.05 are shown in decreasing order of fold enrichment. All affected genes are listed with genes with gain of copy number and/or function abnormalities in cancer (candidate driver genes) shown red. Genes that have been described with deletion or loss of function in cancer are shown in grey (likely passengers).
Enrichment analysis was carried out using DAVID Bioinformatics tool (http://david.abcc.ncifcrf.gov, accessed March 2013) based on information from Biocarta® (http://www.biocarta.com/genes/index.asp, accessed March 2013) and KEGG® (http://www.genome.jp/kegg/pathway.html, accessed March 2013) databases.

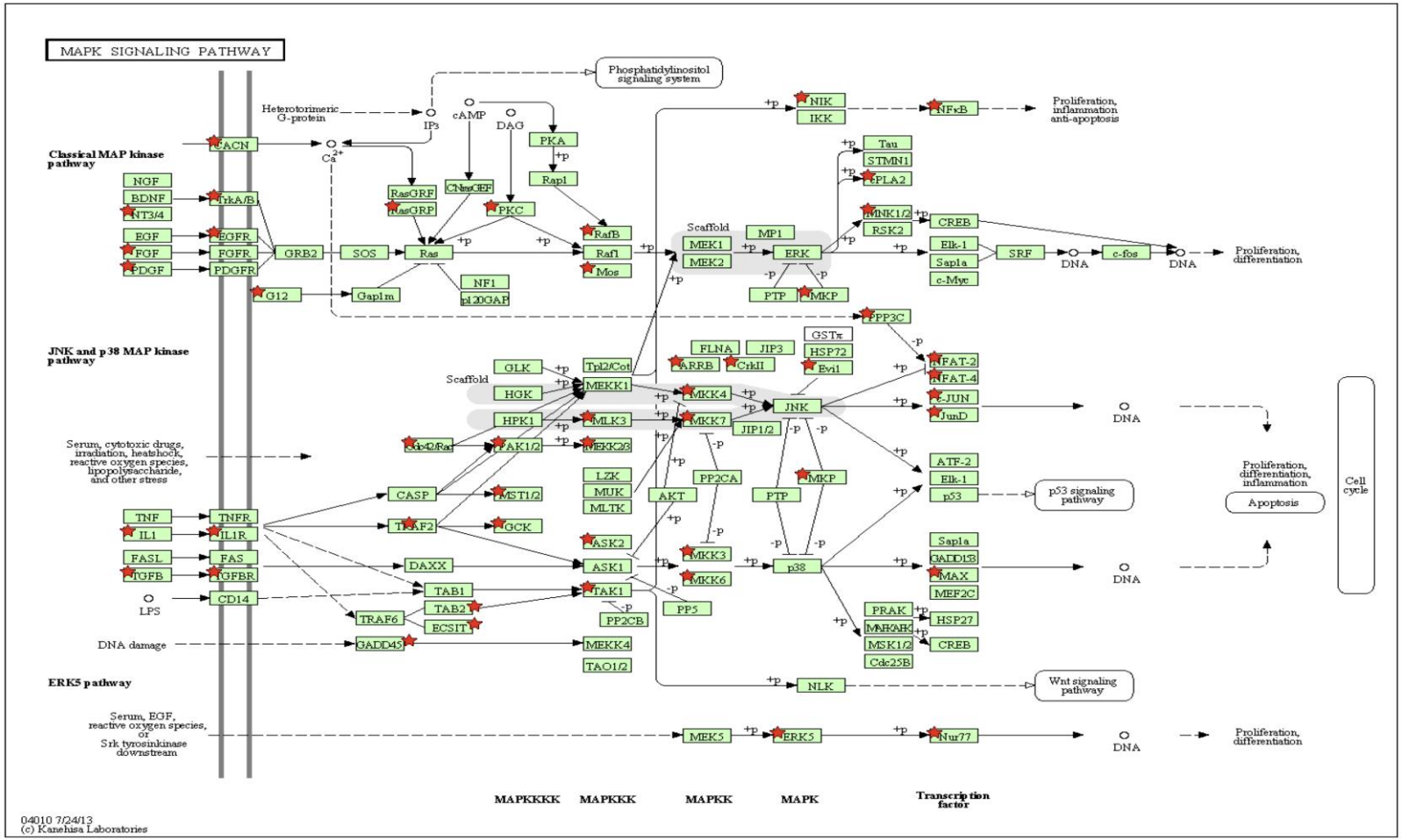

Figure 4.13: Graphic output showing enrichment of members of the KEGG® Mitogen-activated Protein Kinase (MAPK) Signalling pathway among genes frequently deleted in UPS

All genes involved in deleted focal SCNA with a frequency ≥ 20% among UPS were submitted for analysis and the MAPK Signalling pathway was significantly enriched (EASE p-value <0.05). Arrows represent activation while capped lines represent inhibition. $^{+}$p denotes phosphorylation while $^{-}$p denotes dephosphorylation. Pathway members coded for by genes that are frequently deleted in UPS are highlighted by red stars.

Pathway and highlighted genes are graphical results automatically generated (http://david.abcc.ncifcrf.gov, accessed March 2013) based on information from KEGG® database (http://www.genome.jp/kegg-bin/show_pathway?hsa04010, accessed March 2013).

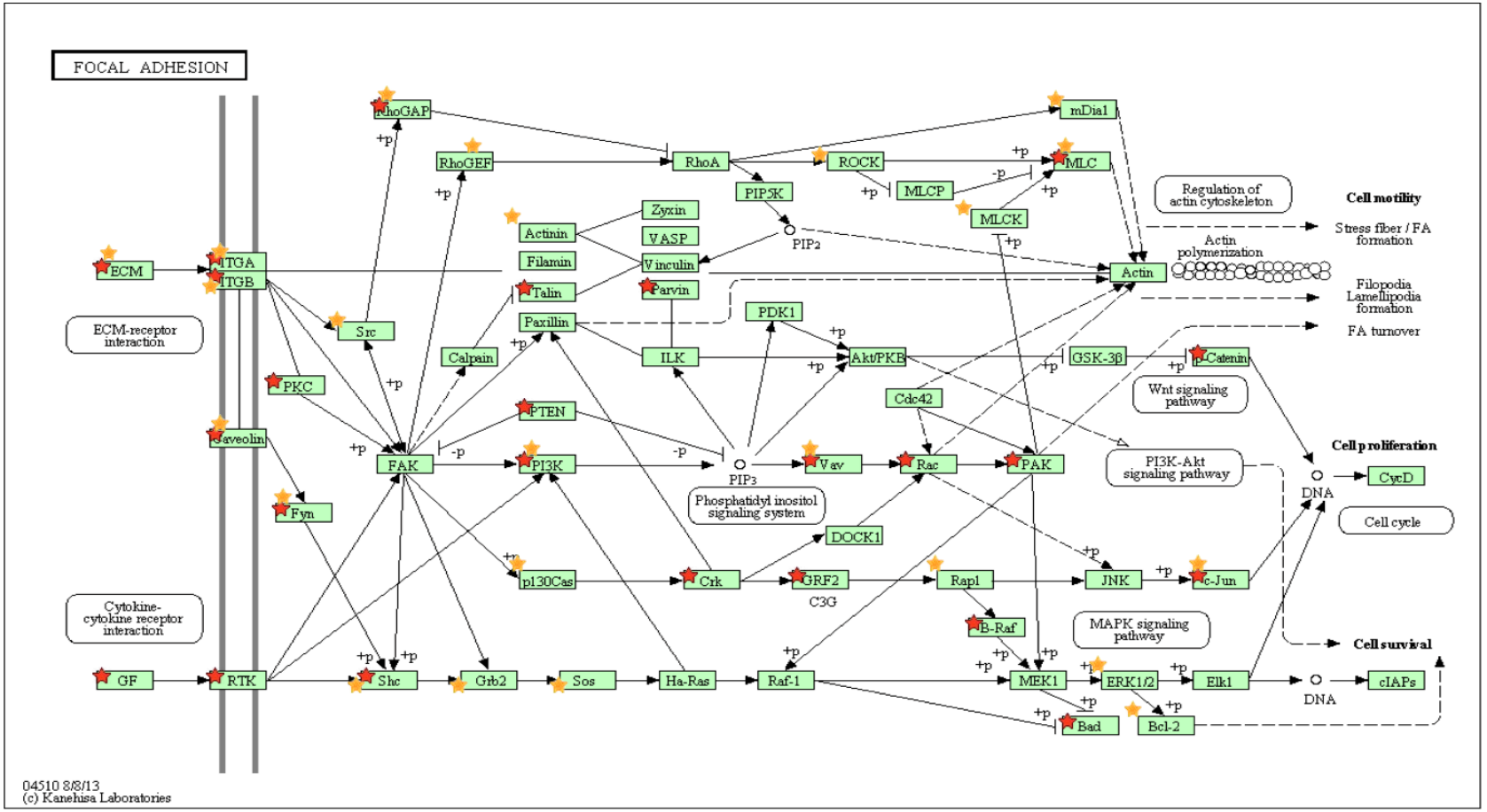

Figure 4.14: Graphic output showing enrichment of members of the KEGG® Focal Adhesion pathway among genes frequently deleted in UPS and LMS.

All genes involved in deleted focal SCNA with a frequency ≥ 20% among UPS and LMS were submitted for analysis and the Focal Adhesion pathway was significantly enriched (EASE p-value < 0.05) in both STS subtypes. Arrows represent activation while capped lines represent inhibition. $^{+}$p denotes phosphorylation while $^{-}$p denotes dephosphorylation. Pathway members coded for by genes that are frequently deleted in UPS and LMS are highlighted by red and yellow stars, respectively showing that important pathway mediators such as integrins (IGTA) and Caveolin were affected in both subtypes.

Pathway and highlighted genes are modified graphical results automatically generated following enrichment analysis by the DAVID analysis tool (http://david.abcc.ncifcrf.gov, accessed March 2013) based on information from KEGG® database (http://www.genome.jp/kegg/pathway/hsa/hsa04510.html, accessed March 2013).

Deleted genes

The commonly deleted focal SCNA in UPS involved 2740 genes, among which 11 biological pathways were significantly over-represented (EASE p-value < 0.05) as shown in Table C2 of Appendix 6. Four of the enriched pathways were cancer-related and these are summarised in Table 4.24.

In addition to the frequent homozygous deletions affecting 13q14 locus of the RB1 gene in UPS, the RB1 tumour suppressor/cell cycle checkpoint pathway was over-represented with almost all key mediators affected (Figure 4.15). This suggests that this SCNA likely plays an important role in the genetic instability as well as malignant phenotype that is evident in these tumours. The other enriched pathways include the closely-related cdc25/chek1 regulatory pathway in response to DNA damage, as well as the FAS signalling and caspase cascade pathways that are important for induction of apoptosis and/or maintenance of genome integrity.

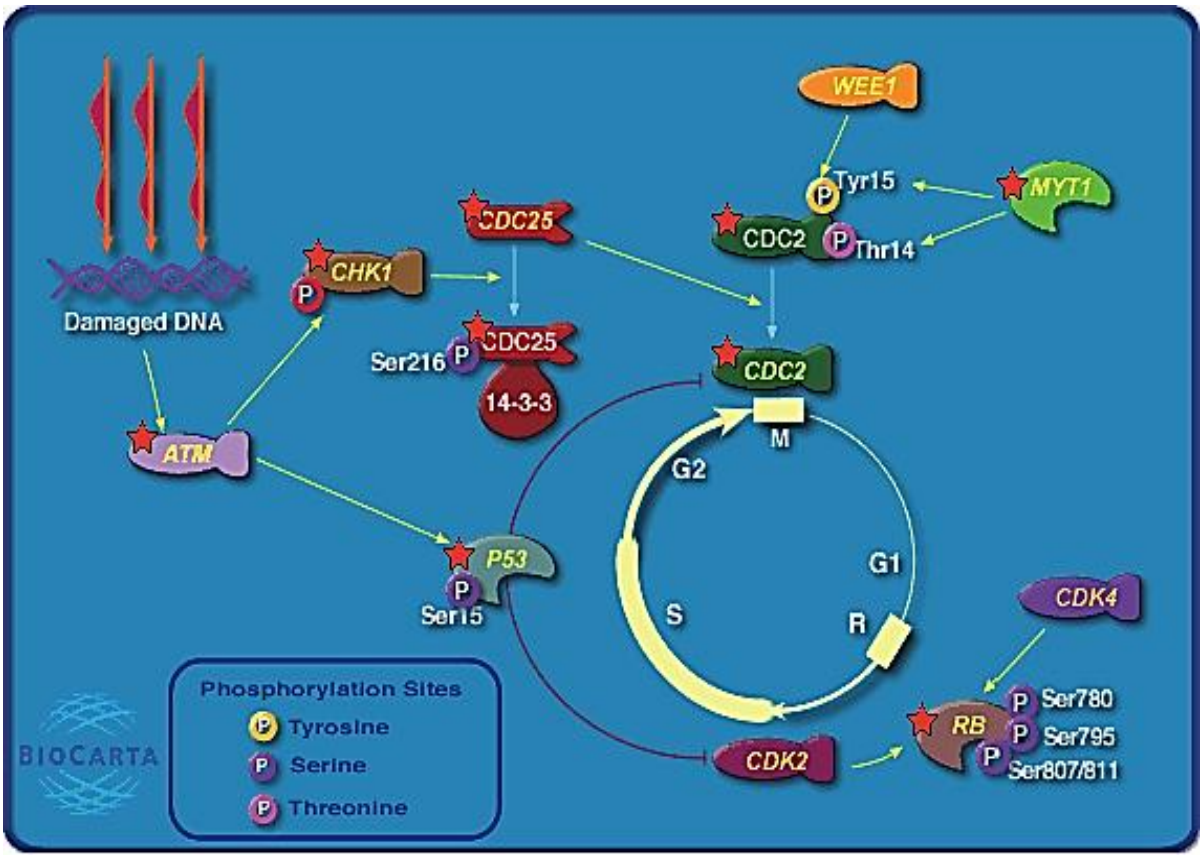

Figure 4.15: Schematic Representation of the Biocarta® RB Tumour Suppressor/Checkpoint Signalling in response to DNA damage pathway.

All genes involved in deleted focal SCNA with a frequency ≥ 20% among LMS were submitted and the PML pathway was significantly enriched (EASE p-value <0.05). Yellow arrows represent activation while red lines represent inhibition. Pathway members coded for by genes deleted in LMS are highlighted by red stars. With the exception of cdc25, loss of all other affected mediators would favour genetic instability and tumour phenotype

Pathway and highlighted genes are graphical results automatically following enrichment analysis by the DAVID analysis tool (http://david.abcc.ncifcrf.gov, accessed March 2013). Pathway information is based on Biocarta® database (http://www.biocarta.com/pathfiles/m_rbPathway.asp, accessed March 2013).

Table 4.24: Cancer-related Molecular Pathways significantly over-represented in commonly deleted focal SCNA among UPS

Category	Term	Gene Count	Fisher's p-value	EASE p-value	Genes Involved	Fold Enrichment
BIOCARTA h_rbPathway:	RB Tumour Suppressor/Checkpoint Signaling in response to DNA damage	7	0.0003	0.0024	*NPAT* *RB1* *CDK1* *MYT1* *CHEK1* *RASGRF1* *TP53*	4.34
BIOCARTA h_cdc25Pathway:	cdc25 and chk1 Regulatory Pathway in response to DNA damage	5	0.0017	0.0140	*NPAT* *CDK1* *MYT1* *CHEK1* *RASGRF1*	4.65
BIOCARTA h_caspasePathway:	Caspase Cascade in Apoptosis	8	0.0099	0.0319	*APAF1* *CASP7* *CASP3* *CASP4* *CASP8* *CASP10* *LMNB1* *PRF1*	2.48
BIOCARTA h_fasPathway:	FAS signaling pathway (CD95)	9	0.0170	0.0449	*CASP7* *CASP3* *CFLAR* *PTPN13* *CASP8* *RB1* *CASP10* *LMNB1* *FAS*	2.16

All genes involved in deleted focal SCNA with a frequency ≥ 20% among UPS were submitted for enrichment analysis. Only cancer-related pathways with EASE p-value < 0.05 are shown in decreasing order of fold enrichment. All affected genes are listed with genes with loss of copy number and/or function abnormalities in cancer (candidate driver genes) shown red. Genes that have been described with amplification or gain of function in cancer are shown in grey (likely passengers).
Enrichment analysis was carried out using DAVID Bioinformatics tool (http://david.abcc.ncifcrf.gov, accessed March 2013) based on information from Biocarta® (http://www.biocarta.com/genes/index.asp, accessed March 2013) database.

4.4.3.6 Differential Aberration Analysis

Areas of the genome with significantly different aberration frequency in UPS compared with other STS with complex genomic profiles (as described in Section 4.4.5) were identified using 40% difference in frequency and Fisher's exact test p-value < 0.05 as thresholds. Thirty-four focal copy number amplifications were found to be significant, affecting a total of 53 genes as summarised in Table D1 of Appendix 6. Nine of those genes have been implicated in various cancers as having gain of copy number and/or function abnormalities (Table 4.25). Of note are the focal amplifications at 1p32.1 and 3p12.1 affecting *JUN* and *VGLL3*, respectively that have previously been shown to be significant using one or both common aberration algorithms (Section 4.4.3.4).

Table 4.25: Differential Amplifications in UPS

	Region	Cytoband	Gene Count	Length (Mb)	Freq. in NOS (%)	Freq. in Others (%)	p-value	Genes in Atlas
Chr 1	17,236,783-17,280,446	p36.13	1	0.04	56.25	16.22	0.0064	-
	33,570,565-33,600,936	p35.1	1	0.03	56.25	10.81	0.0010	-
	34,489,528-34,777,012	p35.1 – p34.3	1	0.29	56.25	16.22	0.0064	-
	54,724,167-54,779,132	p32.3	1	0.05	50.00	8.11	0.0014	-
	59,228,562-59,247,642	p32.1	1	0.02	75.00	29.73	0.0056	***JUN***
	63,085,146-63,097,309	p31.3	1	0.01	50.00	8.11	0.0014	-
Chr 3	63,310,435-63,415,348	p14.2	1	0.10	50.00	8.11	0.0014	-
	85,152,448-86,003,801	p12.1	1	0.69	50.00	5.41	0.0005	-
	86,880,017-87,076,959	p12.1	1	0.20	68.75	13.51	0.0001	***VGLL3***
	87,727,265-88,074,845	p11.2 – p11.1	1	0.35	56.25	8.11	0.0004	-
	89,441,269-89,908,861	p11.1	1	0.47	50.00	5.41	0.0005	*EPHA3*
Chr 7	15,302,504-15,335,686	p21.2	1	0.03	62.50	18.92	0.0034	-
	15,964,268-16,510,513	p21.2 – p21.1	2	0.55	56.25	16.22	0.0064	*SOSTDC1*
	17,985,333-18,169,473	p21.1	1	0.18	62.50	18.92	0.0034	-
	18,349,312-19,091,057	p21.1	1	0.12	68.75	24.32	0.0045	-
	19,808,883-20,020,725	p21.1	1	0.21	62.50	18.92	0.0034	-
	20,390,109-20,400,245	p21.1	1	0.01	62.50	21.62	0.0098	-
	22,696,520-22,969,146	p15.3	3	0.27	62.50	18.92	0.0034	*IGF2BP3*
	23,301,525-23,362,857	p15.3	3	0.06	56.25	16.22	0.0064	*IL6*
	24,593,062-24,738,029	p15.3	2	0.14	62.50	16.22	0.0023	*DFNA5*
	25,241,387-25,274,221	p15.3	1	0.03	68.75	21.62	0.0017	-
	26,387,705-26,400,185	p15.2	1	0.01	56.25	16.22	0.0064	-
	27,792,941-27,803,481	p15.2	1	0.01	56.25	13.51	0.0023	-
	30,417,431-30,561,717	p14.3	2	0.14	56.25	16.22	0.0064	*GGCT*
	116,680,093-116,921,396	q31.2	2	0.24	50.00	8.11	0.0014	*ST7*
	135,033,284-135,102,970	q33	1	0.07	62.50	21.62	0.0098	-
Chr 9	77,295,023-77,307,482	q21.13	1	0.01	56.25	10.81	0.0010	-
	100,186,536-100,230,335	q22.33	1	0.04	50.00	8.11	0.0014	-
Chr 11	101,384,809-101,437,354	q22.1	1	0.05	43.75	0.00	0.0001	*TRPC6*
	102,169,359-102,234,053	q22.2	2	0.06	43.75	2.70	0.0005	*BIRC3* *BIRC2*
	102,393,563-102,740,197	q22.2	9	0.35	43.75	0.00	0.0001	*MMP7* *MMP20* *MMP27* *MMP8* *MMP10* *MMP1* *MMP3* *MMP12*
	102,822,550-103,076,857	q22.2 – q22.3	3	0.25	43.75	0.00	0.0001	*MMP13*
Chr 20	31,009,725-31,128,856	q11.21	1	0.12	56.25	16.22	0.0064	*ASXL1*
	31,866,250-31,890,835	q11.21	1	0.02	56.25	16.22	0.0064	-
			53					22

Genomic regions shown have a differential frequency of amplification > 40% in UPS (compared with average of other STS subtypes with complex karyotypes) and are statistically significant ($p < 0.05$) as judged by the Fisher's exact test. Only affected genes that have been implicated in cancer are shown in the last column. Genes with reported gain of copy number and/or function abnormalities in cancer (candidate driver genes) are in shown in red.
Differential aberration analysis was carried out using Nexus® Copy Number software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

Of the 44 genes affected by 27 significant differentially deleted focal SCNA ($p < 0.05$), only 4 have been implicated in cancer with loss of copy number/function abnormalities (summarised in Table 4.26 and Table D2 of Appendix 6). Focal deletions on 17p involving the *NF1* tumour suppressor gene and *TAOK1*, a gene that is believed to function as part of the spindle checkpoint (Draviam, Stegmeier et al. 2007) were previously highlighted using the GISTIC algorithm (Section 4.4.3.4).

Table 4.26: Differential Deletions in UPS

	Region	Cytoband	Count	Length (Mb)	Freq. in NOS (%)	Freq. in Others (%)	p-value	Genes
Chr 1	213,443,234-213,697,129	q32.3	1	0.25	43.75	2.70	0.0005	-
	213,980,880-214,338,265	q32.3	1	0.36	43.75	2.70	0.0005	*PROX1*
	215,302,569-215,751,141	q41	2	0.45	43.75	2.70	0.0005	-
	216,043,174-216,684,943	q41	1	0.64	43.75	2.70	0.0005	-
	216,731,644-217,224,745	q41	1	0.49	43.75	2.70	0.0005	*ESRRG*
	234,357,539-234,437,898	q42.2	1	0.08	50.00	5.41	0.0005	-
	242,489,058-242,528,211	q43	1	0.39	43.75	2.70	0.0005	-
	245,514,462-245,689,504	q44	1	0.18	50.00	5.41	0.0005	-
Chr 13	70,404,055-70,857,794	q21.33	2	0.21	75.00	32.43	0.0066	-
	19,296,544-19,550,188	q11 - q12.11	1	0.25	75.00	29.73	0.0056	-
	24,831,951-24,864,286	q12.12	1	0.03	68.75	27.03	0.0064	-
	26,218,732-26,263,107	q12.13	1	0.04	81.25	32.43	0.0021	*ATP8A2*
	37,588,232-37,598,673	q13.3	1	0.01	68.75	27.03	0.0064	-
	58,178,513-58,211,706	q21.1	1	0.03	81.25	37.84	0.0063	-
	67,442,419-67,451,208	q21.32	1	0.01	75.00	32.43	0.0066	-
	74,152,253-74,320,849	q22.1	1	0.17	81.25	32.43	0.0021	*KLF12*
	92,058,034-92,596,337	q31.3	1	0.05	68.75	24.32	0.0045	*GPC5*
	96,808,522-97,090,785	q32.1	1	0.28	68.75	24.32	0.0045	-
Chr 15	83,773,547-83,780,826	q25.2	1	0.01	43.75	2.70	0.0005	-
Chr 16	6,127,389-6,503,206	p13.3	1	0.38	50.00	8.11	0.0014	-
Chr 17	26,726,585-27,042,613	q11.2	14	0.31	50.00	5.41	0.0005	*RAB34*
	27,203,586-27,211,735	q11.2	1	0.01	56.25	8.11	0.0004	-
	27,862,668-27,880,882	q11.2	1	0.02	62.50	13.51	0.0006	***TAOK1****
	28,458,050-28,525,378	q11.2	2	0.07	56.25	13.51	0.0023	-
	29,464,657-29,476,900	q11.2	1	0.01	56.25	16.22	0.0064	***NF1***
	8,574,964-8,676,013	p13.1	2	0.1	56.25	16.22	0.0064	-
Chr 21	47,760,929-47,790,354	q22.3	1	0.03	56.25	10.81	0.0010	-
			44					8

Genomic regions shown have a differential frequency of deletion > 40% in UPS (compared with average of other STS subtypes with complex karyotype) and are statistically significant ($p < 0.05$) as judged by the Fisher's exact test. Only affected genes that have been implicated in cancer are shown in the last column. Genes with reported gain of copy number and/or function abnormalities in cancer (likely passenger) is shown in grey.
** - gene with possible role in cell spindle checkpoint.*
Differential aberration analysis was carried out using Nexus® Copy Number software and gene function information was obtained from Atlas of Genetics and Oncology and PubMed literature reviews.

4.4.3.7 Summary of Candidate Genes in UPS

Combined statistical analysis of array CGH data identified 174 candidates in UPS. Summarised in Table 4.27, they comprise 148 amplified and 26 deleted genes. As with LMS, the majority of these were highlighted by pathway enrichment. Strong candidates that were identified by two or more independent analysis methods comprised 9 amplified and 2 deleted genes, summarised in Table 4.28. They included genes located in the most frequent focal amplification and deletion seen in this subtype, *JUN* and *RB1* respectively.

Table 4.27: Summary of Candidate genes identified in UPS

Analysis Tool	Amplified candidate Genes			Deleted Candidate Genes
Common Aberration	*ABCB1* *BIRC2* *BIRC3* *CD44* *CDK6* *COL1A2* *CTNND2*	*DNAH5* *HGF* *JUN* *MMP1* *PDE4DIP* *PDGFRA* *RAD51B* *(n = 21)*	*RUNX1* *SEMA3A* *STEAP1* *STEAP2* *TRIO* *TRPC6* *VGLL3*	*DUX4* *FEZ1* *ING2* *INTS6* *NF1* *OPCML* *RB1* *SORBS2* *TAOK1* *(n = 9)*
Pathway Enrichment	*ARHGAP5* *ARNT2* *ARRB1* *BAIAP2* *BCL2L1* *BRAF* *CACNA1C* *CACNA1D* *CACNA1E* *CACNA1I* *CACNA2D3* *CACNG4* *CACNG5* *CAV3* *CDK6* *COL11A1* *COL1A2* *COL3A1* *COL5A2* *COL6A1* *COL6A2* *COL6A3* *CRK* *CTNNB1* *DUSP7* *DVL1* *ECSIT* *EGFR* *EGLN2* *FGF10*	*FGF20* *FGF20* *FGF7* *FYN* *GADD45A* *GLI3* *GNA12* *HGF* *IL1A* *IL1R1* *IL6* *INSR* *ITGA11* *ITGA9* *ITGB5* *JUN* *JUND* *KIT* *LAMA1* *LAMA3* *LAMA4* *LAMA5* *LAMB1* *LAMB4* *MAP2K3* *MAP2K4* *MAP2K6* *MAP2K7* *MAP3K11* *MAP3K14* *MAP3K2* *(n = 120)*	*MAP3K6* *MAP3K7* *MAP4K2* *MAPK7* *MAX* *MECOM* *MET* *MITF* *MKNK1* *MMP1* *MOS* *MYL12A* *MYL12B* *NFATC2* *NFATC4* *NR4A1* *NTF3* *NTRK2* *PAK1* *PAK4* *PAK6* *PARVA* *PDGFB* *PIK3CG* *PLA2G4E* *PRKCA* *PTPN1* *PTPRF* *RAC2* *RAC3*	*RALB* *RALBP1* *RALGDS* *RAPGEF1* *RASGRP1* *RELA* *RELN* *RET* *SHC3* *SHH* *SMAD3* *SNAI2* *SSX2IP* *TAB2* *TFG* *TGFB1* *TGFB3* *TGFBR1* *TNC* *TNR* *TPR* *TRAF2* *VAV2* *VEGFB* *WASF2* *WASL* *WNT16* *WNT2* *WNT5A*	*APAF1* *CASP10* *CASP3* *CASP4* *CASP7* *CASP8* *CDK1* *CHEK1* *FAS* *LMNB1* *MYT1* *PRF1* *PTPN13* *RB1* *TP53* *(n = 15)*
Differential aberration	*BIRC2* *BIRC3* *EPHA3* *GGCT* *IGF2BP3* *IL6*	*JUN* *MMP1* *MMP10* *MMP12* *MMP13* *MMP20* *(n = 17)*	*MMP27* *MMP3* *MMP7* *MMP8* *VGLL3*	*ATP8A2* *NF1* *PROX1* *TAOK1* *(n = 4)*
Total Individual Genes	**148**			**26**

Genes shown in red have reported gain of copy number and/or function abnormalities in cancer, while those shown in green have reported loss of copy number and/or function abnormalities. Genes that were identified using more than one analysis tool are highlighted in grey.

‡ *- gene with possible role in mesenchymal cell lineage differentiation.*
* *- gene with potential role in maintenance of genome integrity*

Table 4.28: Strong Candidate genes identified in UPS

	Gene	Putative Biologic Role	Frequency (%)	Likely Pathogenetic Role	Previously Identified in STS?
Amplified	*JUN*	Putative Oncogene, p53 inhibition	75	Oncogenesis Tumour progression,	Yes
	HGF	Tumour cell proliferation, survival, motility, and morphogenesis	38	Tumour progression	No
	CDK6	G1-S phase checkpoint	38	Tumour Progression	No
	COL1A2	Matrix Interaction and Integrin signalling	38	Tumour Progression	No
	IL6	Growth factor signalling	56	Tumour progression	No
	VGLL3	Transcription factor co-activator Negative regulation of adipocytic differentiation	69	Tumour progression Mesenchymal differentiation	Yes
	YAP1	Transcription Factor co-activator	50	Tumour Progression	Yes
	BIRC2	Inhibition of apoptosis	50	Tumour progression	No
	BIRC3	Inhibition of apoptosis	50	Tumour progression	No
Deleted	*RB1*	Tumour suppressor, G1-S phase checkpoint	81	Oncogenesis Tumour Progression Genetic instability	Yes
	TAOK1	Spindle cell cycle checkpoint	63	Genetic Instability	No
	NF1	Tumour suppressor	63	Oncogenesis	Yes

All genes shown were identified as significantly aberrant among UPS (n=16) by two or more of the statistical approaches (Common Aberration, Pathway Enrichment and Differential Aberration) used in this study.

4.5 RECLASSIFICATION OF A CASE BASED ON ARRAY CGH PROFILING

Prior to 1998 and the identification of the sensitive and specific KIT (CD117) staining for GIST (Sarlomo-Rikala, Kovatich et al. 1998), many cases were classified as LMS, which sometimes bore histological resemblance and also stained positive for CD34. Based on the differential aberrations identified in GIST, genomic copy number profiles of all LMS cases in this study were examined. One tumour (originally identified as LMS 06 in Chapter 3) had all three differential deletions identified among GIST, but did not possess any of the differential aberrations identified among LMS (Figure 4.16). The tumour, which was resected in 1994 had been classified as LMS based on the knowledge at the time. Subsequent immunohistochemistry of the tumour tissue for CD117 showed strong positivity (Figure 4.17). The tumour was thus reclassified as a GIST (case identified as GIST 11). All data analysis was then repeated with the correct tumour classification before being presented in this chapter.

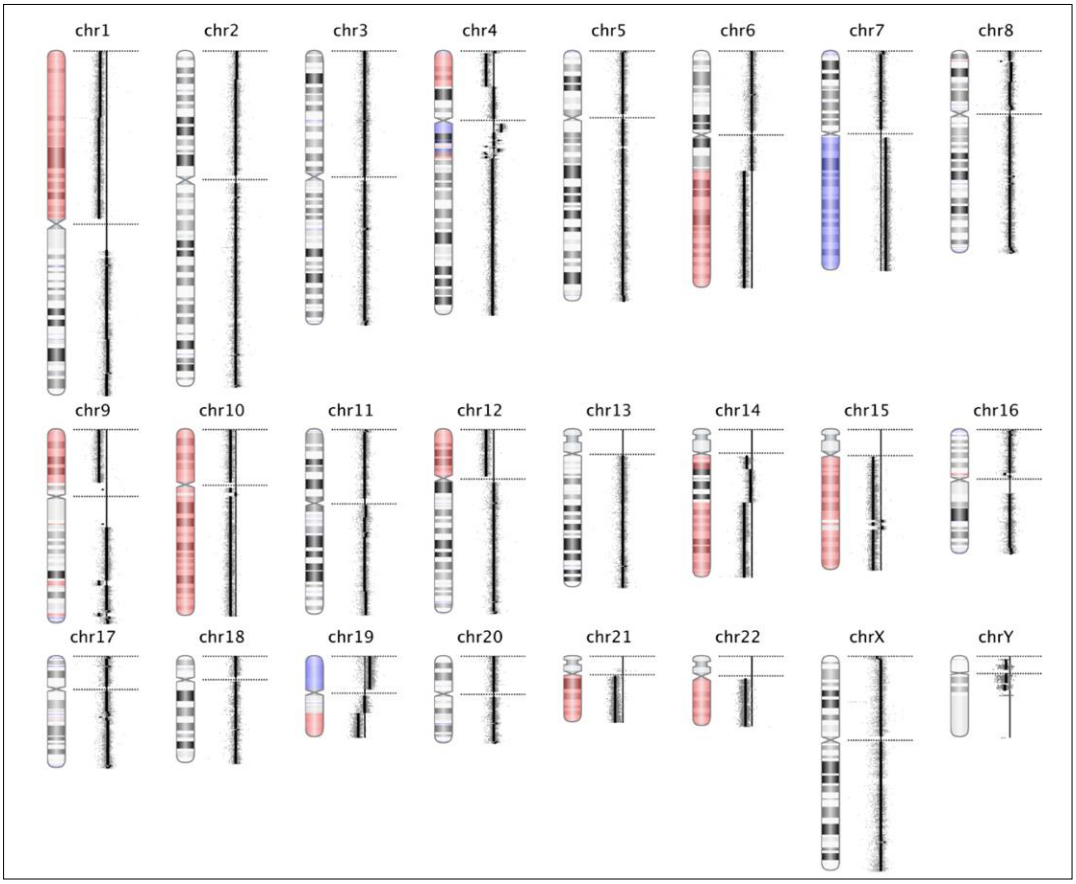

Figure 4.16: Genome View Ideogram of a case initially classified as LMS, which may be a GIST

Differential aberrations found to be significant among GIST (whole or near whole arm losses of 1p, 14 q and 22q) were all present in this tumour and it was considered for reclassification, The Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the amplitude of log ratio.
All SCNAs are detected using the FASST2 algorithm

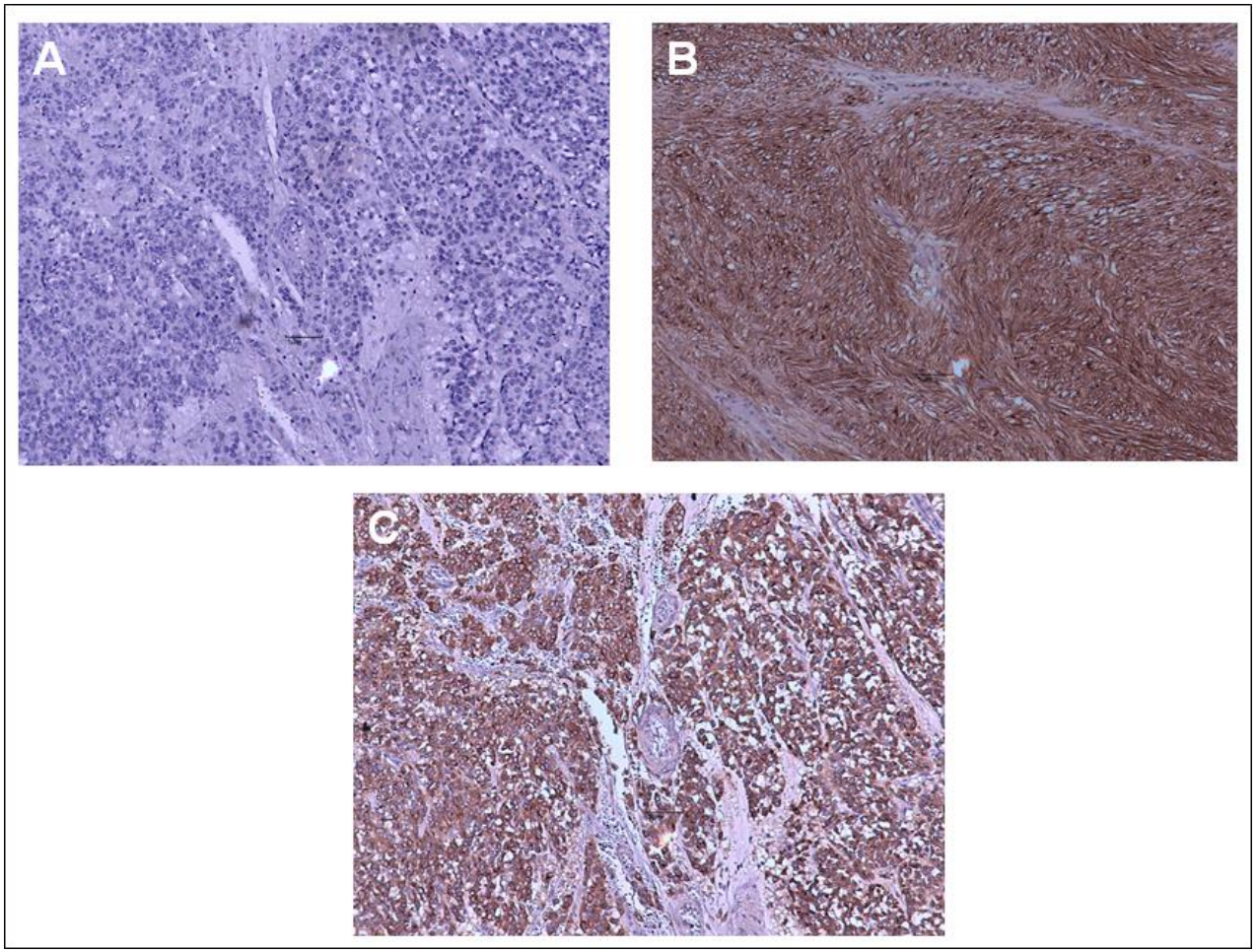

Figure 4.17: Immunohistochemical detection of KIT (CD117) in STS sections

FFPE sections were immuno-stained with antibodies against KIT (brown) and counterstained with haematoxylin (blue). Representative micrographs are shown. A: Negative control (LMS 06 stained without primary antibody). B: LMS 06 staining strongly positive for CD117 leading to its re-designation as GIST 11. C: Positive control (known GIST tumour obtained from Histopathology laboratory), stained concurrently.
Original magnifications: × 200

4.6 DISCUSSION

Since array CGH is only able to detect unbalanced karyotypic abnormalities and not balanced ones, it is of limited use in those tumour subtypes that are almost exclusively driven by balanced chromosomal translocation such as synovial sarcoma, extraskeletal myxoid chondrosarcoma, alveolar soft part sarcoma. This was confirmed by the featureless array CGH profiles seen with these STS subtypes (Section 4.3). This array CGH study aimed to identify subtype-specific 'driver' SCNA in STS and this was reflected in subtype preference of FFPE cases that were sampled, focusing on subtypes with complex karyotypes. The resulting bias meant that identification of SCNA that are common to all STS subtypes was not possible using this sample population.

The majority of LMS and UPS tumours tend to exhibit highly malignant behaviour (Fletcher, Bridge et al. 2013) and this was clearly reflected in the high Trojani grade of the majority of cases in this study. The combination of prospective and retrospective cases also meant that survival data was not available for all cases. Therefore, in addition to the relatively small sample size of STS subtypes, there were few clinical features with which copy number data could be correlated in the context of this project. Contribution of the data to pooled sets in the form of collaborations and online databases such as Oncomine (Rhodes, Yu et al. 2004) however means that beyond this PhD project, more aspects of the array CGH data will be explored sooner rather than later and that the statistical analyses will subsequently have sufficient power. In spite of these limitations, the results from this study confirm the findings of multiple larger-scale studies thus adding validity to the novel observations.

Candidates in Leiomyosarcoma

Analysis of the amplified focal SCNA among LMS resulted in 69 potential driver genes in total (Table 4.11). Among these is *MYOCD*, a transcription cofactor gene located at 17p13 that has been shown to induce smooth muscle differentiation in cells (Wang, Wang et al. 2003, van Tuyn, Knaän-Shanzer et al. 2005). It was amplified in around 43% (nine out of 21) of LMS cases in this study, most of which were intraabdominal tumours. These results concur with those Perot and colleagues, who demonstrated differential amplification and overexpression of *MYOCD* among retroperitoneal LMS (Perot, Derre et al. 2009). They also showed *in vitro* functional effects of its knockdown that have resulted in the current view that *MYOCD* is a putative driver gene for the LMS subtype (Taylor, Barretina et al. 2011). The differentially aberrant candidate gene, *MYH11* is of related interest because it encodes smooth muscle myosin heavy chain (SMMHC), a definitive marker of smooth muscle differentiation (Tamama, Sen et al. 2008, Kurpinski, Lam et al. 2010) whose expression can be induced via overexpression of *MYOCD* in embryonic or mesenchymal stem cells (Raphel, Talasila et al. 2012). It also has a possible role in oncogenesis because it is involved in a fusion protein with CBFβ in acute non-lymphoblastic leukaemia, although its exact role in this tumour is yet to be determined.

Another candidate that was amplified in 15 out of 21 LMS cases (71%) and identified using both common aberration algorithms was *PDE4DIP*, which encodes a golgi/centromere protein called myomegalin. Myomegalin is a putative activator of platelet-derived growth factor receptor-β (PDGFRB), with which it forms a fusion protein in chronic eosinophilic leukaemia (Wilkinson, Velloso et al. 2003). Increased levels of *PDE4DIP* expression were reported in oesophageal carcinoma cells following the detection of antibodies to myomegalin in sera from patients (Shimada, Kuboshima et al. 2007). Another frequently aberrant candidate gene at a nearby locus on chromosome 1 is *ARHGEF2,* which encodes a putative oncogene (Brecht, Steenvoorden et al. 2005). A guanine nucleotide exchange factor for RhoGTPase, the *ARHGEF2* gene product was found to be overexpressed in *p53*-mutant U2OS osteosarcoma cell lines (Mizuarai, Yamanaka et al. 2006) and down-regulated in GIST tumours that were treated with Imatinib, a receptor tyrosine kinase inhibitor that targets PDGFR (Frolov, Chahwan et al. 2003). Neither candidate gene has been previously reported as amplified in LMS and these SCNA may identify a subset of LMS with potential therapeutic susceptibility to Imatinib.

Deletion of the *RB1* locus at 13q14 was one of the most common aberrations detected among LMS in this study. This contrasts with findings from a previous study in the same laboratory where common amplifications detected at 13q21 - 32 among LMS led to the proposal that the nearby *RB1* locus may be co-amplified and extra copies of *RB1* were visualised using FISH (Ul-Hassan, Sisley et al. 2009). Those studies were however based on chromosome CGH, the available technology at the time which has a limited resolution of around 10Mb (Tan, Lambros et al. 2007) and the authors admit that polyploidy could not be ruled out as an explanation of the FISH observations. While this high-resolution aCGH study did show amplifications at regions slightly telomeric to 13q32 in around 20% of LMS cases (Figure 4.5), specific and frequently homozygous deletion of the *RB1* locus was far more prevalent. It was seen in two-thirds of cases analysed and had the highest GISTIC score of all the commonly aberrant regions in LMS (Table 4.7).

Enrichment analysis showed pathogenic pathways for three cancers (including chronic myeloid leukaemia, small and non-small cell lung cancer) were over-represented among the amplified genes in LMS and that all 3 have the PIK3/Akt/PTEN signalling pathway in common. A well-characterised pathway, gain of function mutations and/or amplifications in key genes along the pathway such as the oncogene *PIK3CA* have described in many cancers, including myxoid liposarcoma (Italiano, Chen et al. 2012). In addition to the strong candidate *PIK3R1*, amplifications involving *PIK3CA* were seen in more than half of the LMS cases in this series. When combined with the frequent deletions involving two important genes, *PTEN* and *TP53* (deleted in 52% each of LMS cases) that regulate this pathway via their antagonist effects (Cully, You et al. 2006), all but one case possessed SCNA whose overall effect would in theory, result in increased cell survival via modulation the PIK3/Akt/PTEN pathway. Not only are both antagonist genes known to be frequently deleted in LMS (Italiano, Chen et al. 2012), they are candidates identified in this study. Further, genes in the PTEN pathway that act downstream of its inhibitory effects were overrepresented among amplified genes in

LMS. Taken together, there is a strong overall suggestion that the PIK3/Akt/PTEN pathway may be an oncogenic mechanism for LMS potentially comparable to the *MDM2* driver amplifications currently being exploited for targeted therapy in liposarcoma (Taylor, Barretina et al. 2011). The results are in concordance with *in vivo* observations in transgenic mice by Hernando et al who concluded that dysregulation of the PIK3/Akt/PTEN pathway was vital to development of LMS (Hernando, Charytonowicz et al. 2007).

The other molecular signalling pathways with overrepresented genes in LMS, Mitogen-activated Protein Kinase (MAPK) and Epidermal growth factor receptor (EGFR) pathways have significant cross talk with the PIK3/Akt/PTEN pathway and with each other. They result in the activation of transcription proteins that drive cell proliferation, activation of matrix metalloproteinases and induction of angiogenesis (Wagner and Nebreda 2009) with an overall effect of contributing to tumour progression. Actin modification via amplification of members within the Rho signalling pathway and ECM matrix interactions could also contribute to tumour progression in LMS. Further investigation of these signalling pathways with a proteomic approach, as well as specific inhibitors could lead to identification of valuable targets within these druggable pathways that could be used for molecular therapeutic development in LMS.

Deleted SCNA yielded a final list of 26 potential driver genes, four of which were strong candidates. Focal deletion at 1p36.32 involving *PRDM16* was significant using both common and differential aberration approaches. Although the exact functional relevance in cancer is still unclear, knockdown of *PRDM16* was shown to favour differentiation of cells into a myogenic lineage instead of brown fat (Seale, Bjork et al. 2008), suggesting a possible role in the myogenic differentiation of LMS. Another candidate gene of interest, *CEP55* encodes a centrosome protein that interacts with *BRCA2* and mediates the cytokinesis step of mitosis (Fabbro, Zhou et al. 2005, Mondal, Rowley et al. 2012). Its deletion in around half of the LMS cases may contribute to the aneuploidy and chromosomal instability that is characteristic of these tumours.

The only enriched cancer-relevant pathway among deleted genes was the promyelocytic leukaemia protein (PML) pathway that participates in tumour suppression and apoptosis. It involves the interaction of several well-known genes including *RB1*, *TP53* and *FAS*. Other genes that were involved in deleted SCNAs identified as relevant by both common aberration approaches include the pro-apoptotic genes *TNFSRF14* and *BOK,* which have been implicated in follicular lymphoma (Launay, Pangault et al. 2012) and colorectal cancer (Zeilstra, Joosten et al. 2011), respectively. Taken together, these results strongly suggest that inhibition of apoptosis via targeted gene deletion may be relevant to the pathogenesis of LMS.

Candidates in GIST

The majority of STS subtypes with known characteristic genetic aberrations have simple karyotypes (Taylor, Barretina et al. 2011) with few or no SCNA. This may account for the relatively simple genomic profiles seen among the cKIT mutation-driven GISTs. Non-focal genomic losses were the predominant aberration and the most frequent of them (those on 14q, 1p and 22q) were also found to be specific for GIST. This observation led to the re-classification of one of the older LMS cases and confirmation by immunohistochemistry for cKIT (Section 4.5).

Among only five candidate genes affected by focal amplifications was the proto-oncogene *JUND*, which encodes a member of the activator protein 1 (AP-1) transcription factor complex. Overexpression of *JUND* has been shown to be involved in inhibition of apoptotic signals in prostate cancer (Zerbini, de Vasconcellos et al. 2011) and the induction of matrix metallo-proteinases in skin-invasive Adult T-cell leukaemia (Nakachi, Nakazato et al. 2011). Amplified in nearly five out of the 11 (nearly 50%) cases, this focal SCNA may therefore represent a secondary abnormality that contributes to tumour progression among GIST.

Candidates in UPS

The high degree of karyotypic complexity of UPS was evidenced by the involvement of around 33% of the genome in common focal SCNA, compared with an average of around 10% genomic involvement in focal SCNA observed in a wide range of cancers (Beroukhim, Mermel et al. 2010), and 20% in LMS. Statistical analysis identified nearly 200 potential drivers, including 12 strong candidate genes (Tables 4.27 and 4.28). Among the amplified candidate genes, not only was *JUN* the most frequently affected, it was identified by differential aberration and pathway enrichment analyses as well. This makes it statistically one of the strongest candidate genes identified in this study. A well-known putative oncogene, it encodes a key member of the AP-1 transcription factor complex that has been shown to promote cell proliferation via activation of cyclin D and inhibit apoptosis via inhibition of p53 and p21 (Shaulian 2010). Recently, Mariani and colleagues based on observations from array CGH results demonstrated a functional association between *JUN* amplification/overexpression and de-differentiated liposarcomas (Mariani, Brennetot et al. 2007). They concluded that the aberration blocked adipocytic differentiation and resulted in an aggressive phenotype. Interestingly, the *JUN* amplification observed was independent of amplifications on 6q or 12q involving *ASK* or *MDM2*, respectively as reported in de-differentiated liposarcoma (Mariani, Brennetot et al. 2007). This suggests that *JUN* amplification may play a more prominent role in early oncogenesis of pleomorphic sarcomas than previously thought, with a secondary role in preventing lineage differentiation and should be the subject of further study.

Other candidate genes are of potential relevance to lack of specific lineage differentiation that is characteristic of UPS. Overexpression of the strong candidate *VGLL3* was recently shown to inhibit differentiation of mesenchymal cells into an adipocytic lineage (Halperin, Pan et al. 2013) while *CDK6* expression is known to inhibit the differentiation of osteoblasts (Grossel and Hinds 2006). The deleted candidate gene *DUX4*, on the other hand has been shown to be involved in promoting

myogenic differentiation (Winokur, Chen et al. 2003). It also participates in the characteristic *CIC-DUX4* fusion in undifferentiated paediatric primitive sarcomas that is believed to result in loss of its function (Graham, Chilton-MacNeill et al. 2012). Nearly all the UPS cases showed aberrations affecting one or more of these three genes.

Further, amplification of *VGLL3* was highlighted as significant using common aberration as well as the differential aberration approaches making it a strong, subtype-specific candidate gene for UPS. Along with *YAP1,* another candidate gene located at 11q22.2, it has previously been reported as amplified and overexpressed in a subset of sarcomas that included UPS (Helias-Rodzewicz, Perot et al. 2010). Both genes encode a cofactor of the TEAD family of transcription cofactors and knockdown of their expression resulted in reduced *in vitro* tumour cell migration, suggestive of possible roles in tumour progression (Helias-Rodzewicz, Perot et al. 2010).

Copy number amplification of regions on 5p have been the subject of previous studies in STS and a number of candidate genes were proposed e.g. *CTNND2,* which encodes a cell adhesion protein called δ-catenin whose overexpression is functionally implicated in metastatic cancers (Taylor, Barretina et al. 2008). Another gene within the same amplicon, *TRIO* encodes a Rho Guanine-nucleotide Exchange Factor (RhoGEF) that regulates actin cytoskeleton modification via which it is involved in triggering cell motility, proliferation and differentiation. *TRIO* amplification has been reported various cancers including breast cancer (Lane, Martin et al. 2008) and glioblastoma (Salhia, Tran et al. 2008), where it was functionally associated with tumour progression (Salhia, Tran et al. 2008). Over-representation (among the amplified genes in UPS) of the focal adhesion signalling pathway (Table 4.23), which acts via RhoGTPase signalling adds validity to the selection of this gene as a candidate for tumour progression in UPS.

The 26 deleted candidate genes in UPS include some very important and well-known tumour suppressor genes such as *RB1* and *NF1*. Germ-line deletion in both these genes causes a predisposition to STS (discussed in chapter 1) and frequent somatic deletions have been described in various cancers including ovarian cancer (The Cancer Genome Atlas Research Network 2011), glioblastoma (The Cancer Genome Atlas Research Network 2008) and pleomorphic STS (Taylor, Barretina et al. 2008). Neurofibromin, the product of *NF1* is a negative regulator of the *RAS* oncogene, an upstream kinase in the overrepresented PIK3 and MAP kinase pathways seen among amplified genes in UPS. Inactivating events of this gene therefore probably facilitates increased proliferation among and evasion of apoptosis by UPS tumour cells.

The *RB1* product in addition to it tumour suppressor role, functions as a key member of a G1/S phase cell cycle checkpoint pathway, many members of which were frequently deleted in UPS (Table 4.24). The important cdc25/Chk1 cell cycle checkpoint, as well as multiple apoptosis-regulating pathways was also over-represented among deleted genes. Frequent deletions were detected affecting *TAOK1*, a kinase that has been shown to function in the G2/M phase in response to both ultraviolet and ionising radiation acting via the p38 MAPKinase pathway (Raman, Earnest et al. 2007). With these multiple abnormalities in genome maintenance machinery, it is not surprising

that genomic instability is highly characteristic of UPS tumours. In addition, *TAOK1* deletion alone or in combination with deletion of multiple pro-apoptotic caspase genes may confer resistance to radiotherapy on a subset of UPS and warrants further investigation.

Summary

Taken together, it appears that both LMS and UPS may utilise similar signalling pathways to drive cell proliferation and that ECM interaction and Rho signalling is probably relevant to tumour progression in both subtypes. However, while LMS favour inactivation of the PTEN and p53 tumour suppressor pathways, UPS tend to favour *RB1* and *NF1* inactivation. Even more exciting is the identification with very high levels of statistical significance of candidate genes that are likely to be of importance in determination of the phenotypic appearance and thus classification of specific STS subtypes e.g. *MYH11* in LMS and *VGLL3* in UPS. These genes also appear to contribute to tumour progression and may be targetable with therapeutics.

Since copy number aberrations frequently but do not always correlate with gene expression, the observations for this array CGH study require further investigation using expression and functional approaches in model systems. They however serve as a useful starting point for areas of focus in further studies, an approach that has led to elucidation of pathogenesis in various cancers.

CHAPTER FIVE

RESULTS: PRIMARY TISSUE CULTURE

5.1 INTRODUCTION

Functional testing of molecular genetic candidates using *in vitro* disease models is usually the first step in therapeutic target validation and serves as a screening tool before their evaluation in more complex *in vivo* models and subsequent clinical trials (Taylor, Barretina et al. 2011). Commercial cell lines are a widely used *in vitro* model for initial functional validation studies because they are a readily available, endlessly replicating source of tumour material from which results obtained are usually reproducible (Cree, Glaysher et al. 2010).

There has however, been increasing recognition of the limitations of commercial cell lines as a disease model stemming from reports of poor correlation of the response in these cell lines with *in vivo* tumour response (Cree, Glaysher et al. 2010, Kamb 2010, Smith, Stewart et al. 2010). With cellular adaptation to artificial culture conditions, cell lines have been shown to grow more rapidly than parent tumour cells and acquire phenotypic changes that may alter their therapeutic response, such as dependence on growth factors in culture media or adherence to plastic (Kato, Espinoza et al. 2008, Pan, Kumar et al. 2009). In addition, the heterogeneity of tumour cell clones characteristic of many cancers is lacking in cell lines in which single clones have been selected for and results may not reflect resistant tumour cell clones or cancer 'stem' cells that are believed to be responsible for tumour recurrence and late therapeutic failure.

These observations resulted in a paradigm shift that led to the use of a panel of cell lines for drug testing in each cancer that presumably represent the variability of *in vivo* therapeutic responses. Notable examples include the National Cancer Institute (NCI-60) panel that includes 60 cell lines that represent nine different cancers (Shoemaker 2006) and a similar Japanese panel of 39 cell lines representing nine cancers (Yamori 2003). Cell lines included in these panels have been used in research for nearly 30 years and proven useful for elucidation of important pathogenetic pathways as well as cytotoxic and targeted therapy development in many cancers (Gillet, Varma et al. 2013) and more recent genomic and transcriptomic studies of these cell lines support their continued relevance to research (Beroukhim, Mermel et al. 2010, Bignell, Greenman et al. 2010).

In STS however, there is a limited number of *in vitro* disease models (tumour cell lines) available for functional testing and target validation. Data from large-scale cancer cell line studies such as the Cancer Cell Line Encyclopaedia (Barretina, Caponigro et al. 2012) and Sanger Cancer Cell Line Project (Forbes, Bindal et al. 2011) showed that less than 2% of commercially available cell lines are derived from STS, the majority of which are translocation-driven. With improvements in the technology for genomic aberration mapping e.g. high-resolution array CGH, combined SNP and oligonucleotide arrays and more recently, next generation sequencing leading to an increased rate of candidate identification (Barretina, Taylor et al. 2010, Taylor, Barretina et al. 2011), there is a growing need to establish a wider range on STS cell lines (Taylor, Barretina et al. 2011).

While it is widely accepted that the best alternative to established cell lines as *in vitro* disease models is the use of cells cultured directly from tumours (primary cell cultures), this is fraught with many problems (Luca, Privitera et al. 2007). Fresh tumour tissue has to be donated by patients and obtained during surgery with the associated ethical and logistic constraints. When it is available, the

behaviour of tumour cells is generally unpredictable with a variable rate of successful establishment in culture (Luca, Privitera et al. 2007). Further, parent tumours contain many normal cells e.g. fibroblasts that may grow instead of or alongside tumour cells in culture (Brouquet, Taleb et al. 2011). These normal cells need to be distinguished from tumour cells using biomarkers if available as their presence may confound experimental results, as demonstrated in array CGH studies (Neill, Torchia et al. 2010). As a result, many researchers now believe that successful bench to bedside translation of *in vitro* results is well worth the effort of obtaining primary cell cultures for all *in vitro* work. Others use results from primary cultures to augment those from readily available commercial cell lines, instead of replacing them entirely (Cree, Glaysher et al. 2010).

Primary cell culture is also gaining relevance in the selection of cytotoxic drugs for cancer treatment. Historic practice has been the selection of cytotoxics that statistically are efficacious in cancer types based on their histopathology. For rare cancers or patient who have failed to respond to conventional cytotoxics, an emerging strategy is drug selection based on high throughput chemosensitivity assays on primary cells cultured from biopsy samples (Kamiyama, Rauenzahn et al. 2013, Mitra, Mishra et al. 2013). This personalised medicine strategy was demonstrated, to be effective in a case of rare pancreatic acinar cancer (Armstrong, Von Hoff et al. 2011) and other researchers are developing methods that increase the efficiency of primary cell culture to be used for such purposes (Hidalgo, Bruckheimer et al. 2011, Kamiyama, Rauenzahn et al. 2013).

In this study, establishment of primary cell cultures was attempted with all fresh STS tissue collected. A number of methods were utilised to examine their genetic and phenotypic characteristics of the cultured cells in relation to the parent tumours and thus their suitability for functional validation studies. Cultures were grown as long as possible and frozen down at intervals at early passages. This chapter presents the results of the attempts at primary tissue culture and describes the characteristics of long-term cell cultures established during this PhD project.

5.2 RESULTS

5.2.1 STS Cases

Fresh tissue samples were collected from 53 patients with suspected STS (based on clinical, radiological or biopsy findings) over a 30-month period from February 2010 to August 2012. Cultures were set up as described in Section 2.3.1.1 with laboratory designations based on their chronological order and the year they were obtained, with 'STS' as a prefix. For example, cultures derived from the second tumour obtained in 2010 were referred to as STS 02/10 and the sixth tumour in 2012 was called STS 06/12, etc. Detailed clinical features of all the fresh STS included in this study are summarised in Appendix 1.

Six cases subsequently received definitive diagnoses of benign or epithelial malignant tumours including one leiomyoma (STS 03/10), one melanoma (STS 08/10), one breast carcinoma (STS 07/11), two lipomas (STS 12/11 and 23/11) and a benign chondroma (STS 24/11). They were therefore excluded from the study. The remaining 47 cases had confirmed STS diagnoses comprising 16 different subtypes (summarised on Table 5.1). All but five cases were primary tumours. Three cases, all well-differentiated liposarcomas, were local tumour recurrences while two others were metastatic tumours (one pleomorphic Liposarcoma and one angiosarcoma). Four out of five primary angiosarcomas sampled occurred at sites that had previously been exposed to radiotherapy for breast cancer, while the fifth occurred secondary to chronic lower limb lymphoedema (Stewart-Treves syndrome).

In four cases (STS 07/10, 16/11, 20/11 and 05/12), neo-adjuvant radiotherapy was given locally to reduce the tumour sizes and render them operable. Two other patients received systemic preoperative chemotherapy as part of their treatment protocol for a rhabdomyosarcoma (STS 12/10) and Ewing's sarcoma (STS 13/11). As a result, all six cases that had been exposed to neoadjuvant therapy, as well as the two alveolar soft part sarcomas and the myxoid liposarcoma (nine cases in total) could not be graded histologically using the Trojani system (Coindre 2006). Of the remaining 38 cases , over 60% (25 cases) were Trojani grade 3; three cases Trojani grade 2; and 10 cases low grade, including seven well-differentiated liposarcomas that are Trojani grade 1 by default (Fletcher, Bridge et al. 2013).

5.2.2 Primary Tissue Culture

5.2.2.1 Washes

Seventy per cent of STS cases (33 out of 47) established adherent cells in culture, which reached confluence within the flasks. Cells settled either as individual cells that had been mechanically dissociated from the tumour tissue or from tissue explants that were adherent to the plastic (Figure 5.1A). In some cases, adherent cultures were eventually established in washes of viable non-adherent cells that were set up (as described in Section 2.3.1.2) in addition to the original flasks or slopes. Cell cultures from washes were designated with a 'w' prefix to the passage number. For example, cells from the second wash of an original culture and currently in their third passage were designated were designated w2p3 while cells from a wash of the original slopes at the same

passage were designated wsp3. Cells from original cultures and washes were maintained as separate variants with all relevant precautions to prevent cross-contamination.

5.2.2.2 Senescence

Proliferating cells were cultured until they became senescent (Figure 5.1B). After senescence was observed, cells were kept in culture for at least 6 weeks, during which they were media-changed regularly and observed for any change in proliferation rate or morphology before they were eventually discarded.

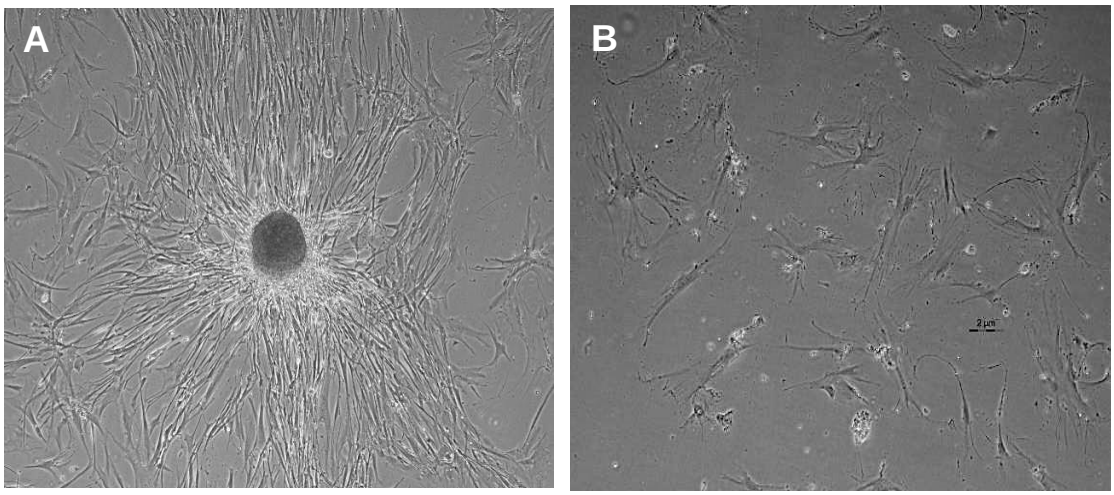

Figure 5.1: Phase-contrast Micrographs of Proliferating and Senescent cells in primary Soft Tissue Sarcoma cultures.
A – An example of early culture showing proliferating spindle-shaped cells growing radially outwards from an adherent piece of tissue (tumour explant) shown by the white arrows. **B** – a representative example of senescent cells in a primary STS culture.

Images were captured at x40 magnification.

5.2.2.3 Outcome

Table 5.1 summarises the outcome of culture in all the STS cases by subtype. Early senescence was observed within the first or second passage (p1 – p2) in ten cases, typically in the first two to three weeks of culture. Later senescence occurred in a third of cases and was generally observed at two points. The first was around the passage 5 (p4-p7) and the second, around passage 10 (p9 – p12). All cells lines that grew beyond passage 12 went on to become long term cultures, i.e. maintained proliferative cultures for one year or more. Cells were frozen every 3 - 4 passages, within the first 15 passages to provide cells with less cultural adaptation for future experiments.

Well-differentiated liposarcomas had the poorest tissue culture outcome overall with only one of the seven cases able to establish cells in culture, which subsequently underwent early senescence. Both cases where the patient had neo-adjuvant chemotherapy established cells that reached the tenth passage, approximately before they became senescent. One of the 4 tumours that received neo-adjuvant radiotherapy, (STS 20/11) formed a long term culture while the others failed to establish any cells in culture. Undifferentiated pleomorphic sarcomas were the most successful in culture overall with all cases that had not been exposed to neo-adjuvant therapy establishing cells in culture and a long term culture outcome in three out of five cases. Table 5.2 summarises the features of the tumours that established long term primary cell culture.

Table 5.1: A Summary of Fresh Soft Tissue Sarcoma Subtypes obtained and their Primary Cell Culture Outcomes.

STS Subtype	Number of Cases					
	Obtained	Not Established in Culture	Early Senescence	Senescent around p5	Senescent around p10	Long Term Culture
Alveolar Soft Part Sarcoma	2	-	-	1	1	-
Angiosarcoma	6	-	3‡	2	1	-
Dedifferentiated Liposarcoma	7	3	2	-	-	2*
Ewing's Sarcoma	1	-	-	-	1*	-
Extraskeletal Myxoid Chondrosarcoma	1	-	-	1	-	-
Leiomyosarcoma	3	1	-	-	1	1
Low grade Myofibroblastic Sarcoma	1	-	-	1	-	-
Malignant Peripheral Nerve Sheath Tumour	1	-	-	1	-	-
Malignant Solitary fibrous Tumour	1	-	-	1	-	-
Myxofibrosarcoma	5	-	3	-	1	1
Myxoid Liposarcoma	1	1	-	-	-	-
Pleomorphic Liposarcoma	2	1*	-	1‡	-	-
Pleomorphic Rhabdomyosarcoma	1	-	-	-	1*	-
Synovial Sarcoma	1	-	-	1	-	-
Undifferentiated Pleomorphic Sarcoma	7	2**	1	-	1	3
Well-differentiated Liposarcoma	7	6§	1	-	-	-
Total	**47**	**14**	**10**	**9**	**7**	**7**

** - One case received neo-adjuvant radiotherapy or chemotherapy*
*** - Two cases received neo-adjuvant radiotherapy or chemotherapy*
§ - Three cases were recurrent tumours
‡ - One case was a metastatic tumour

Table 5.2: Characteristics of Soft Tissue Sarcoma tumours that established Long-term Primary Cell Cultures

Laboratory Designation	Age/Gender	STS Subtype	Site	Size	Date Obtained	Duration in Culture (months)	Number of Passages (by April 2013)	Current Clinical Status (as of April 2013)
STS 09/10	68y/F	De-differentiated Liposarcoma	Retroperitoneal	300mm	June 2010	33	70	Dead
STS 14/10	53y/F	Undifferentiated pleomorphic sarcoma	Lower Limb	230mm	August 2010	31	70	NP
STS 02/11	62y/F	Leiomyosarcoma	Pelvis	135mm	January 2011	26	65	NP
STS 06/11	76y/M	Undifferentiated pleomorphic sarcoma	Lower Limb	170mm	February 2011	25	85	Dead
STS 09/11	66y/F	Undifferentiated pleomorphic sarcoma	Lower Limb	115mm	April 2011	23	35	Dead
STS 20/11	70y/F	De-differentiated Liposarcoma	Lower Limb	170mm	October 2011	17	24	Metastatic Disease
STS 21/11	73y/M	Myxofibrosarcoma	Upper Limb	50mm	November 2011	17	30	Metastatic Disease

NP – no disease progression

5.2.3 Characterisation of Cells in Culture

5.2.3.1 DNA Profiling

Routine precautions were taken in cell culture to avoid cross-contamination and cultures were routinely monitored by phase contrast microscopy. The appearance of cell clones with distinct morphology raised the possibility of cross-contamination. One instance of cross-contamination was suspected in cultures from the first wash of STS 04/11. Prior to the tenth passage, the cultures consisted of spindle-shaped fibroblast-like cells with many senescent-looking cells (Figure 5.2A). A morphologically-distinct clone of polygonal, histiocyte-like cells was observed in one of the flasks at passage w1p10 (Figure 5.2B). By the thirteenth passage (w1p14), the new clone of cells had completely replaced the original cells the culture flask (Figure 5.2C).

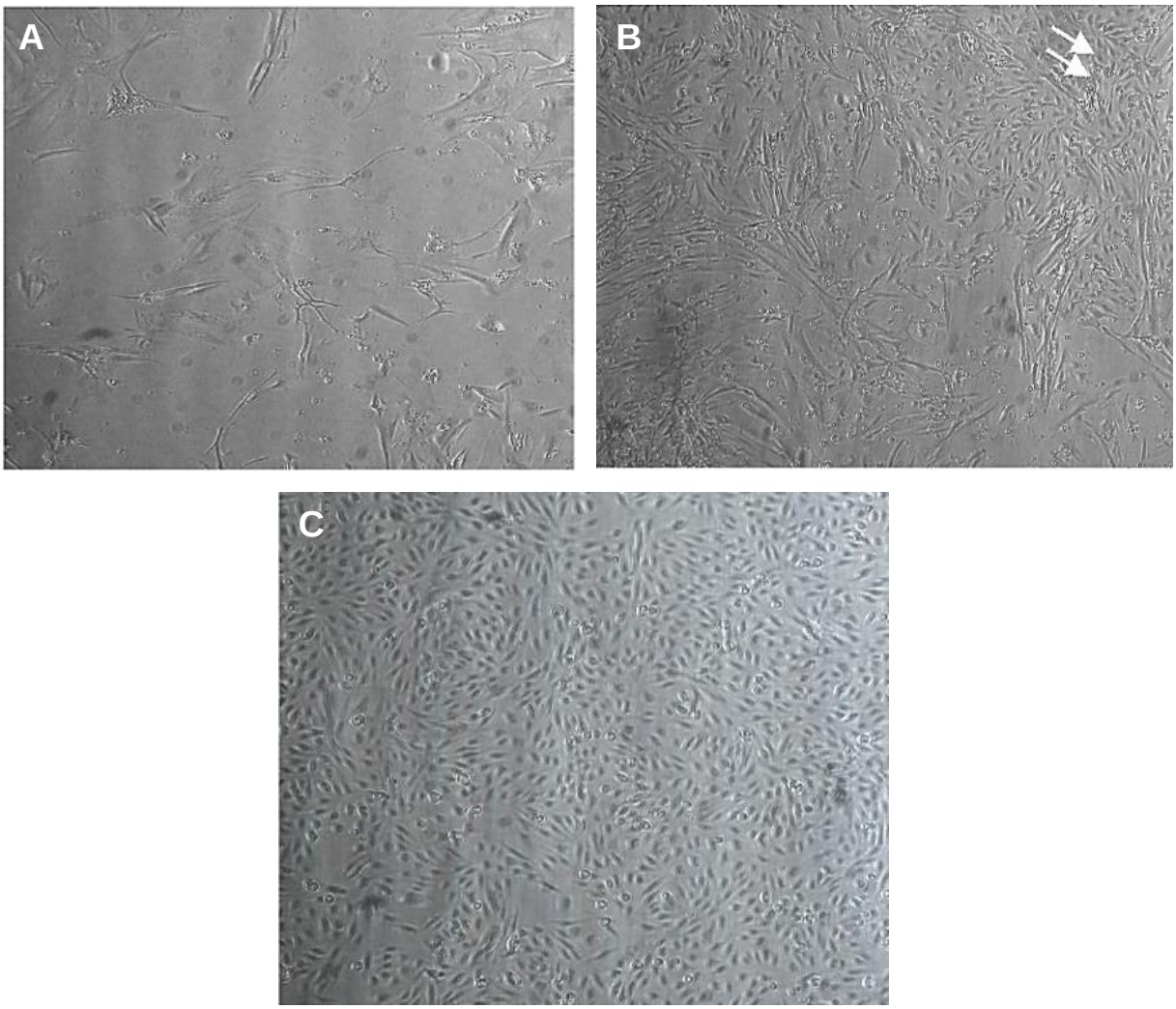

Figure 5.2: Phase-contrast micrographs of cells from STS 04/11 showing suspected cross-contamination

A – Original morphological appearance of STS 04/11 cells in culture before the ninth passage (w1p9).
B – Emergence of a morphologically-distinct clone of cells (white arrows) in STS 04/11 culture at w1p10.
C – Complete overgrowth of STS 04/11 culture with new clone of cells by w1 p13.
All images were captured at x40 magnification

Such potential cross-contamination was further investigated by DNA profiling of cells in culture by Short Tandem Repeat (STR) analysis and done using services provided by the University of Sheffield Core Genomic Facility service. Briefly, it involves the amplification of 10 STR loci in genomic DNA extracted from the cells in culture by multiplex PCR with fluorescent-labelled primers. The PCR product sizes are then determined using capillary electrophoresis and correspond to the number of nucleotide repeats in the STR (alleles). A graphical profile of all 10 STR alleles for each cell culture is then generated and compared with those of other cell lines in the laboratory to rule out or confirm a genetic match indicative of contamination.

Large studies have shown 99% confirmation of relatedness of cell lines by STR analysis when the allele match is 75 - 80% with up to 98% confirmation when allele match is >50% (American Type Culture Collection Standards Development Organization Workgroup 2010, Capes-Davis, Reid et al. 2013). The Sheffield Core Genomics Facility utilises a 70% allele match threshold for determination of probable relatedness. Comparison is also made with cell lines in the Children's Oncology Group Cell Culture and Xenograft Repository Database (www.COGcell.org) which also includes STR profiles for commercial cell line collections such as the American Type Culture Collection (ATCC), German Collection of Microorganisms and Cell Cultures (DSMZ), Japanese Collection of Research Bioresources (JCRB), and the RIKEN Research database (RIKEN).

STR analysis was performed on all the cell cultures in the laboratory at the time of suspected cross contamination and again after a further 15 months in culture to ensure that the cultures remained uncontaminated. The results are summarised on Table 5.3. Cells in STS04/11 culture by w1p13 had a profile that was completely different from those in w1p9, but identical to both variants of STS 02/11 (discussed in detail in Section 5.2.6). This confirmed that cells from STS 02/11 had cross-contaminated the STS 04/11 cultures and because they had a faster proliferative rate, were able to completely replace the native cells.

Comparison of the remaining cell cultures with profiles from the COGcell database showed no allele matches that exceeded 72%. None of the cell lines identified within the database with allele match over 70% had ever been used in the Sheffield laboratory, making them an unlikely source of contamination. Further, array CGH analysis confirmed close similarity between all the primary cell lines and their parent tissue (discussed in detail in subsequent sections). STR analysis of the same primary tumour cultures after further 30 – 40 passages showed that the profiles were largely maintained, with only three instances of allelic disparity identified (Table 5.3). All 3 instances were loss of heterozygosity (LOH) events as would be expected among cells in culture (Capes-Davis, Reid et al. 2013) and were not surprising considering the level genomic instability that is inherent among the primary tumours.

Table 5.3: Short Tandem Repeat (STR) Profiles of Soft Tissue Sarcoma Primary Cell Cultures

Lab ID	Passage	THO1	D21S11	D5S818	D13S317	D7S820	D16S539	CSFIPO	AMEL	vWA	TPOX	Allele Match	Cell line(s)
*STS 09/10	W2p35	6	29,32	12,13	8,14	9,10	9,10	10,14	X	17,18	11	72.2%	HE55
	W2P70	6	29,32	12,13	8,14	9, 10	9,10	10,14	X	17,18	11	< 70%	-
*STS 14/10	p31	6,7	27,30	12,13	8,11	8	14	12	X	16	8	< 70%	-
	p68	6,7	27,30	12,13	8,11	8	13,14	12	X	16	8	< 70%	-
*STS 02/11	W1p16	9	27,30	11	14	10,11	11	10,12	X	16	8	72.2%	HCC3153 MKN1
	W1p54	9	27,30	11	14	10,11	11	10	X	16		< 70%	-
	Wsp27	9	27,30	11	14	10,11	11	10,12	X	16	8	72.2%	HCC3153 MKN1
	Wsp63	9	27,30	11	14	10,11	11	10,12	X	16	8	72.2%	
STS 04/11	**W1p13**	**9**	**27,30**	**11**	**14**	**10,11**	**11**	**10,12**	**X**	**16**	**8**	**72.2%**	HCC 3153 MKN1
	W1p9	7,9	29,30.2	11	12	10	10,12	11,14	X	14,17	8,11	72.2%	HUC-F2
*STS 06/11	p41	6,9.3	29,31	9,13	14	8,11	11	10,11	X,Y	17,18	8	<70%	-
	p83	6,9.3	29,31	9,13	14	8,11	11	10,11	X	17,18	8	<70%	-
STS 09/11	P34	9.3	28,31.2	9,12	11	10	9	10	X	17	8	72.2%	H647 H865 HCC4018
STS 20/11	p2	9	29	9	14	10,12	12	12	X	20	9,11	< 70%	-
STS 21/11	W1p35	8	30,31.2	12,13	13	8,9	12	10	X,Y	14,16	11	< 70%	-

Profiles comprise the number of repeats in 10 STR loci (10 alleles) and were compared with one another and STR profiles for cell lines in the COGcell database. Only cell lines with >70% allele match are shown in the last column

STS 04/11 w1p13 cells (shown in red) had identical alleles to STS 02/11 cells of both variant passages (in blue), but different from STS 04/11 w1p9 cells (in green), confirming the contamination and overgrowth of STS 04/11 w1p13 by STS 02/11 cells.

** - STR profiling was repeated after 15 months (30-40 subsequent passages) and any disparities in allele matching is highlighted in grey*

5.2.3.2 Karyotyping

In order to avoid cultural artefacts, attempts were made to obtain karyotype information before cells became too highly adapted to culture. This was however very technically challenging and in the majority of cases proved unsuccessful. In cases where available metaphase spreads were available, Giemsa staining for chromosome banding (G-banding) carried out on metaphase spreads and chromosome counts were obtained. Chromosome counting was in some cases was carried out by Kristin Wright or Dr Leslie Hoh. Full karyotype analysis by G-banding (when eventually successful) was done by Dr Karen Sisley.

Repeated attempts at Giemsa staining did not always yield clear chromosome banding. Detailed structural karyotype analysis was therefore not possible even though chromosome enumeration was carried out and gross structural chromosomal abnormalities were observed. An example is shown in Figure 5.3 below.

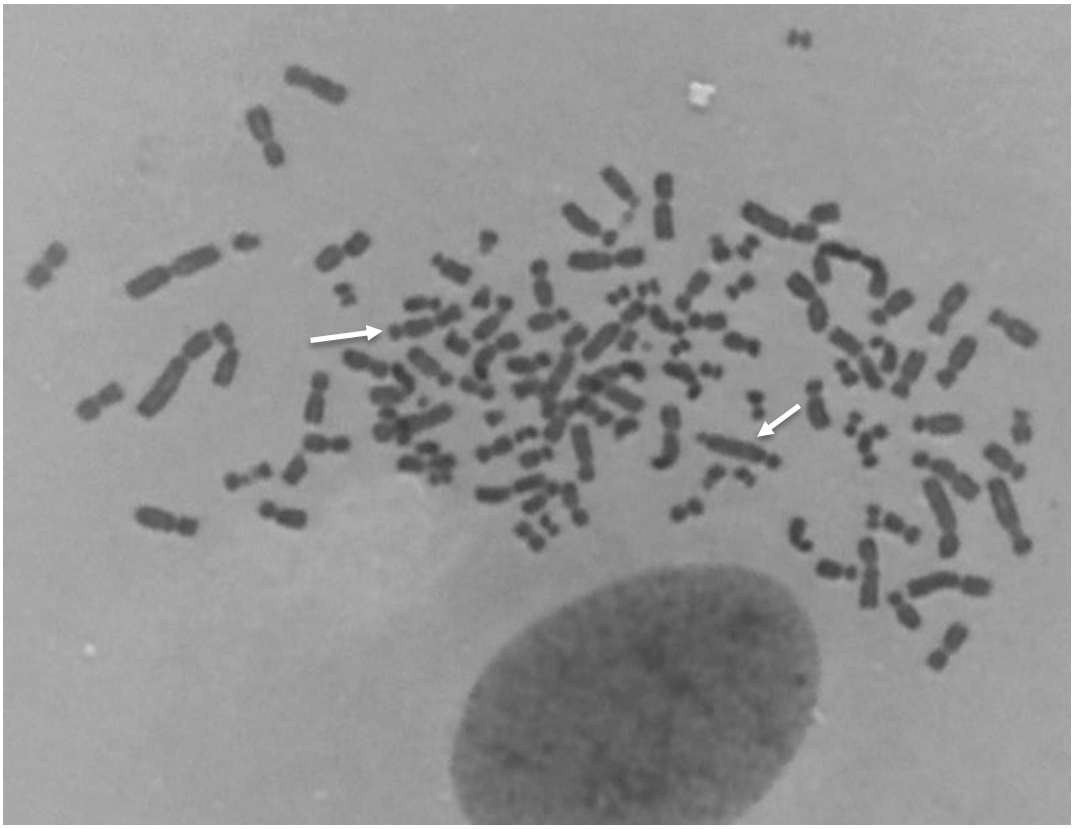

Figure 5.3: Metaphase Chromosome Spread from STS 06/11 cells

Chromosome Preparations were obtained from cells at passage 65. The chromosomes and adjacent intact nucleus are stained with Leishmann's stain. The tumour cells are polyploid with the metaphase spread containing of 105 chromosomes in total, some of which have gross structural abnormalities such as double centromeres/constrictions (white arrows).

Picture was captured n at X63 magnification

Metaphase chromosomes were obtained from early cultures in only two cases (STS 04/11 and STS 20/11). Average chromosome counts obtained from 11 metaphases in STS 04/10 (synovial sarcoma) cells following their first passage suggest a diploid karyotype, while those from STS 20/11 (dedifferentiated liposarcoma) at passage two had a near triploid karyotype (70 chromosomes on average) based on 12 metaphases counted. Chromosome spreads at later stages in culture were available from cells in all the long-term primary cultures. Table 5.4 summarises the chromosome counts obtained from primary STS cultures in this study.

Table 5.4: Chromosome counts of cells from Soft Tissue Sarcoma Primary Cell Cultures

Lab ID	Passage	Number of Metaphases counted	Range	Mean	Median	Mode
STS 04/10	P1	11	40 - 50	46	46	46
STS 09/10	W2P70	23	43 - 106	77	93	48
STS 14/10	P68	21	63 - 74	68	68	68
STS 02/11	Wsp62	22	101 - 147	125	130	130
	W1p51	23	110 - 138	132	134	136
STS 06/11*	P15	10	56 – 66	60	59	56
	P65	28	60 - 214	120	112	105
	P82	13	95 - 135	117	116	115
STS 09/11	P17	31	47 - 68	57	58	50
STS 20/11*	p2	12	63 - 73	70	72	73
	P12	31	51 - 154	109	128	69
STS 21/11	W1p22	23	63-179	111	99	66

* - *metaphase chromosomes were obtained at more than one passage*

5.2.3.3 Array CGH

Array CGH was carried out using DNA from cultured primary cells and their genomic copy number profiles were compared to those of the parent tumour tissue. A total of fourteen cases compared in this way are summarised on Table 5.5. Apart from known, non-pathologic copy number variations (CNV), the genomic profiles from both tumour tissue and cultured cells were featureless in four cases and thus uninformative. All 4 cell cultures (STS 02/10, 04/10, 05/10 and 15/10) became senescent around the fifth passage.

Table 5.5: Primary Soft Tissue Sarcoma Cell Cultures analysed by Array CGH

Lab ID	STS Subtype	Passage from which DNA was extracted for aCGH	Genomic Profile of Parent Tumour Tissue	Genomic Profile of Cultured Cells
STS 02/10	Extraskeletal Myxoid Chondrosarcoma	P2	No SCNA	No SCNA
STS 04/10	Synovial Sarcoma	P1	No SCNA	No SCNA
STS 05/10	Angiosarcoma	P2	No SCNA	No SCNA
STS 09/10	De-differentiated Liposarcoma	W2p35	Complex SCNA	Complex SCNA
STS 12/10	Pleomorphic Liposarcoma	P6	Complex SCNA	No SCNA
STS 14/10	Undifferentiated Pleomorphic Sarcoma	P5	Complex SCNA	Complex SCNA
STS 15/10	Alveolar soft Part Sarcoma	W1p1	No SCNA	No SCNA
STS 16/10	Myxofibrosarcoma	P6	Complex SCNA	No SCNA
STS 01/11	Malignant Peripheral Nerve Sheath Tumour	P1	Complex SCNA	No SCNA
STS 02/11	Leiomyosarcoma	Wsp27	Complex SCNA	Complex SCNA
STS 06/11	Undifferentiated Pleomorphic Sarcoma	P41	Complex SCNA	Complex SCNA
STS 09/11	Undifferentiated Pleomorphic Sarcoma	P26	Complex SCNA	Complex SCNA
STS 20/11	De-differentiated Liposarcoma	P2	Complex SCNA	Complex SCNA
STS 21/11	Myxofibrosarcoma	W2p28	Complex SCNA	Complex SCNA

For cases shown in blue, no SCNA were detected in either cultured cells or parent tissue,

In another 3 cases, tumour tissue showed complex genomic profiles but the corresponding cultured cells (STS 12/10, 16/10 and 01/11) had featureless array CGH profiles, suggesting that cells in culture were unlikely to represent the tumour and were probably normal cell 'contaminants' that had managed to establish in culture. Not surprisingly, none of the cell cultures went on to grow in the long term. An example of such disparate tissue *vs*. cell culture array CGH profiles (for STS 01/11) is shown in Figure 5.4. Array CGH profiling was carried out on DNA from cells in at least one variant of each long-term primary cell culture. In all cases, the genomic profiles showed features that were remarkably similar to those of the parent tumour tissue. These cases are described in detail in the subsequent sections.

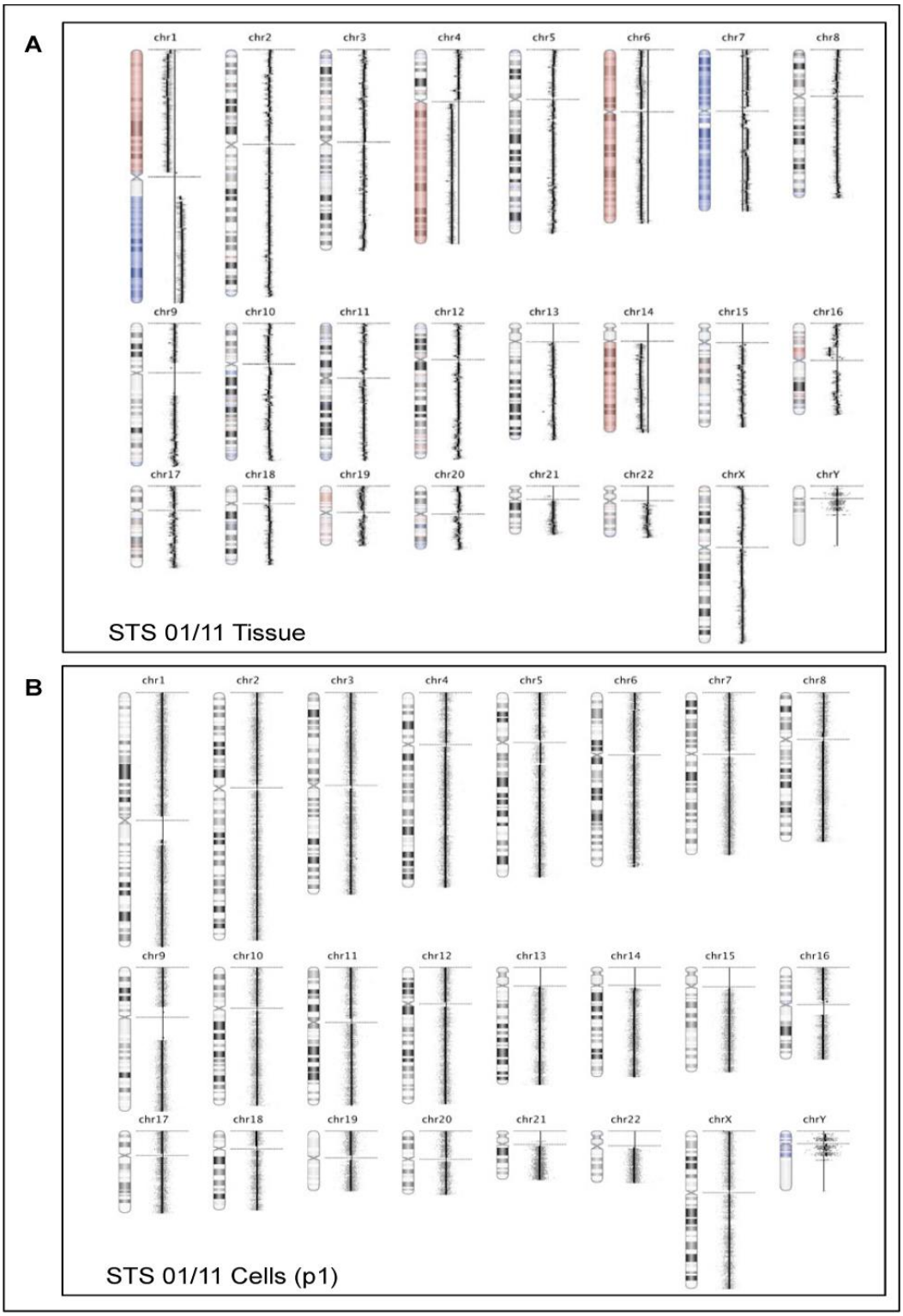

Figure 5.4: Genome View Ideograms of STS 01/11 comparing cultured cells with parent tumour tissue

Panel A – whole genome ideogram for STS 01/11 parent tumour tissue DNA showing multiple SCNA detected across the genome, compared with **Panel B** – whole genome ideogram of DNA from cells cultured from STS 01/11 tissue (at passage 1) showing no SCNA detected across all the autosomes.
Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Small dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the magnitude of aberration (log ratio). *All CNAs were detected using the FASST2 algorithm*

5.2.4 STS 09/10: De-differentiated Liposarcoma

5.2.4.1 Tumour Characteristics

Tumour samples were kindly donated by a 68-year old female patient whose tumour was diagnosed as a Trojani Grade 3 retroperitoneal dedifferentiated liposarcoma with a maximum dimension of approximately 300mm. No distant metastases were evident at the time of surgical excision, but the tumour was locally invasive.

5.2.4.2 Primary Tissue Culture Characteristics

Tissue culture was set up as described in Section 2.3.1.1. Two washes were set up because it appeared that viable cells remained in suspension. Cells from the original flasks, slope and first wash became senescent before the tenth passage.

Cells from the second wash, given the specific designation 'STS 09/10 W2' however became established in culture after approximately four weeks. The cells were mostly spindle-shaped, a few being slightly polygonal (Figure 5.5). They have remained viable in culture since August 2010 and been passaged over 70 times. The doubling time, based on MTT proliferation assay was estimated at 49.5 hours (Figure 5.6).

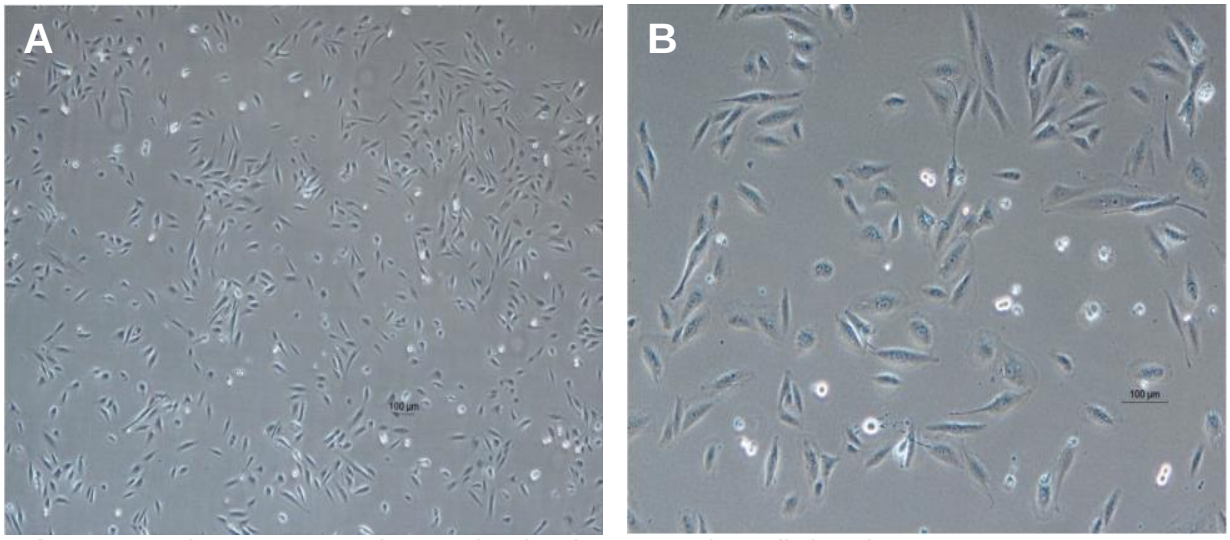

Figure 5.5: Phase-contrast Micrographs showing STS 09/10 cells in culture

Cells shown were at passage 71 and were derived from the second wash of cultures set up from STS 09/10 tissue. **A** – Images captured at x40 magnification showing the pattern of growth without distinct colonies. **B**- Images captured at x100 magnification showing the fibroblast-like spindle shaped morphology of most cells.

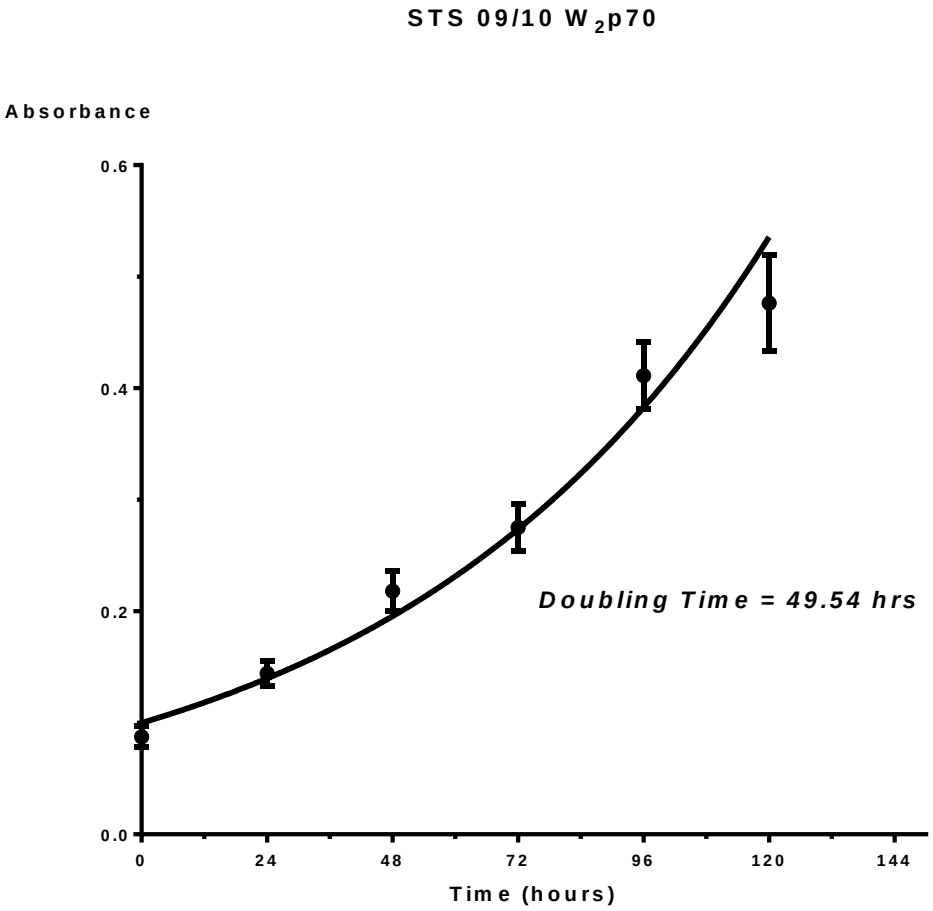

Figure 5.6: STS 09/10 cell Growth evaluated by MTT proliferation assay

Cells were at passage 70 and were derived from the second wash of cultures set up from STS 09/10 tissue. Data is representative of experiments done in quadruplicate. Curve fitting was done using GraphPad® Prism software (version 6) and the doubling time was calculated using the Exponential Growth Equation in the same software.

5.2.4.3 Genetic Characterisation

Karyotype

None of the metaphase spreads obtained from the cells in culture had a normal chromosome number. The majority were however near-diploid, having around 48 chromosomes (Table 5.4).

Array CGH

Array CGH was carried out using DNA extracted from cells after 35 passages and the genomic copy number profile compared to that from the fresh frozen parent tissue as shown in Figure 5.7. Both profiles showed very similar SCNA patterns on most chromosomes, including identical discrete amplicons on 12q13 –15 that involve the putative driver genes, *MDM2* and *CDK4*. Another ten chromosomes showed near-identical aberration breakpoints such as those detected in chromosomal regions such as 1q, 2q-ter, 5p, 6q, 7q, 8p, 8q, 9p, 10p, 14q, and 16p (Figure 5.7). Two chromosomes (3 and 21) showed no SCNA in either sample. A number of genomic regions showed dissimilar copy number profiles, most notable of which was chromosome 11. These dissimilarities however were relatively few (on six chromosomes) compared to the remarkable similarities seen on 13 chromosomes (Figure 5.7).

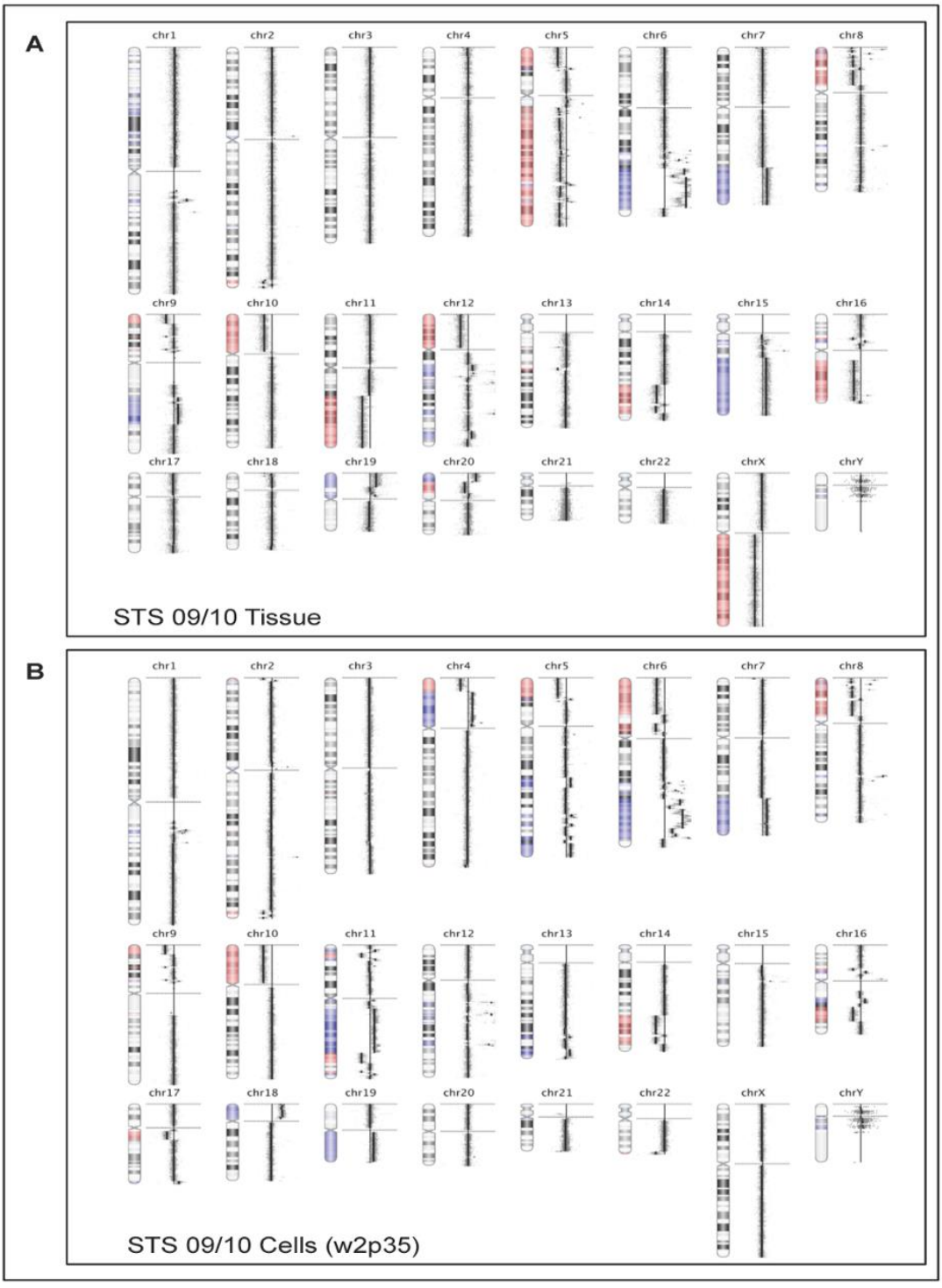

Figure 5.7: Genome View Ideograms of STS 09/10 comparing cultured cells with parent tumour tissue

Panel A – whole genome ideogram for STS 09/10 parent tumour tissue DNA showing multiple SCNA detected across the genome, compared with **Panel B** – whole genome ideogram of DNA from cells (at passage 35) derived from a wash taken off original cultures of STS 09/10 tissues that were set up in flasks, showing multiple SCNA detected across all autosomes and similar average log ratio patterns overall.
Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Small dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the magnitude of aberration (log ratio). *All CNAs were detected using the FASST2 algorithm*

5.2.5 STS 14/10: Undifferentiated Pleomorphic Sarcoma

5.2.5.1 Tumour Characteristics

Tumour samples were obtained from a large (230mm maximum dimension) tumour that was widely excised from the right thigh of a 54-year old female patient. Distant metastases were not evident at the time of surgical resection and no neo-adjuvant therapy had been administered. The tumour was diagnosed as an undifferentiated pleomorphic sarcoma of Trojani grade 3.

5.2.5.2 Primary Tissue Culture Characteristics

Primary cultures were set up as described in Section 2.3.1.1. No washes were required as most cells became established as adherent cultures within a few days. Cultures are composed of a morphologically homogenous clone of fibroblast-like spindle-shaped cells (Figure 5.8). The cells proliferate with an estimated doubling time of 40.5 hours (Figure 5.9) and have been passaged more than 70 times in over 30 months of culture.

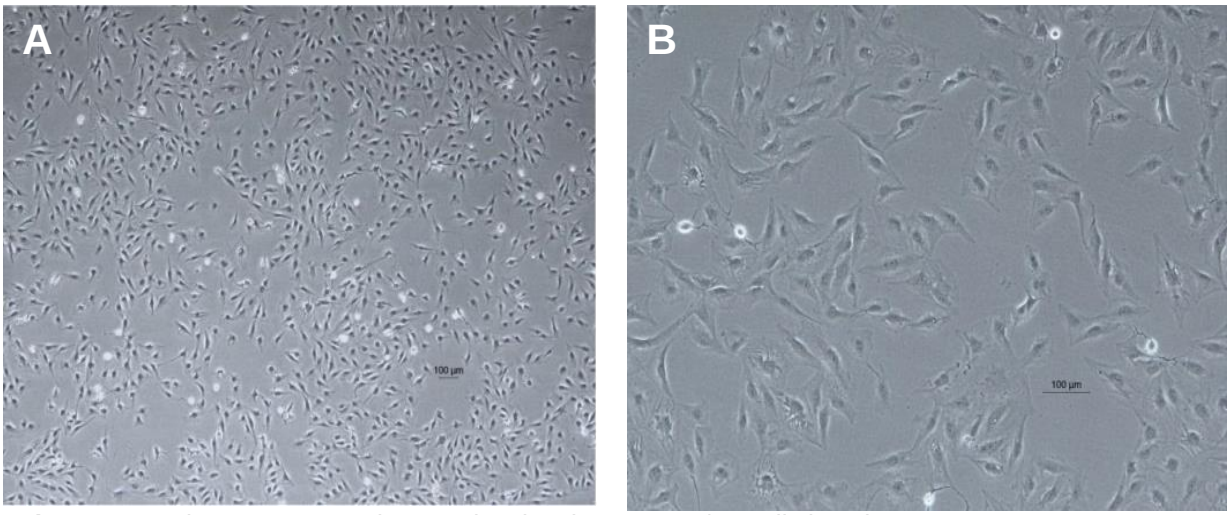

Figure 5.8: Phase-contrast Micrographs showing STS 14/10 cells in culture

Cells shown were at passage 72 and were derived from cultures set up from STS 14/10 tissue.
A – Images captured at x40 magnification showing the pattern of growth without distinct colonies.
B- Images captured at x100 magnification showing the fibroblast-like spindle shaped morphology of most cells.

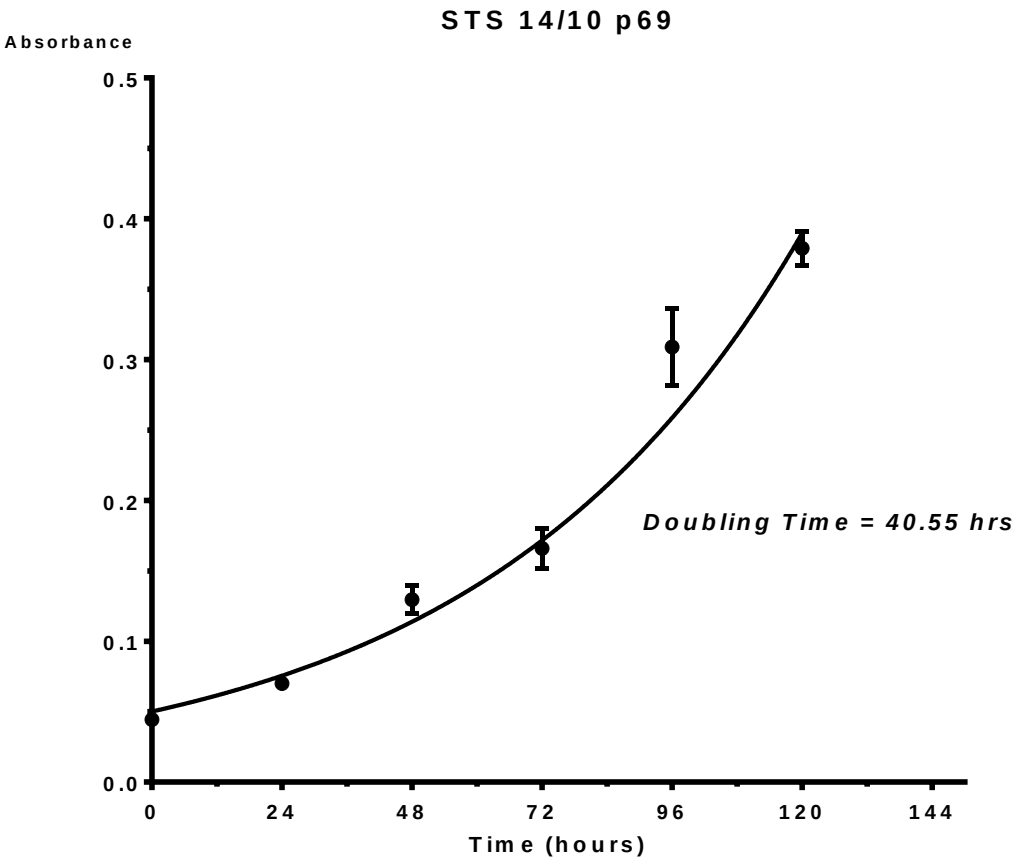

Figure 5.9: STS 14/10 Cell Growth evaluated by MTT proliferation assay

Cells were at passage 69 and were derived from cultures set up from STS 14/10 tissue. Data is representative of experiments done in quadruplicate. Curve fitting was done using GraphPad® Prism software (version 6) and the doubling time was calculated using the Exponential Growth Equation in the same software.

5.2.5.3 Genetic Characterisation

Karyotyping

Metaphase spreads were obtained from STS 14/10 cells at passage 68. A pseudo-triploid karyotype was seen in most metaphases spreads with cells containing between 63 and 74 chromosomes and a mean chromosome count of 69 (Table 5.4).

Array CGH

The copy number profile of DNA obtained from STS 14/10 cells at the fifth passage was compared to that from the parent tissue. Profiles showed near-identical SCNA breakpoints on 9 different chromosomes, including a specific deletion on 13q14 that involved the locus of the *RB1* gene (Figure 5.10). Three chromosomes (2, 3 and 11) showed no SCNA in either cultured cell or parent tissue copy number profiles. Only the aberrations detected on 4p, 6p, 7q and 20p were notably dissimilar between the profiles, which are shown in whole genome ideograms in Figure 5.10.

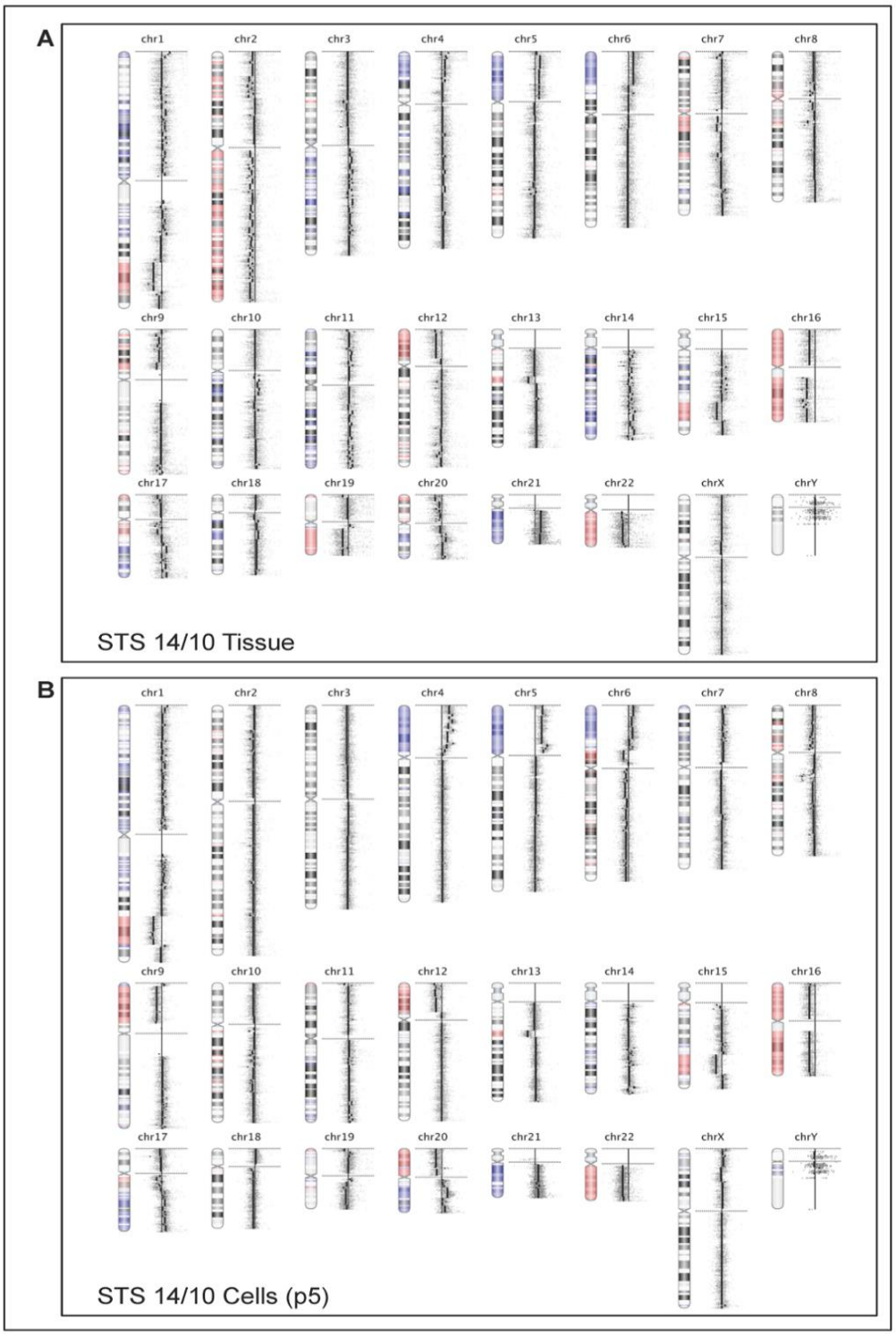

Figure 5.10: Genome View Ideograms of STS 14/10 comparing cultured cells with parent tumour tissue

Panel A – whole genome ideogram for STS 14/10 parent tumour tissue DNA showing multiple SCNA detected across the genome, compared with **Panel B** – whole genome ideogram of DNA from cells cultured from STS 14/10 tissue (at passage 5) showing multiple SCNA detected across all autosomes and similar average log ratio patterns overall.

Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Small dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the magnitude of aberration (log ratio). *All CNAs were detected using the FASST2 algorithm.*

5.2.6 STS 02/11: Leiomyosarcoma

5.2.6.1 Tumour Characteristics

Tumour samples were a kindly donated from by a 62-year old female patient who had wide resection as initial treatment for a 135mm Trojani grade 3 vaginal leiomyosarcoma.

5.2.6.2 Primary Tissue Culture Characteristics

Cultures initially established slowly proliferating spindle-shaped cells and washes set up from the slopes and flasks were designated 'w_s' and 'w_1', respectively. An appearance suggestive of early senescence was observed in all the cultures but they were maintained in culture as per protocol.

Variant 1 – 'STS 02/11 w_s'

After around 8 weeks in culture during which they were passaged only twice, a single clone of polygonal cells with distinct nuclei was observed among the largely senescent cells in the cultures designated 'w_s'. This clone eventually outgrew all other cells in the flask and has now been proliferating since March 2011 without notable further change in morphology. The cells grew in tight colonies that increased in size until confluence (Figure 5.11). The cells have so far been sub-cultured over 65 times and have an estimated doubling time of approximately 45 hours (Figure 5.12).

Variant 2 – 'STS 02/11 w_1'

Following an even longer lag phase (around 12 weeks in culture), another distinct clone of cells was observed in cultures designated w_1. Compared to the clone that formed the w_s variant, these cells were spindle-shaped rather than polygonal and had less distinct nuclei. They however also grew in tight colonies (Figure 5.13). Initial suspicion of contamination of STS 02/11 cell cultures by cells from another cell line was allayed by STR profiling of cells from both variants that showed identical alleles for both clones, which were distinct from all other cell lines used in the same laboratory (Table 5.3). Cells from the W_1 variant have now been in passaged over 55 times and have an estimated doubling time of 27.4 hours (Figure 5.14)

5.2.6.3 Genetic Characterisation

Karyotyping

Metaphase spreads from STS 02/11 w_s variant cells were highly polyploid and contained between 101 and 147 chromosomes each, many of which had gross structural abnormalities (not shown). Similar chromosome counts were obtained from the w_1 variant cells that were examined, which had 110 – 138 chromosomes per metaphase spread (Table 5.4).

Array CGH

Genome view ideograms of the DNA copy number profiles of STS 02/11 w_s cells at passage 27 and that from the parent tumour tissue were compared as shown in Figure 5.15. A number of SCNA detected in the cultured cells were different from those in parent tissue. For example, deletions on 5p seen in the parent tissue were not identified in this clone the cultured cells and conversely, deletions on 7q observed in the cultured cells but not in the parent tissue (Figure 5.15). However, moving average patterns of probe log ratios on many chromosomes were very similar in both profiles and had near identical breakpoints as seen on 9p, 11q, 19q and 20q. Two chromosomes (2 and 21) bore no SCNA in either sample.

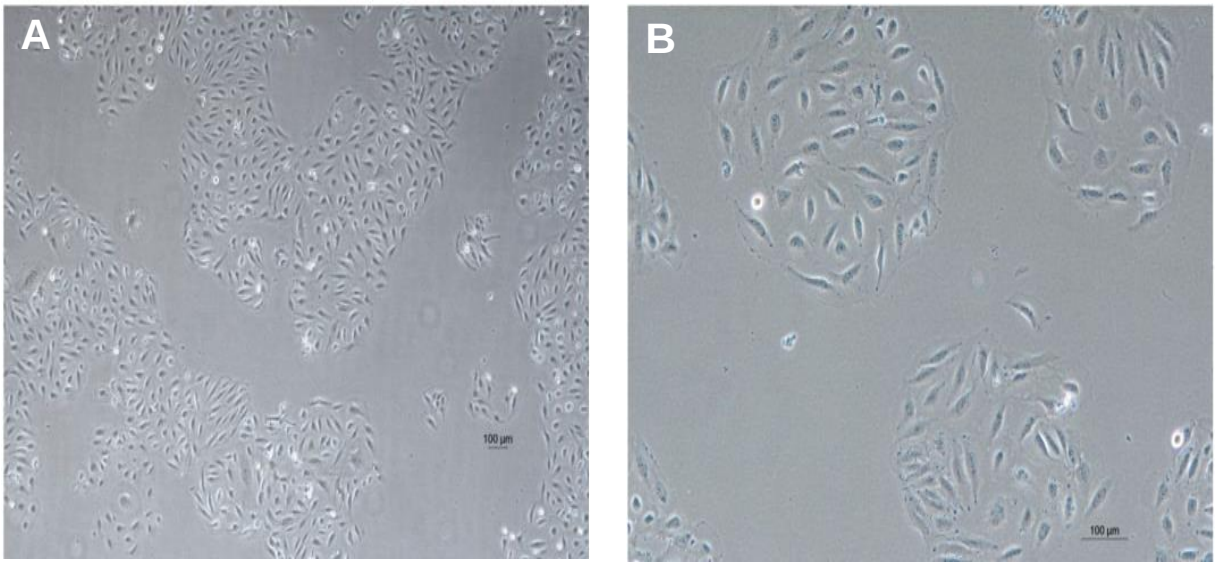

Figure 5.11: Phase-contrast Micrographs showing the 'ws' variant of STS 02/11 cells in culture

Cells shown were at passage 69 and were derived from a wash taken off cultures of STS 02/ 11 tissues that were set up in slopes. **A** – Images captured at x40 magnification showing the pattern of growth in tightly-pack colonies. **B**- Images captured at x100 magnification showing the polygonal morphology of most cells with distinct nuclei.

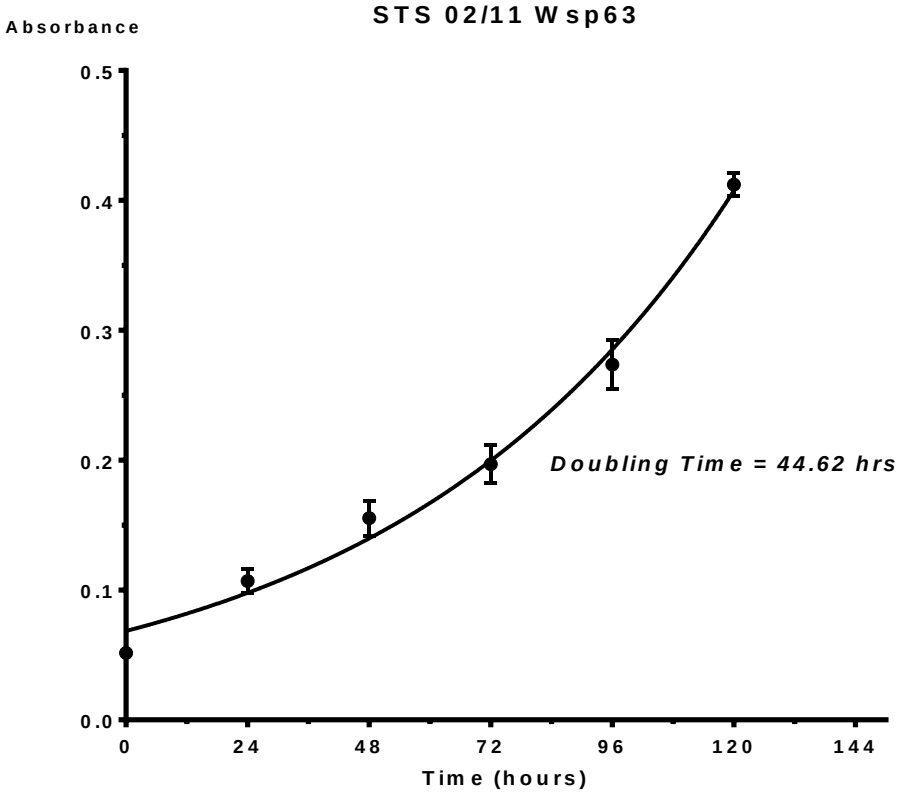

Figure 5.12: STS 02/11 'w_s' variant Cell Growth evaluated by MTT Proliferation Assay

Cells were at passage 63 and were derived from a wash taken off cultures of STS 02/ 11 tissues that were set up in slopes. Data is representative of experiments done in quadruplicate. Curve fitting was done using GraphPad® Prism software (version 6) and the doubling time was calculated using the Exponential Growth Equation in the same software.

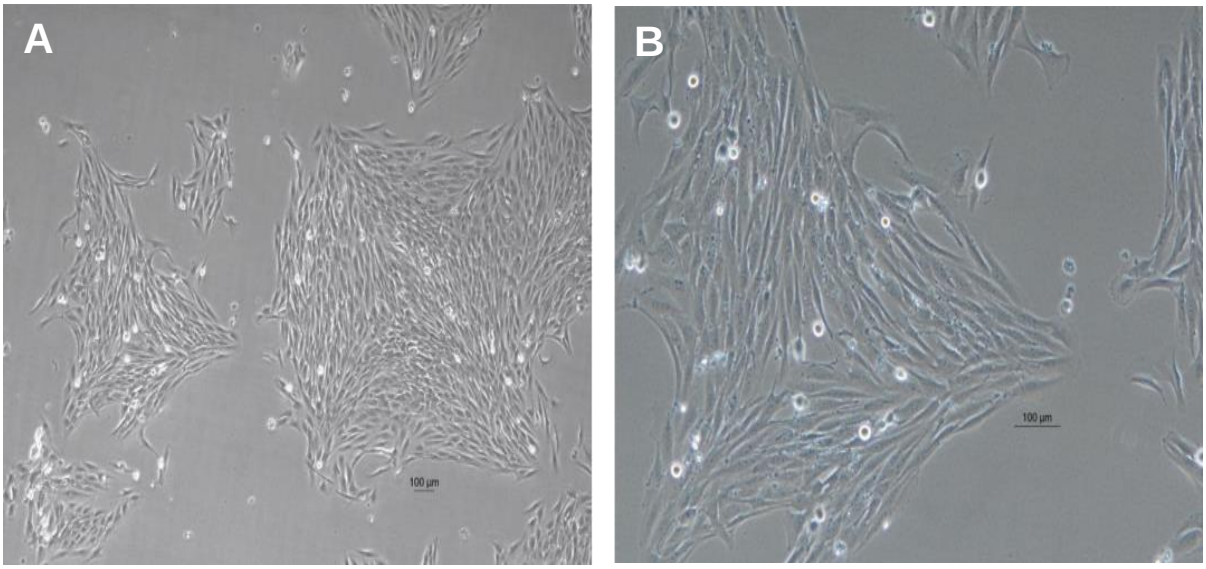

Figure 5.13: Phase-contrast Micrographs showing STS 02/11 cells in culture

Cells shown are at passage 56 and were derived from the first wash taken off cultures of STS 02/ 11 tissues that were set up in flasks. **A** – Images captured at x40 magnification showing the pattern of growth in tightly-packed colonies. **B**- Images captured at x100 magnification showing the fibroblast-like spindle shaped morphology of most cells.

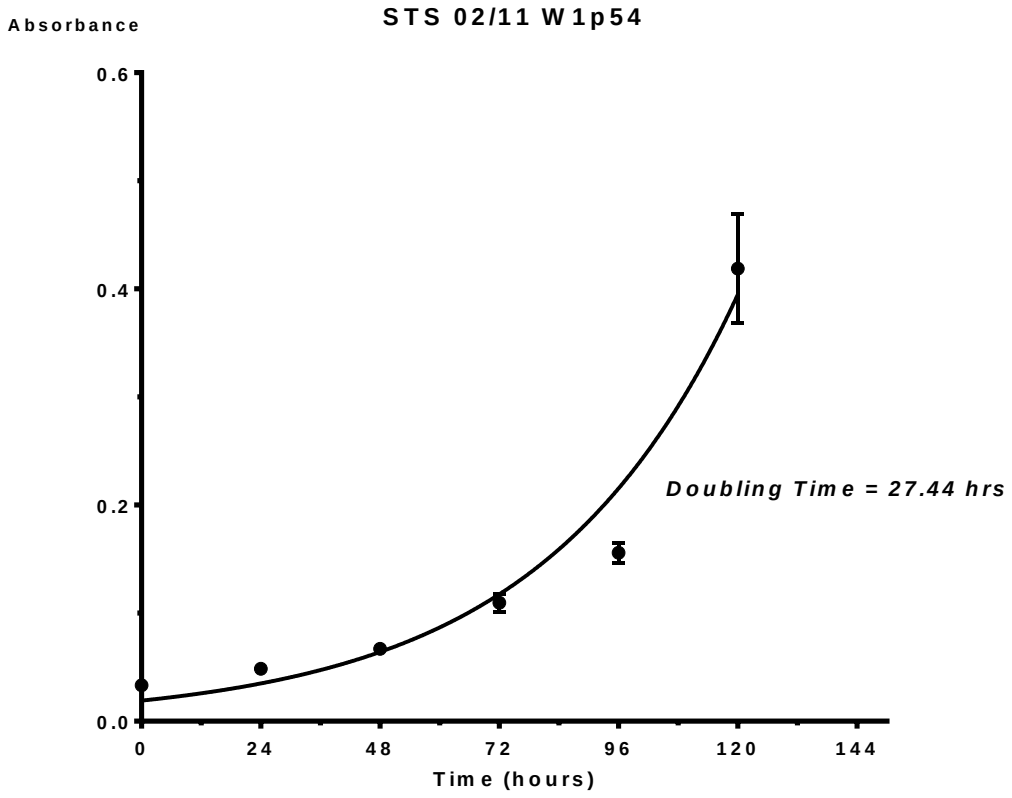

Figure 5.14: STS 02/11 'w1' variant cell Growth evaluated by MTT proliferation assay

Cells were at passage 70 and were derived from the first wash taken off cultures of STS 02/ 11 tissues that were set up in flasks. Data is representative of experiments done in quadruplicate. Curve fitting was done using GraphPad® Prism software (version 6) and the doubling time was calculated using the Exponential Growth Equation in the same software.

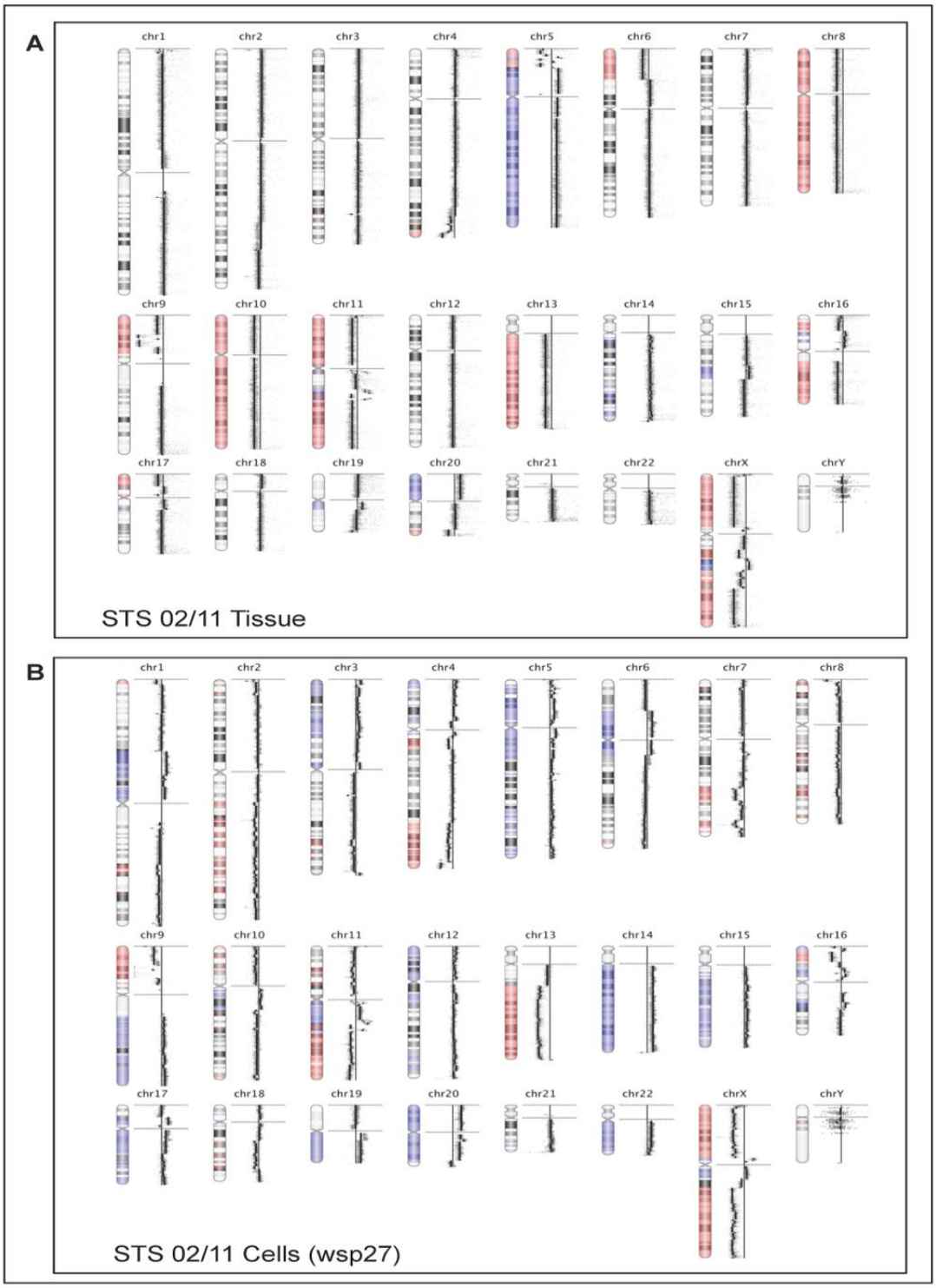

Figure 5.15: Genome View Ideograms of STS 02/11 comparing cultured cells of the 'w_s' variant with parent tumour tissue

Panel A – whole genome ideogram for STS 02/11 parent tumour tissue DNA showing multiple complex SCNA detected across the genome, compared with **Panel B** – whole genome ideogram of DNA obtained from cells (at passage 27) derived from a wash taken off cultures of STS 02/ 11 tissues that were set up in slopes, showing multiple SCNA detected across all autosomes and similar average log ratio patterns overall.
Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Small dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the magnitude of aberration (log ratio). *All CNAs were detected using the FASST2 algorithm*

5.2.7 STS 06/11: Undifferentiated Pleomorphic Sarcoma

5.2.7.1 Tumour Characteristics

An 83-year old male patient kindly donated the tumour samples from which cultures were derived. With a maximum dimension of 170mm, the tumour was located in his right thigh and was surgically excised without neo-adjuvant therapy. It was subsequently diagnosed as an undifferentiated pleomorphic sarcoma of Trojani grade 3 and no metastases were evident at the time of surgery.

5.2.7.2 Primary Tissue Culture Characteristics

Cultures set up according to normal protocol became established within a few days. Designated STS 06/11, the cells have remained slightly pleomorphic in culture but majority have a spindle-shaped morphology (Figure 5.16). They grow in loose colonies and lack contact inhibition, able to form multiple layers in culture that if left long enough, can be seen macroscopically (not shown).

In culture since February 2011, the rapidly proliferating STS 06/11 cells have an estimated doubling time of 35.38 hours (Figure 5.17) and have been passaged over 90 times. They also exhibit *in vitro* migratory and invasive capabilities when used in micro-chemotactic Boyden chamber assay (discussed in detail in chapter 6).

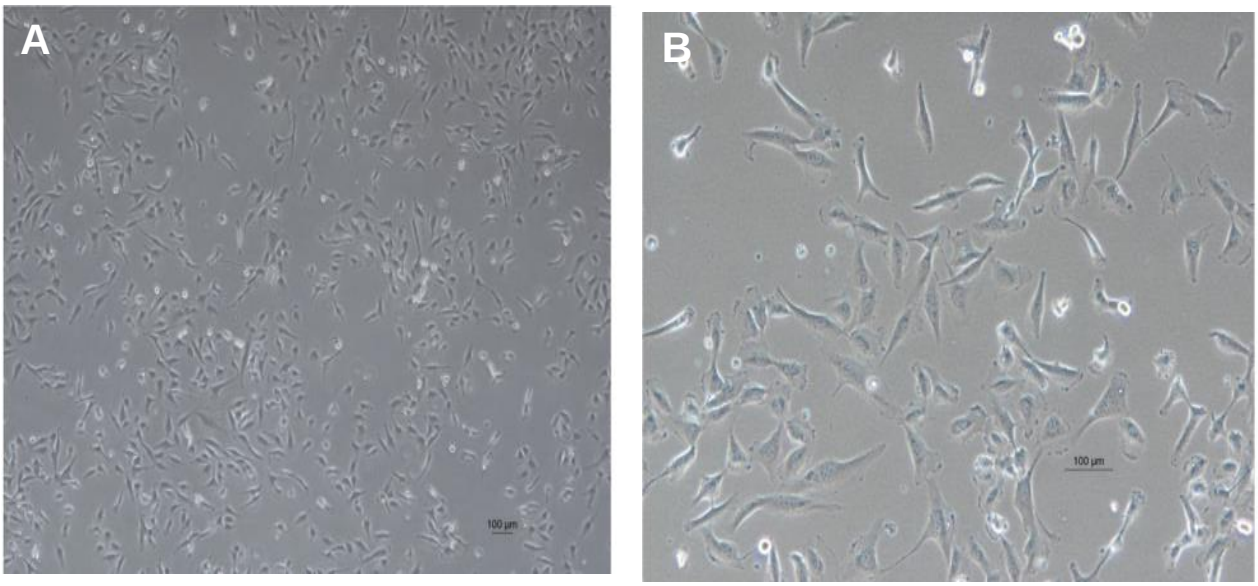

Figure 5.16: Phase-contrast Micrographs showing STS 06/11 cells in culture

Cells shown are at passage 91 and were derived from cultures set up from STS 06/11 tissue.
A – Images captured at x40 magnification showing the pattern of growth without distinct colonies.
B- Images captured at x100 magnification showing the fibroblast-like spindle shaped morphology of most cells.

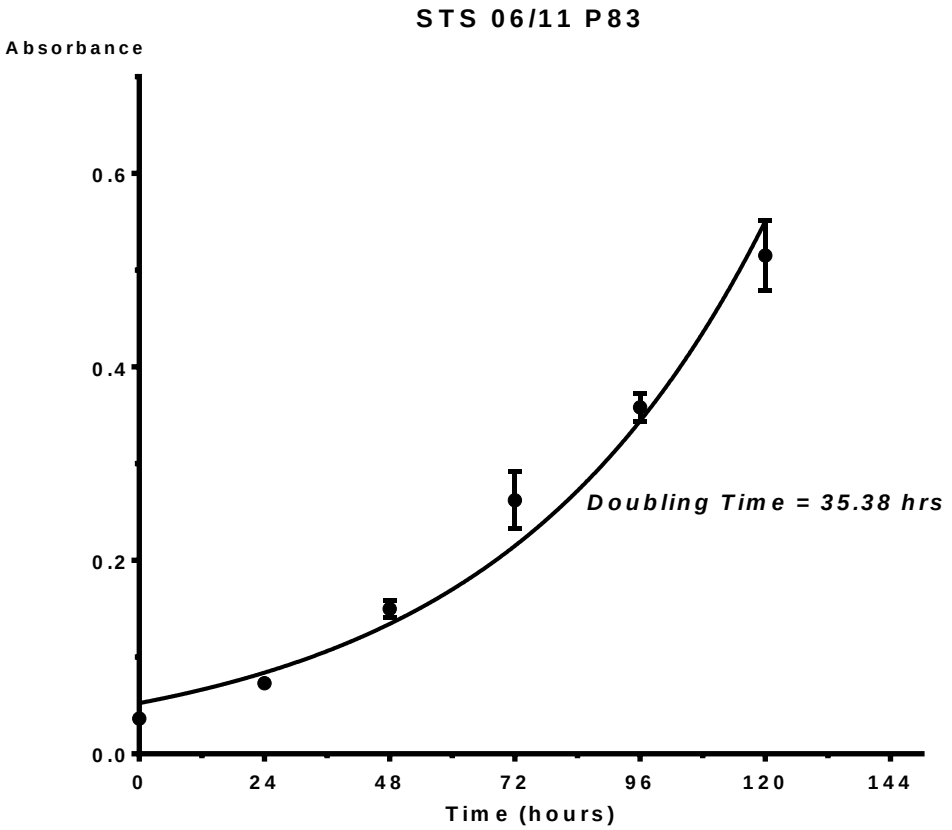

Figure 5.17: STS 06/11 cell Growth evaluated by MTT proliferation assay

Cells were at passage 83 and were derived from cultures set up from STS 06/11 tissue. Data is representative of experiments done in quadruplicate. Curve fitting was done using GraphPad® Prism software (version 6) and the doubling time was calculated using the Exponential Growth Equation in the same software.

5.2.7.3 Genetic Characterisation

Karyotyping

Metaphase spreads were initially obtained from STS 06/11 cells at passage 15, when the chromosome counts were between 56 and 66 (Table 5.4). With prolonged culture, the number of chromosomes per cell increased and became more variable. Metaphase chromosomes obtained from cultures at passage 82 ranged from 95 to 135 per cell (n = 20; mean = 117; median = 116; mode = 115). Full karyotype analysis is currently on-going.

Array CGH

Genomic copy number profile of STS 06/11 cells after 41 passages in culture bore remarkable similarity to that of the parent tumour tissue as shown in Figure 5.18. Nearly all affected chromosomes showed SCNA with identical breakpoints, including complex aberrations such as those on 11q. Log ratio amplitudes of SCNA were higher in the cultured cells than the parent tumour tissue in certain genomic regions such as those on the short arm of chromosome 9. However, the overall pattern of log ratio moving averages in this region remained similar (Figure 5.18).

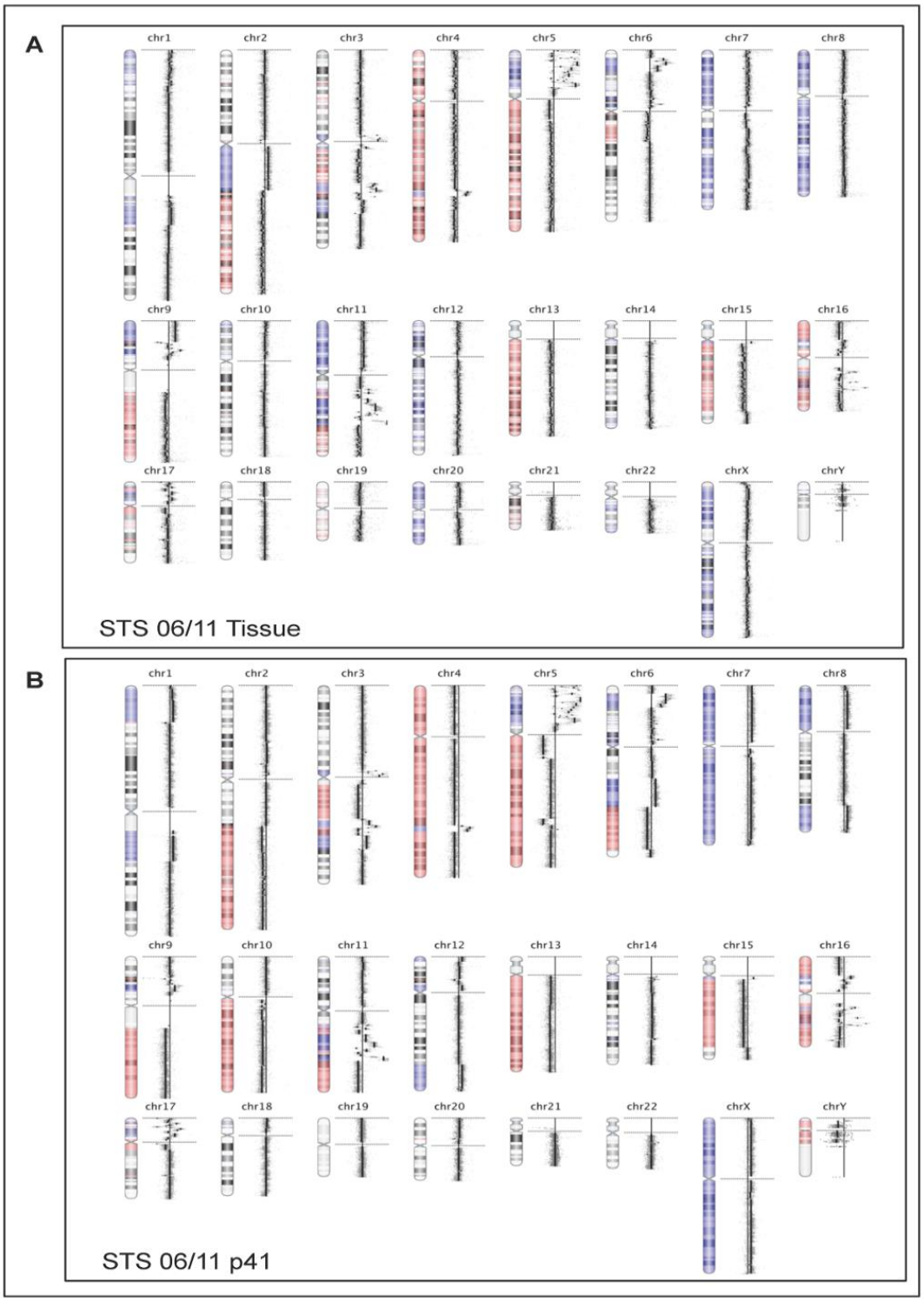

Figure 5.18: Genome View Ideograms of STS 06/11 comparing cultured cells with parent tumour tissue

Panel A – whole genome ideogram for STS 06/11 parent tumour tissue DNA showing multiple SCNA detected across the genome, compared with **Panel B** – whole genome ideogram of DNA from cells cultured from STS 09/10 tissue (at passage 41) showing multiple SCNA detected across all autosomes and similar average log ratio patterns overall.

Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Small dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the magnitude of aberration (log ratio). *All CNAs were detected using the FASST2 algorithm*

5.2.8 STS 09/11: Undifferentiated Pleomorphic Sarcoma

5.2.8.1 Tumour Characteristics

Tumour samples were a kind donation from a 76-year-old female patient who had wide surgical excision of a right lower limb tumour. Measuring 115mm in maximum dimension, the tumour was subsequently diagnosed as a Trojani grade 3, undifferentiated pleomorphic sarcoma. Distant metastases were not evident at the time of surgery and neo-adjuvant therapy was not administered.

5.2.8.2 Primary Tissue Culture Characteristics

Cultures were set as per normal protocol (Section 2.3.1.1) and washes were not required. Adherent cells in culture had a long, fibroblast-like morphology and grew in loose colonies that formed monolayers (Figure 5.19). STS 09/11 cells showed a relatively slow rate of proliferation with their doubling time estimated at approximately 63 hours using the MTT assay (Figure 5.20). Even though the cells have been in culture since June 2011, they have only been passaged around 35 times.

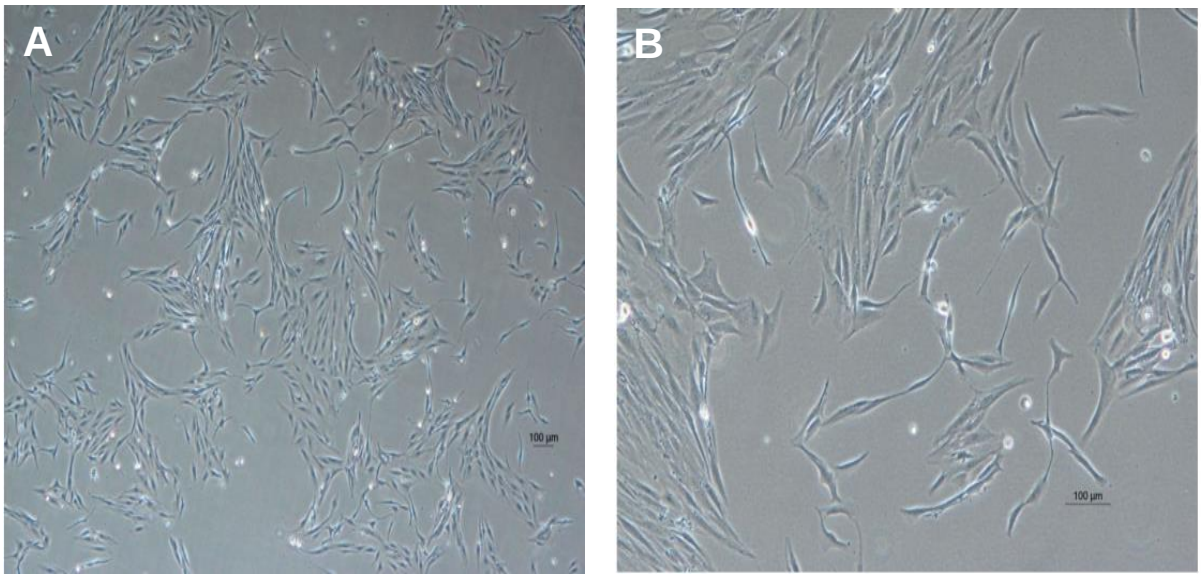

Figure 5.19: Phase-contrast Micrographs showing STS 09/11 cells in culture

Cells shown are at passage 35 and were derived from cultures set up from STS 09/11 tissue.
A – Images captured at x40 magnification showing the pattern of growth of cells in loose colonies.
B- Images captured at x100 magnification showing the long fibroblast-like spindle-shaped morphology of most cells.

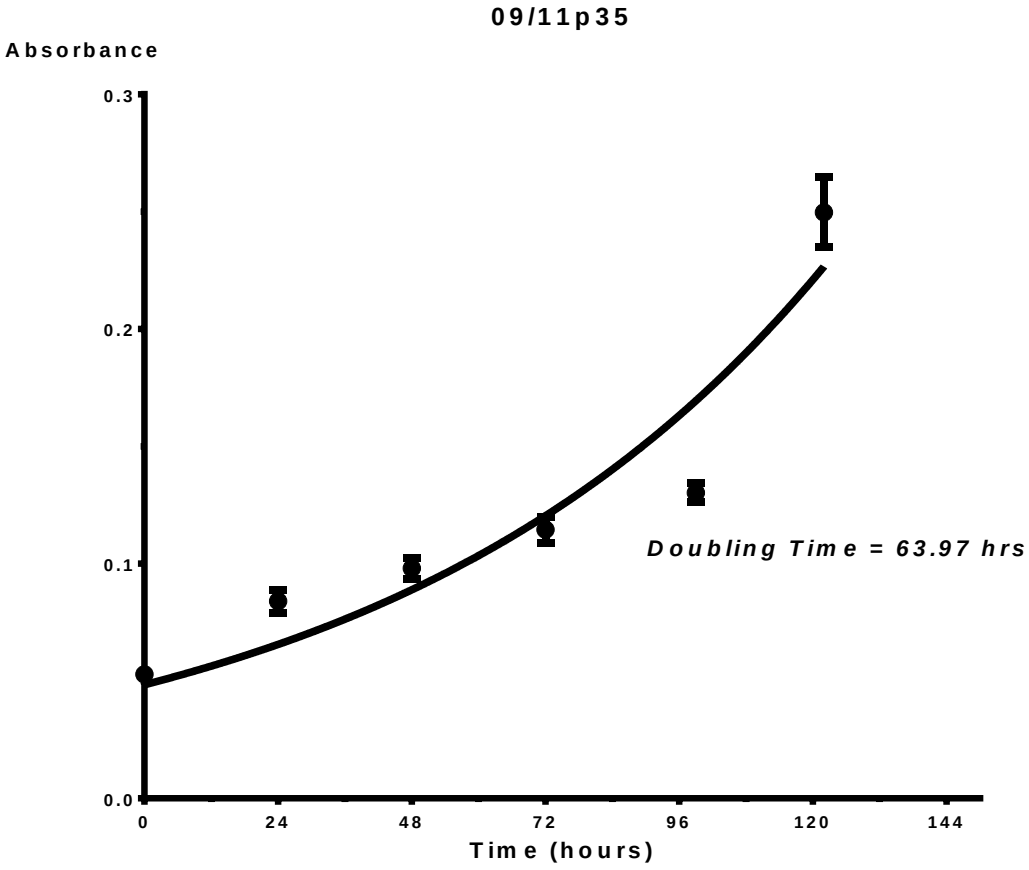

Figure 5.20: STS 09/11 cell Growth evaluated by MTT proliferation assay

Cells were at passage 35 and were derived from cultures set up from STS 09/11 tissue. Data is representative of experiments done in quadruplicate. Curve fitting was done using GraphPad® Prism software (version 6) and the doubling time was calculated using the Exponential Growth Equation in the same software.

5.2.8.3 Genetic Characteristics

Karyotyping

Metaphase spreads were obtained from STS 09/11 cells at passage 17. None of the metaphases examined had a diploid chromosome number and karyotypes were composed of between 47 and 68 chromosomes each (Table 5.4). Detailed structural analysis of the chromosomes is currently on-going.

Array CGH

Comparison of the array CGH profile of STS 09/11 cells at passage 26 with that of the parent tissue showed close similarity on the majority of autosomes (Figure 5.21). Few genomic regions, such as chromosomes 3p, 5 and 9 showed SCNAs in the cultured cells that were different form the parent tumour. Examination of the moving average pattern of log ratios even in regions with complex aberrations, showed close similarity in both profiles including regions such as that on 17q where the calling algorithm identified differential aberrations (Figure 5.21).

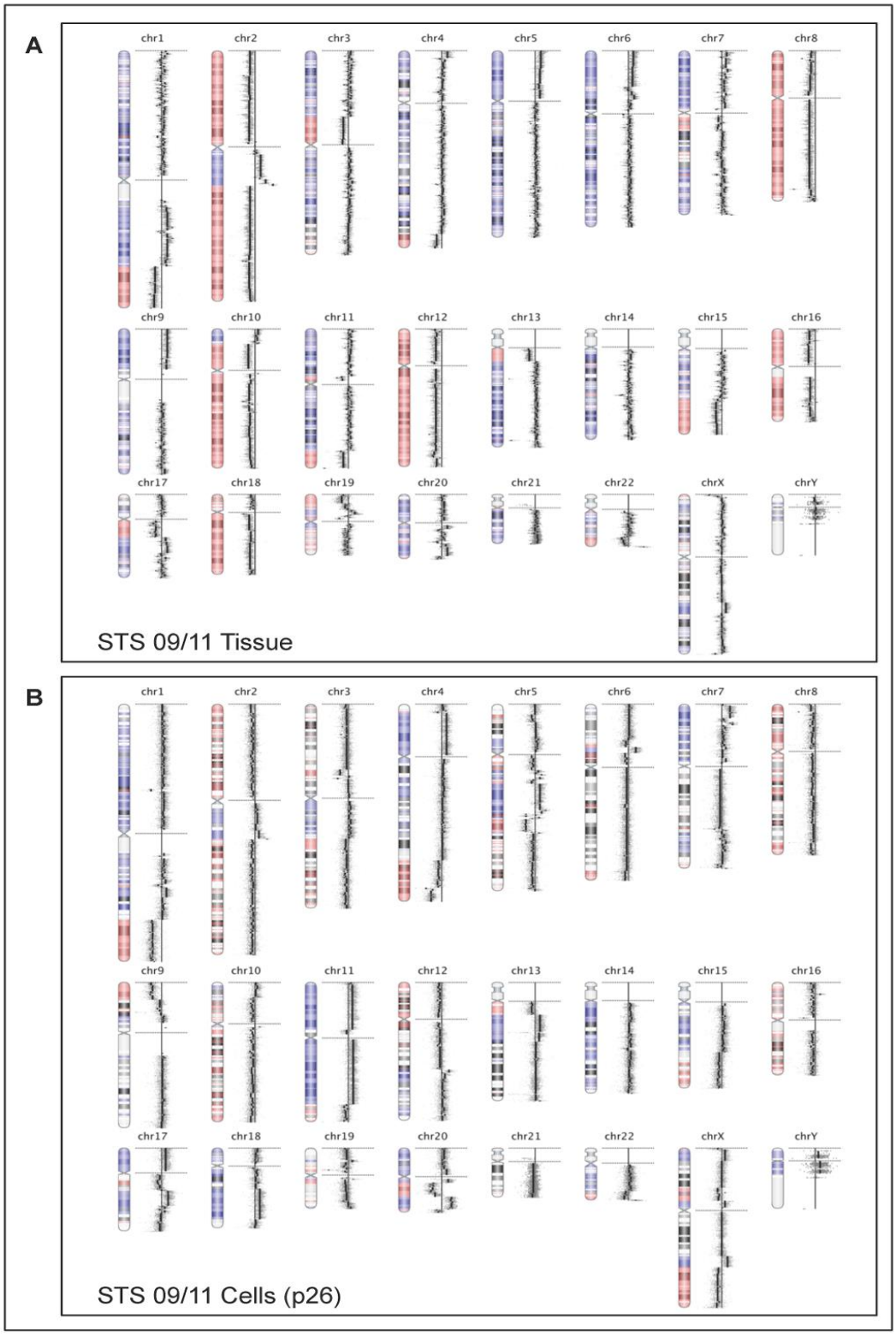

Figure 5.21: Genome View Ideograms of STS 09/11 comparing cultured cells with parent tumour tissue

Panel A – whole genome ideogram for STS 09/11 parent tumour tissue DNA showing multiple SCNA detected across the genome, compared with **Panel B** – whole genome ideogram of DNA from cells cultured from STS 09/10 tissue (at passage 26) showing multiple SCNA detected across all autosomes and similar average log ratio patterns overall.

Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Small dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the magnitude of aberration (log ratio). *All CNAs were detected using the FASST2 algorithm*

5.2.9 STS 20/11: Dedifferentiated Liposarcoma

5.2.9.1 Tumour Characteristics

The 70 year old female patient who donated the tumour designated as STS 20/11 had a large (170mm maximum dimension) tumour widely excised from her left thigh. Distant metastases were not evident at the time of surgery, but the tumour was locally invasive and required neo-adjuvant radiotherapy before it could be safely resected. Subsequent histopathological analysis led to its classification as a de-differentiated liposarcoma with around 30% viability. Trojani grading was not applicable because of the high level of radiation-induced necrosis (Coindre, Nguyen et al. 1988). However, dedifferentiated liposarcomas are generally regarded as high grade tumours (Fletcher, Bridge et al. 2013).

5.2.9.2 Primary Tissue Culture Characteristics

Cultures were set up according to normal protocol (Section 2.3.1.1). Washes were not required because viable cells formed adherent cultures in the original flasks. This was surprising considering the high level of necrosis in the tumour as a whole. The cells formed loose colonies in culture (Figure 5.22A) and consisted of a combination of spindle-shaped fibroblast-like cells and more polygonal cells with distinct nuclei (Figure 5.22B). They proliferate at a slow rate relative to other STS cultures in this study and have an estimated doubling time of around 58 hours (Figure 5.23). They have only been passaged around 25 times in the last 18 months.

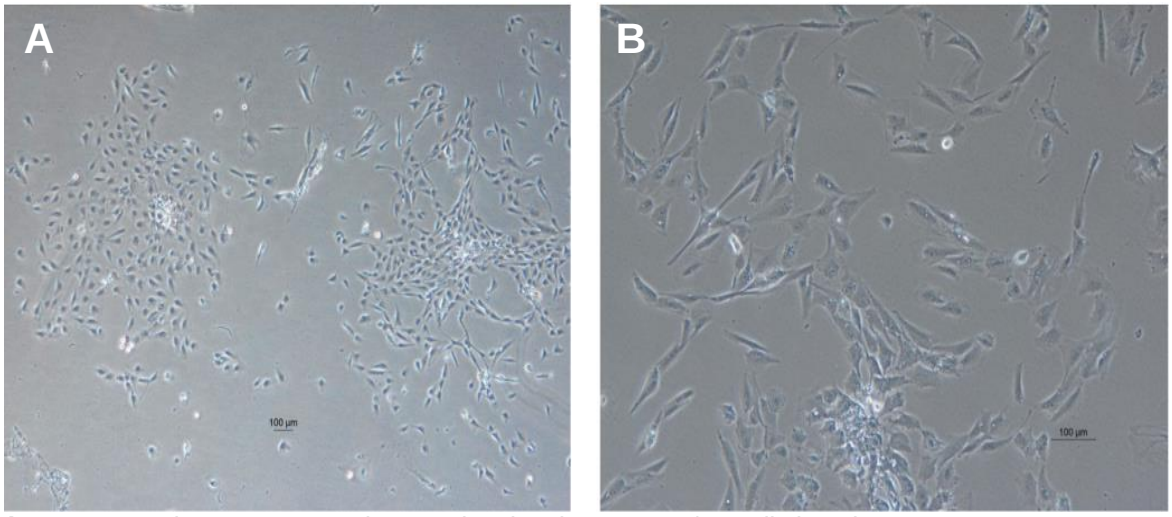

Figure 5.22: Phase-contrast Micrographs showing STS 20/11 cells in culture

Cells shown are at passage 22 and were derived from original cultures set up from STS 20/11 tissue. **A** – Images captured at x40 magnification showing the pattern of growth in loose colonies. **B**- Images captured at x100 magnification showing the mixture of fibroblast-like spindle-shaped morphology of most cells and an admixture of polygonal cells with more distinct nuclei.

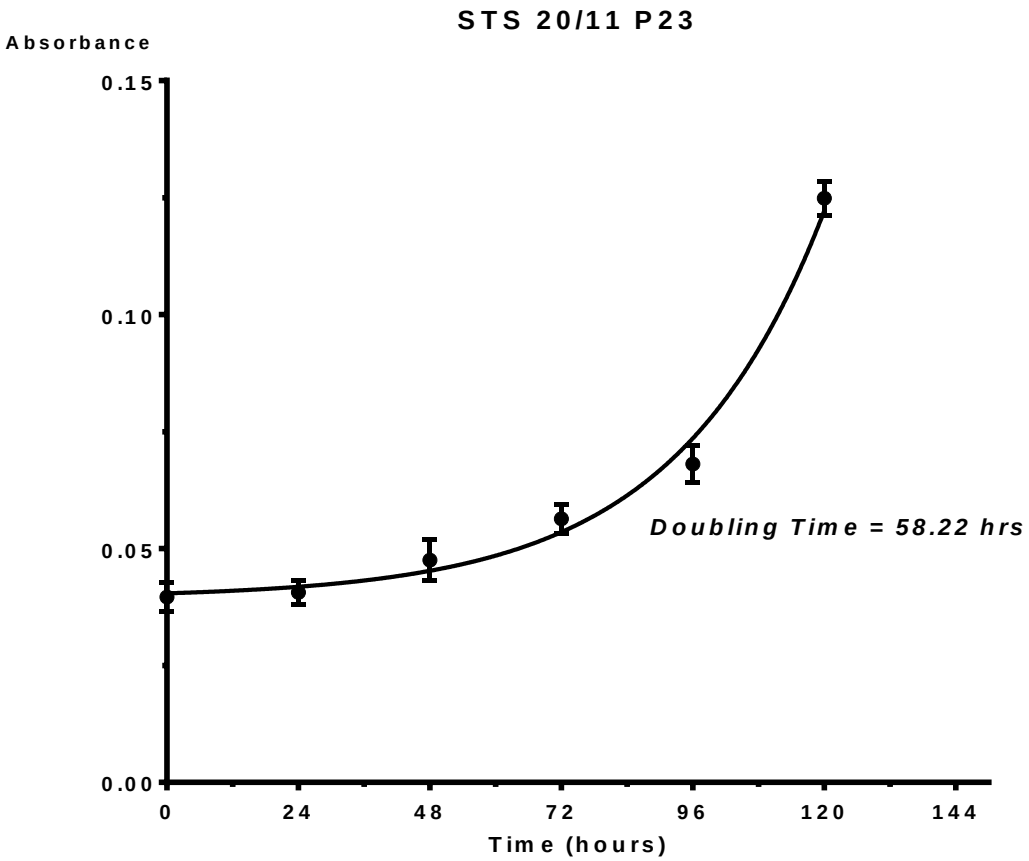

Figure 5.23: STS 20/11 Cell Growth evaluated by MTT proliferation assay

Cells were at passage 23 and were derived from original cultures set up from STS 20/11 tissue. Data is representative of experiments done in quadruplicate. Curve fitting was done using GraphPad® Prism software (version 6) and the doubling time was calculated using the Exponential Growth Equation in the same software.

5.2.9.3 Genetic Characteristics

Karyotyping

Metaphase spreads were initially obtained from STS 20/11 cells at passage 2, when chromosome counts were around 70 on average and ranged between 63 and 73. Subsequent chromosome harvests carried out around passage 15 showed a much wider range of chromosome counts of 51 – 154. However, the modal count of 69 was close to that from the previous chromosome harvest (Table 5.4).

Array CGH

Comparison of the array CGH profile of STS 20/11 cells with that of the parent tissue was carried out at the second passage. The genomic profiles showed close similarity on the majority of autosomes (Figure 5.24) with very few regions such as the short arm of chromosome 5 and long arm of chromosome 12 showing differential aberrations. Regions with very complex aberration patterns such as chromosome 20 retained near-identical breakpoints in the cultured cells (Figure 5.24).

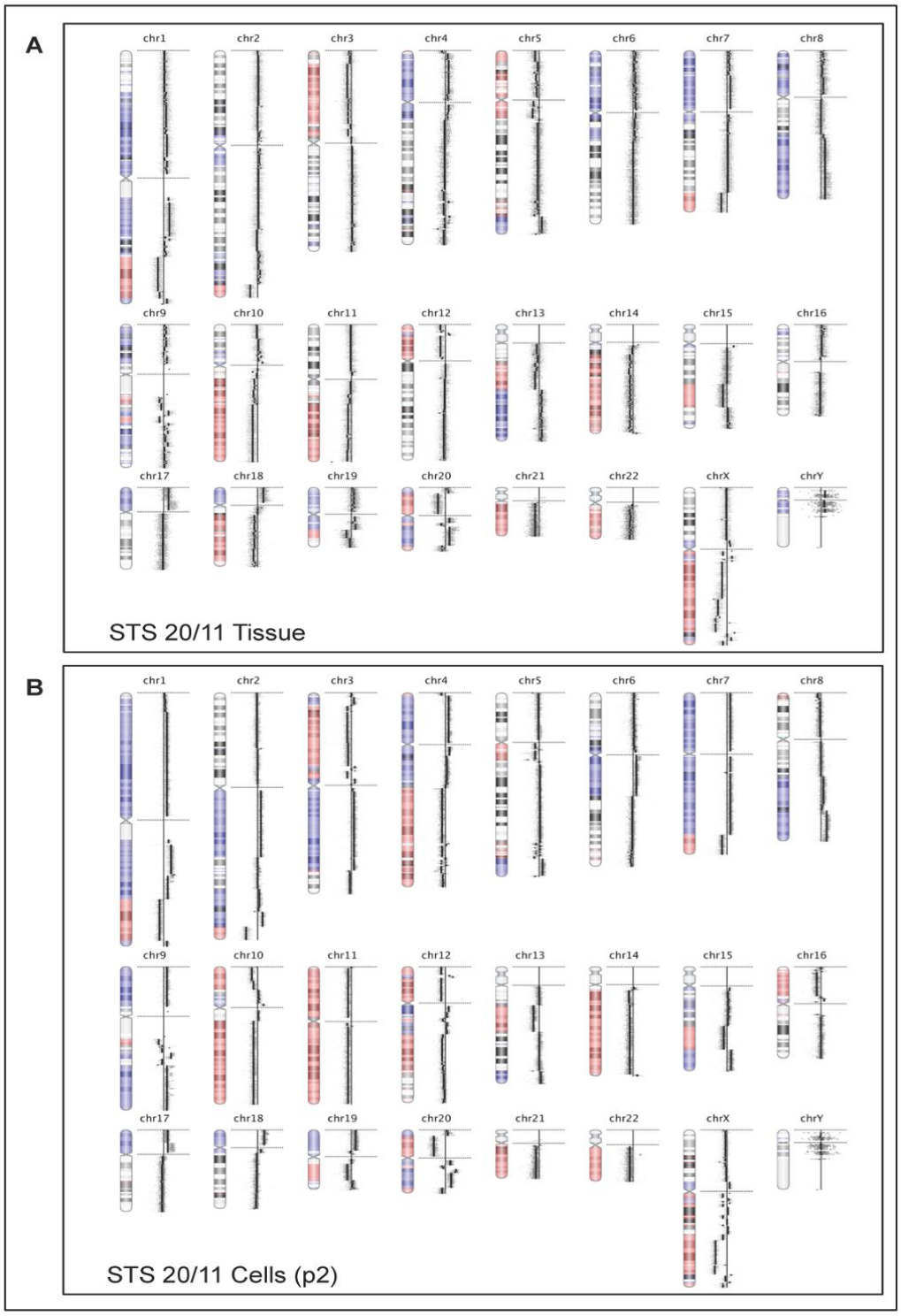

Figure 5.24: Genome View Ideograms of STS 20/11 comparing cultured cells with parent tumour tissue

Panel A – whole genome ideogram for STS 20/11 parent tumour tissue DNA showing multiple SCNA detected across the genome, compared with **Panel B** – whole genome ideogram of DNA from cells cultured from STS 20/11 tissue (at passage 2) showing multiple SCNA detected across all autosomes and similar average log ratio patterns overall.

Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Small dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the magnitude of aberration (log ratio). *All CNAs were detected using the FASST2 algorithm*

5.2.10 STS 21/11: Myxofibrosarcoma

5.2.10.1 Tumour Characteristics

Tumour tissue was kindly donated by a 73 year old gentleman who was having a high grade myxofibrosarcoma removed from his left arm. Tissue samples were obtained following surgical excision, the initial treatment modality for this 50mm tumour. Distant metastases were not evident at the time of surgery.

5.2.10.2 Primary Tissue Culture Characteristics

Cultures were set up according to normal protocol (Section 2.3.1.1). A wash was taken off the original culture flasks because viable non-adherent cells were observed in media. Adherent cell cultures were established within the wash, designated 'w_1', while the cells in the original flasks became senescent after two to three passages. Similar to STS 02/11 w_s cells, the cell form STS 21/11 w_1 cultures had a histiocyte-like polygonal morphology, but their nuclei are not as distinct (Figure 5.25B). The cells grow in distinct colonies (Figure 5.25A) and have an estimated doubling time of approximately 60 hours (Figure 5.26). They have been passaged around 35 times in the approximately 18 months of culture.

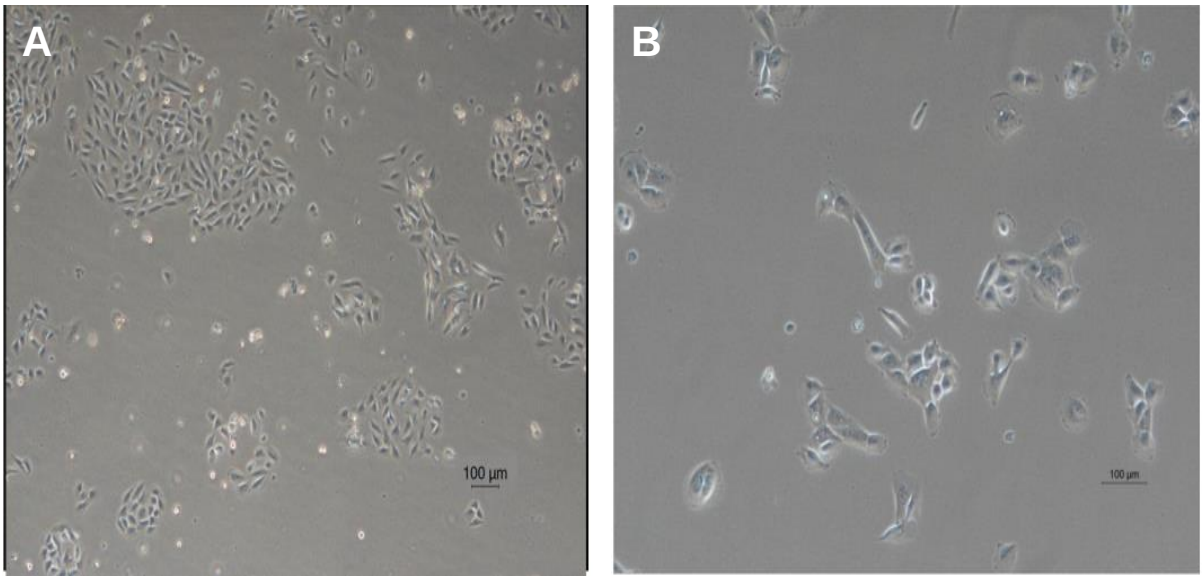

Figure 5.25: Phase-contrast Micrographs showing STS 21/11 cells in culture

Cells shown are at passage 35 and were derived from washes taken off original cultures set up from STS 21/11 tissue. **A** – Images captured at x40 magnification showing the pattern of growth in distinct colonies. **B**- Images captured at x100 magnification showing the polygonal morphology of most cells.

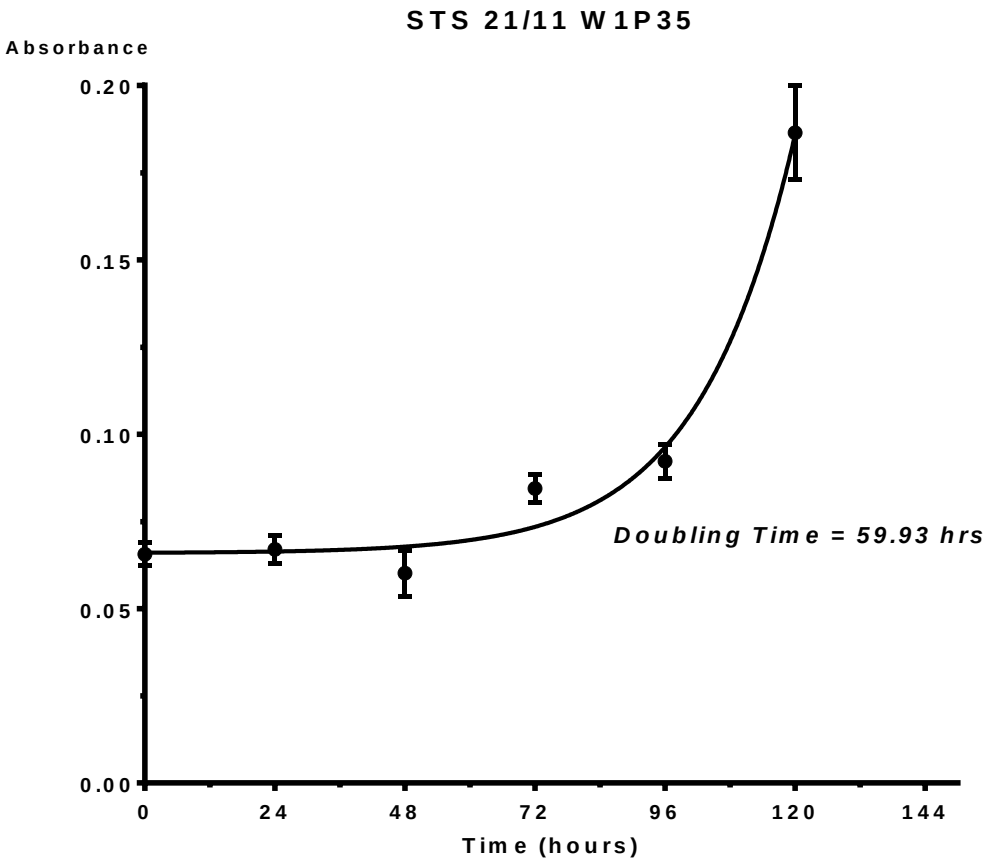

Figure 5.26: STS 21/11 cell Growth evaluated by MTT proliferation assay

Cells were at passage 35 and were derived from the first wash taken off original cultures set up from STS 21/11 tissue. Data is representative of experiments done in quadruplicate. Curve fitting was done using GraphPad® Prism software (version 6) and the doubling time was calculated using the Exponential Growth Equation in the same software.

5.2.10.3 Genetic Characteristics

Karyotyping

Metaphase spreads obtained from STS 21/11 cells at passage 28 comprised between 44 and 118 chromosomes per cell. Most cells had around 88 chromosomes (Table 5.4) and the karyotypes included numerous small, likely abnormal chromosomes (not shown).

Array CGH

DNA from STS 21/11 cells at passage 35 was analysed by array CGH and the genomic profile compared to that from the fresh frozen parent tissue (Figure 5.27). The moving average pattern of copy number aberrations was similar across the majority of chromosomes although the positive and negative amplitudes were greater in the more homogenous cultured cells than the parent tumour tissue with possible adjacent normal cell content. This was most evident in regions with complex aberrations on 4q, 5p 10q and 20q. Differential aberrations were noted on only two chromosome arms 8p and 21q.

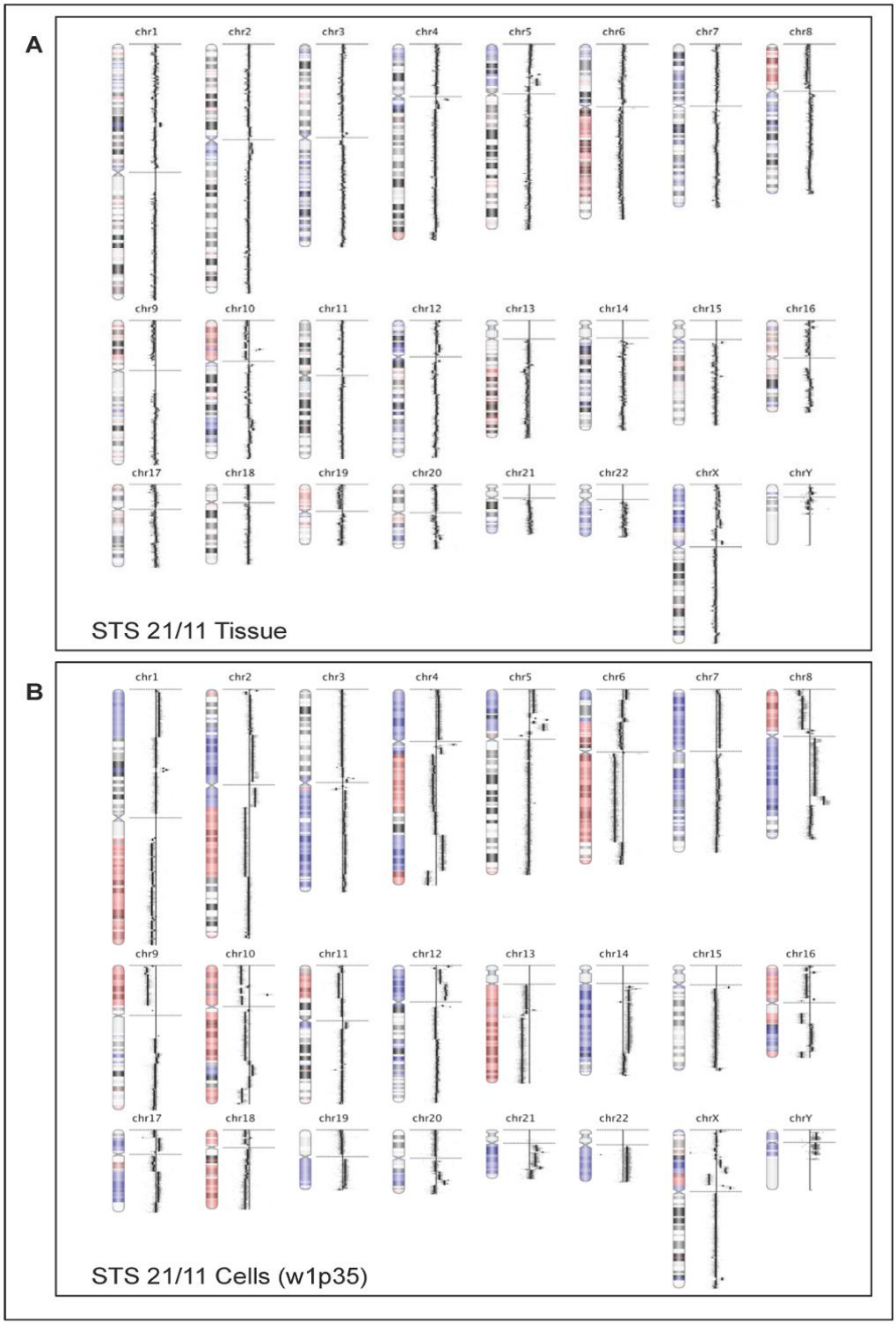

Figure 5.27: Genome View Ideograms of STS 21/11 comparing cultured cells with parent tumour tissue

Panel A – whole genome ideogram for STS 21/11 parent tumour tissue DNA showing multiple SCNA detected across the genome, compared with **Panel B** – whole genome ideogram of DNA from cells (at passage 35) derived from a wash taken off cultures of STS 21/ 11 tissues that were set up in flasks, showing multiple SCNA detected across all autosomes and similar average log ratio patterns overall.
Aberrant regions are shown as coloured shading on the chromosome (blue represents amplification and red represents deletion). Small dots represent individual probes. Black vertical lines represent mean log ratio for the corresponding region on the chromosome. Horizontal distance to the right (amplification) or left (deletion) of the vertical capped zero line represents the magnitude of aberration (log ratio). *All CNAs were detected using the FASST2 algorithm*

5.3 DISCUSSION

In this study, the success rate for establishment of confluent cell cultures from primary STS tissue was over 70% (33 out of 47 cases) and more than half of the established cultures were able to undergo four passages or more (Table 5.2). This rate is very comparable to previous studies which reported success rates of between 5 and 33% when cultures were attempted from primary solid tumours as in this study, and slightly higher success rates from tumour metastases (McBain, Weese et al. 1984, Gazdar, Kurvari et al. 1998, Nayak, Kakati et al. 2000) or xenograft-derived cultures (Dangles-Marie, Pocard et al. 2007, Kamiyama, Rauenzahn et al. 2013). Seven tumours (15% of cases) have grown as long-term cultures.

Among the 14 cases that failed to establish cultures, eight were well-differentiated or myxoid liposarcomas that had high fat and/or myxoid components relative to cell number, which reduced the likelihood of adherent culture using the manual mechanical tissue dissociation, applied in this study. Simultaneous use of the explant method of culture establishment (Mitra, Mishra et al. 2013) in some of these cases only yielded slow-growing, fibroblast-like cells that failed to reach confluence (data not shown). Application of the principle of differential attachment described by Nayak and colleagues (Nayak, Kakati et al. 2000) led to establishment of long term cultures in washes in three cases (STS 09/10, 02/11 and 21/11), while cells in the original culture flask became senescent at earlier stages. It also led to the establishment of long term cultures of two morphologically distinct but genetically related variants of STS 02/11 (Figure 5.10, 5.12 and Table 5.2). Since both variants have been exposed to otherwise identical culture conditions, and it is most likely that the variants represent separate clones present within the parent tumour rather than differential adaptation to culture conditions. Similar observations have been made in multiple cases of primary breast carcinoma tissue culture (McBain, Weese et al. 1984).

Characterisation of primary cells cultures for use as models in target validation studies is essential due to the potential for fibroblast overgrowth in early cultures (Mitra, Mishra et al. 2013) and cross contamination by other established cell lines in long term cultures (Gillet, Varma et al. 2013). Morphological characterisation may be unreliable because of the potential effects of an artificial *in vitro* microenvironment on tumour cell morphology. In most cancers therefore, detection of biomarker expression by immunochemistry or flow cytometry are commonly used for characterisation. The common biomarkers of tumour cell lineage such as CD34, cytokeratin (CK), smooth muscle actin (SMA), desmin and S100, however are not specific for many STS subtypes (Coindre 2003, Fisher 2011). For example, none of these markers is commonly expressed in UPS or myxofibrosarcomas, while others such as epithelial cell marker CK may be expressed in some LMS and synovial sarcomas. Immunohistochemistry for panels of these markers however remains relevant to diagnostic practice as an adjunct to histology is effective for the exclusion of benign and non-mesenchymal tumours; identification of certain STS subtypes and choosing appropriate molecular genetic tests for difficult subtype-specific STS diagnoses (Fisher 2011). It however requires interpretation with an appropriate level of expertise.

Since the majority of STS are characterised by complex molecular genetic profiles, the cell lines developed form these tumours were characterised using array CGH, a method that has contributed significantly to the recent WHO classification of STS (Fletcher, Bridge et al. 2013). This method is unable to detect balanced translocations that are known to drive some STS subtypes, including three of the four cases (Table 5.5) that gave inconclusive featureless array CGH profiles in both cultured cells and parent tissue. The fourth such case - STS 05/10 was diagnosed as an epitheloid angiosarcoma, an STS subtype that is described in literature as having complex genomics (Guillou and Aurias 2009). However, array CGH did not detect any SCNA in any of the three angiosarcoma cases analysed in this study. All three tumours were very haemorrhagic, as is typical of these tumours (Fletcher, Bridge et al. 2013) and macroscopically sampled thus raising the possibility that their percentage tumour cell content was below the threshold (around 25%) that is readily detectable by array CGH (Neill, Torchia et al. 2010). In three other cases of STS subtypes known to have complex genomics that were confirmed by array CGH such as the malignant peripheral nerve sheath tumour designated STS 01/11, the absence of SCNA in the cultured cells (Figure 5.4) implied that were not representative of their parent tumours and were probably fibroblasts that subsequently became senescent before the tenth passage when they presumably reached their Hayflick's number (Macieira-Coelho 1998).

STR profiling, the currently recommended standard for cell line identification (American Type Culture Collection Standards Development Organization Workgroup 2010) was used to resolve issues of potential cross contamination and reliably confirm all seven long term cultures in this study as separate entities. Subsequent characterisation on the basis of genomic copy number profile comparison showed overall similarity in log ratio patterns to the corresponding parent tumours in all seven long-term cell cultures when thresholding of the aberration detection algorithm are taken into account (discussed in section 3.3). A number of real SCNA differences were noted. In line with the clonal evolution model of cancer, inherent genetic instability and rapid proliferation of tumour cells is expected to result in heterogeneity of tumour cell clones, which would be selected for in culture but be a part of the whole in the parent tissue (Anderson, Lutz et al. 2011, Greaves and Maley 2012). With prolonged *in vitro* cell culture, these factors would also result in accumulation of genomic copy number or structural karyotypic aberrations, most of which are functionally neutral, i.e. 'passengers' (Greaves and Maley 2012, Gillet, Varma et al. 2013).

Overall however, these differences are few when compared to the genomic regions that show similar SCNA with near-identical breakpoints even after 40 passages in culture (Figure 5.13). In addition, recent large-scale studies evaluating the relevance of established cell lines in various cancers have shown that 'driver' genomic aberrations are nearly always retained in well-established commercial cell lines and *vice versa* (Beroukhim, Mermel et al. 2010, Gazdar, Gao et al. 2010, Barretina, Caponigro et al. 2012). Moreover, early passages of all the long term cultures that presumably more bear a closer genomic and phenotypic resemblance to parent tumour have been banked and can be used to confirm promising finding in target validation studies.

All seven long term cell cultures established have remained proliferative in culture for between one and three years, and been passaged at least 25 times each with doubling times that are comparable

to those of commercial sarcoma cell lines such as SKLMS1 (leiomyosarcoma), U2OS (osteosarcoma) and other primary cancer cell lines (McBain, Weese et al. 1984) , thus reflecting their possession of the cancer hall marks of increased proliferation and survival (Hanahan and Weinberg 2011). All seven showed abnormal chromosome numbers and in some cases, significant changes in chromosome numbers with increased time in culture, further reflecting their genomic instability. One of the cultures, STS 06/11 was also shown to have migratory and invasive properties (discussed in detail in Chapter 6).

Among the seven tumours that established long term cultures, three were undifferentiated pleomorphic sarcomas, representing half of the cases that were obtained of this characteristically aggressive STS subtype (Fletcher, Bridge et al. 2013). The other STS subtypes that formed long term cultures were also of a high grade. This suggests that overall, high grade and aggressive clinical course in STS correlates with amenability to *in vitro* growth, as was observed in primary cultures of breast (Gazdar, Kurvari et al. 1998) and colorectal cancer tissue (McBain, Weese et al. 1984). However, other aggressive tumour subtypes such as angiosarcoma and myxofibrosarcoma with high grade tumours represented in this study did not show any such correlation. Again, this may explained by other factors such the result of the physical and biologic properties of these tumours as observed in the angiosarcomas, where high blood cell content of samples makes reliable macroscopic sampling of tumour cell-rich areas difficult.

Three of the cell lines represent undifferentiated pleomorphic sarcomas with very different morphology and chromosome numbers; the leiomyosarcoma-derived cultures comprise two morphological variants of the same tumour, and one of the de-differentiated liposarcoma cultures was derived from a patient who received neo-adjuvant radiotherapy with partial response and probably represents a radio-resistant clone of cells present within that tumour. Given this range of features, these seven long term cultures are potential much-needed additions to the number of cell lines of STS with complex genomics available that could comprise panels for drug testing akin to the NCI-60 panel that includes only more-prevalent cancers (Shoemaker 2006, Taylor, Barretina et al. 2011). In addition, the previously-irradiated and therefore likely radio-resistant STS cell line, STS 20/11 may be utilised for the study of tumour response to radiation, either alone or in combination with sensitising agents, as demonstrated with PARP inhibitors in Ewings' sarcoma cell lines (Garnett, Edelman et al. 2012).

The results of primary tissue culture in this study show the potential for fairly reliable establishment of short- and long term cell cultures from STS tissue that can serve as a source of tumour material for reliable translational research. Even though time constraints, as well as karyotypic complexity did not permit the completion of cytogenetic analysis of the cultured cells within this PhD project, molecular genetics was used to establish that the vast majority of cultured cells lines in this study were representative of their parent tumours. When combined with their rapid proliferation and abundance, the long term cell lines represent an excellent model *in vitro* validation of genomic and transcriptomic targets in STS. The combination for banking of early passages of these characterised cell lines as well as the potential to develop prospective short term cultures from other tumours of the

same STS subtypes will serve to mitigate some of the concerns that have been raised with existing cell lines.

Full characterisation of the malignant phenotype of the long term cultures using *in vitro* assays such as soft agar colony-forming; migration and invasion; and telomere length assays would be useful. In addition, assessment of their in-vivo tumour xenograft-forming ability could potentially lead to the development of *in vivo* models suitable for robust preclinical target validation. Although time and cost constraints meant that these assays could not be carried out within this PhD study, they serve as areas for additional work in the near future.

CHAPTER SIX

RESULTS: AMPLIFICATION OF *TRIO* IS A POSSIBLE TARGET IN UPS

6.1 INTRODUCTION

Tumour metastasis remains an important cause of cancer mortality, accounting for up to 90% of deaths from solid tumours (Gupta and Massague 2006). In order to metastasize, cancer cells must undergo a complex series of processes including invasion of, and migration through surrounding tissues; intravasation and survival within the bloodstream; extravasation at distant sites; and subsequent colonisation (Steeg 2006). Elucidation of molecular mechanisms that are responsible for these processes in specific cancers is therefore required for development of effective novel targeted therapies (Brabletz 2012).

The initial steps in metastasis - tumour cell migration and invasion require cell motility, which is regulated at the molecular level by the Rho subfamily of small guanine nucleotide-binding proteins (g-proteins) called rhoGTPases. As illustrated in Figure 6.1, specific rhoGTPases promote the formation of actin-containing structures that cells require for migration (Gupta and Massague 2006) including lamellopodia and membrane ruffles (mediated by Rac1) (Ridley, Paterson et al. 1992), filopodia (mediated by Cdc42) (Kozma, Ahmed et al. 1995), and actomyosin stress fibres (mediated by RhoA) (Ridley and Hall 1992).

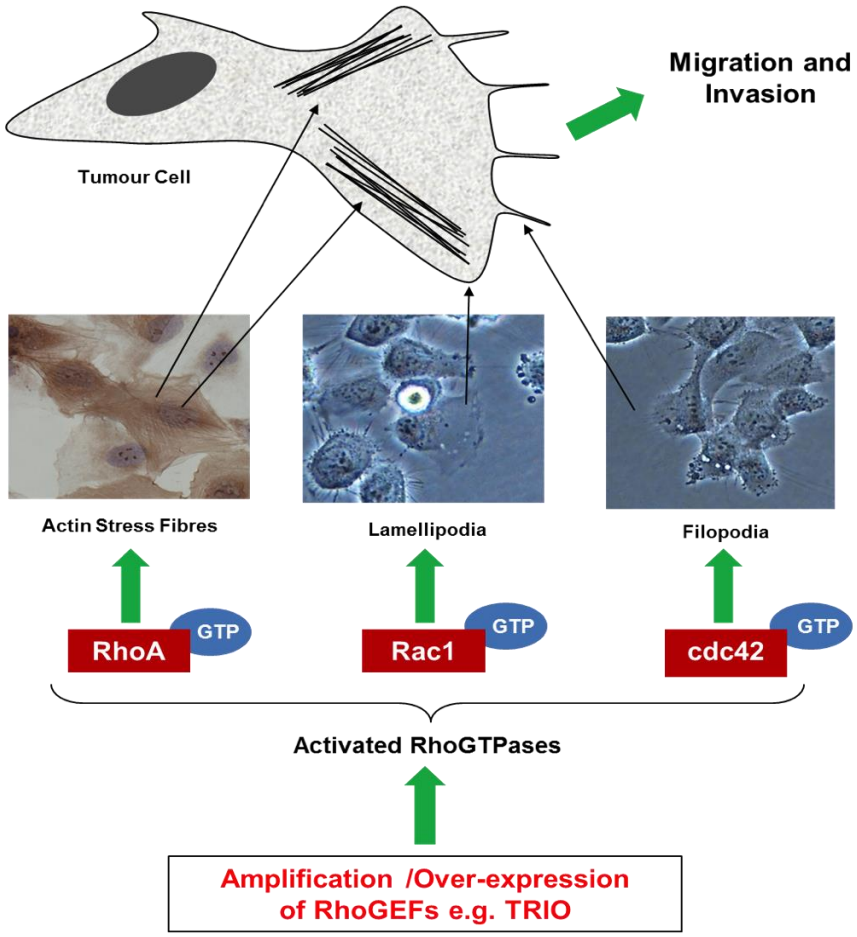

Figure 6.1: Simplified Representation of the Role of RhoGTPases in Tumour Cell Migration and Invasion.

Activation of Specific RhoGTPases (red highlights) leads to the formation of cytoskeletal structures that are used for tumour cell migration and invasion.

More recently, rhoGTPases have also been shown to be involved in other cancer-relevant roles such as cell cycle progression; DNA transcription; cell-cell and cell-matrix interactions (Keely, Westwick et al. 1997, Chan, Coniglio et al. 2005, Jaffe and Hall 2005). While activating mutations of rhoGTPase genes have rarely been reported in cancer (Lazer and Katzav 2011), increased levels of Rac1 and RhoA activity and/or expression have been shown to correlate with tumour progression in various cancers (Fritz, Just et al. 1999, Abraham, Kuriakose et al. 2001, Fritz, Brachetti et al. 2002, Kamai, Yamanishi et al. 2004, Pan, Bi et al. 2004).

RhoGTPase activation occurs via a number of mechanisms, important among which are a group of proteins referred to as Rho Guanine-nucleotide Exchange Factors (rhoGEF). By specific interaction with rhoGTPase proteins in their inactive (GDP–bound) state, they cause a conformational change that results in GDP dissociation and its subsequent replacement by the more physiologically-abundant GTP, thus switching the g-protein to its active state (Bos, Rehmann et al. 2007). This is illustrated schematically in Figure 6.2.

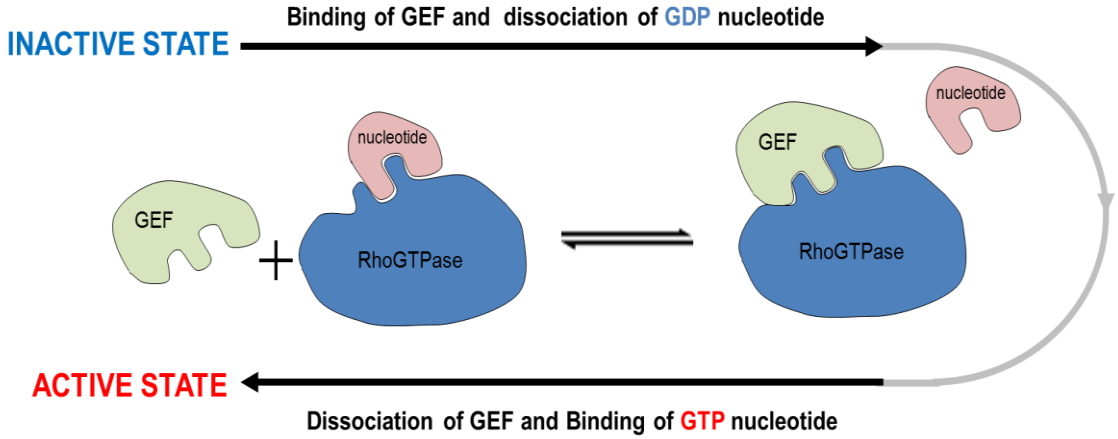

Figure 6.2: Mechanism of GEF Activation of RhoGTPases

Binding of the Rho Guanine-nucleotide Exchange Factors (GEF) to the inactive (GDP-bound) rhoGTPase leads to conformational change and dissociation of the guanidine di-phosphate (GDP) nucleotide. This interaction is transient however, and the GEF soon dissociates and is replaced by the more physiologically abundant guanidine tri-phosphate (GTP) nucleotide, switching the rhoGTPase molecule to its active state

The first rhoGEF identified in mammalian cells, Dbl was isolated from B-cell lymphoma cDNA on the basis of its oncogenic (transforming) activity in NIH3T3 fibroblast cells (Hart, Eva et al. 1991). Over 70 different human rhoGEF proteins have been identified to date, the majority of which possess a Dbl-conserved domain (DH) (Lazer and Katzav 2011) that was shown to be required for GEF activity (Hart, Eva et al. 1994). Some of these, including Ect2 and Ost were also initially identified as oncogenes capable of transforming fibroblasts (Miki, Smith et al. 1993, Yamanaka, Blumenthal et al. 2001) and subsequently found to possess the characteristic DH- domains and GEF activity (Lazer and Katzav 2011).

In addition to their putative oncogenic properties, abnormally increased levels of RhoGEFs expression have also been demonstrated in various cancers and the sum total of this evidence has led to the prevailing theory among researchers that activation of rhoGEFs causing dysregulation of

rhoGTPases and actin cytoskeleton organisation is responsible for tumour progression and metastases (Sahai and Marshall 2002, Lazer and Katzav 2011). This emerging target has thus been the focus of recent anti-cancer drug development with encouraging preclinical results (Lu, Chan et al. 2009, Routhier, Astuccio et al. 2010, Lazer and Katzav 2011, Barrio-Real and Kazanietz 2012).

One rhoGEF, Trio is unusual because it has not one, but two putative DH – domains with specific GEF activity termed GEF1 and GEF2 (Seipel, Medley et al. 1999). The N-terminal GEF1 interacts specifically with RhoG and Rac1, via which it mediates mesenchymal-type cell motility and axon guidance, respectively (van Rijssel, Hoogenboezem et al. 2012) while the C-terminal GEF2 activates RhoA and thus regulates a different form of cell motility known as amoeboid cell movement (Sanz-Moreno, Gadea et al. 2008). Its encoding gene, *TRIO* is located on the short arm of chromosome 5 and was initially identified in adult T-cell leukaemia cells as its truncated cDNA isoform termed Tgat, which encodes only the RhoA-specific GEF domain. Like a number of other rhoGEFs, Tgat was found to be a putative oncogene, able to induce transformation in NIH3T3 cells with subsequent tumour formation in nude mice via activation of RhoA (Yoshizuka, Moriuchi et al. 2004).

Copy number gains affecting the *TRIO* locus have been described as frequent aberrations in various cancers including soft tissue sarcoma (Adamowicz, Radlwimmer et al. 2006), bladder cancer (Zheng, Simon et al. 2004), cervical cancer (Kloth, Oosting et al. 2007, Ng, Winder et al. 2007), oral cancer (Baldwin, Garnis et al. 2005), squamous oesophageal cancer (Chattopadhyay, Singh et al. 2010) and small cell lung cancer (Coe, Henderson et al. 2005). Gene expression studies in brain tumours found that *TRIO* was expressed at higher levels in aggressive glioblastoma multiforme when compared with low-grade gliomas (Salhia, Tran et al. 2008) and silencing its expression led to reduced proliferation of glioblastoma cells in culture, as well as impairment of their invasive capabilities *in vitro* and *ex-vivo* (Salhia, Tran et al. 2008, Kwiatkowska, Didier et al. 2012). Furthermore, protein levels of Trio and other rhoGEFs have been correlated with poor patient outcomes in both glioblastoma (Salhia, Tran et al. 2008) and breast cancer (Lane, Martin et al. 2008), supporting a possible role for *TRIO* in tumour progression.

In view of this accumulating evidence, inhibitors directed specifically against the Trio rhoGEF domains have been under development. The first of these include TRIPα and other aptamer-derived peptides, which were shown to specifically inhibit the GEF2 domain and prevent RhoA activation (Schmidt, Diriong et al. 2002). They also abrogated oncogenic effects of Tgat in cell lines and tumour xenografts (Bouquier, Fromont et al. 2009). More recently, a small molecule inhibitor named ITX3 was shown to specifically inhibit the GEF1 domain, blocking the activation of RhoG and Rac1. ITX3 in normal mammalian cells showed comparable effects to complete Trio knockdown (using siRNA) and the downstream Rac1-specific inhibitor NSC23766, but is believed to avoid their off-target effects due to its specificity for the GEF1 domain (Bouquier, Vignal et al. 2009). Its effects on human tumour cells expressing Trio are however, yet to be examined (Lazer and Katzav 2011).

Using array CGH, *TRIO* was identified as a candidate gene among UPS, where it was amplified in around 60% of tumours examined in this study (Figure 6.3 and Section 4.7.3.2). FISH, utilising custom probes on 3 primary cell cultures derived from UPS as well as FFPE tumour sections was used to visualise the copy numbers of the *TRIO* locus. Immunohistochemistry was used to examine the expression of Trio rhoGEFs protein in STS samples. Further, the effects of ITX3 treatment on proliferation, migration and invasion were examined. The preliminary results of these studies are presented in this chapter.

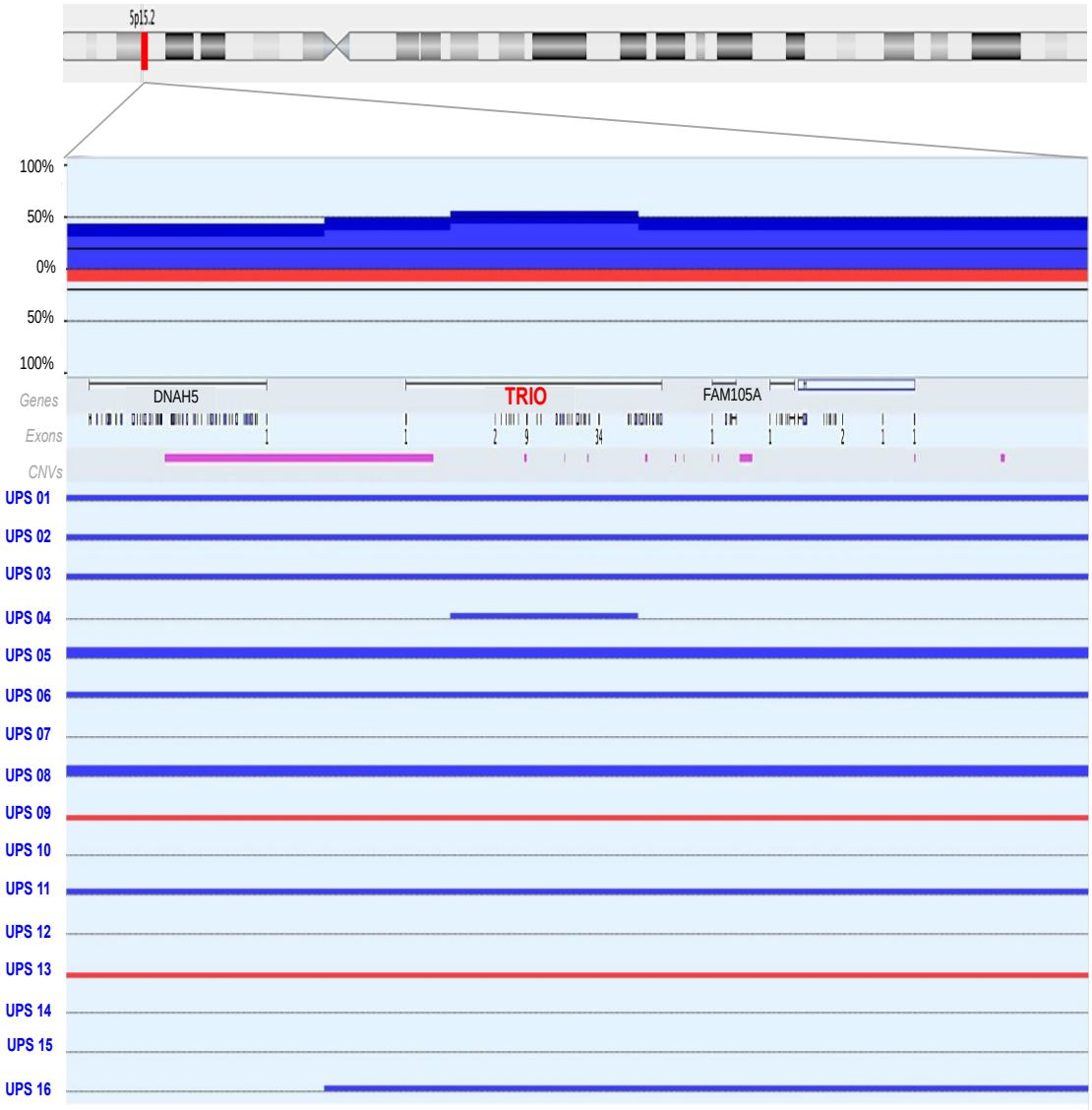

Figure 6.3: Frequency plot of SCNA affecting the *TRIO* gene locus among 16 UPS cases.

Top shows the chromosomal region and corresponding frequency plot of aberrations for all 16 cases, plotted as percentages along the y-axis. Blue shading above zero line represents amplification frequency and red shading below zero line represents deletion frequency. Known CNV region is shown as the purple track. Horizontal lines below represent individual samples and identities are shown on the left. Grey segments have normal copy number, blue segments represent gains and red segments represent losses. Vertical lines show exact alignment of SCNA breakpoints.

All CNAs are detected using the FASST2 algorithm

6.2 RESULTS

6.2.1 Visual Confirmation of *TRIO* Amplification by FISH

Fluorescence *in situ* Hybridisation (FISH) was used to visualise the copy number status of the *TRIO* gene locus at 5p15.2. No commercial FISH probes specific for that gene locus were available, so custom probes were prepared from specific human DNA sequences cloned into Bacterial Artificial Chromosomes (BAC clones). BAC clones that mapped to the *TRIO* gene locus were identified using the University of California at Santa Cruz (UCSC) genome browser (Figure 6.4) and chosen from well-annotated libraries including the RPCI human BAC library 11 (RP11) and the CalTech human BAC library D (CTD).

The *TRIO* gene consists of 30 exons, with 24 known splice variants. It spans a region of around 400 kb in length and hence no single BAC clone insert (average size < 200kb) reliably mapped to the entire gene. A pair of clones including RP11-586K23 and CTD-2505B6 that mapped to the 5' and 3' ends of the gene respectively, with a small overlap in the middle of the gene was therefore chosen (Figure 6.4). Nick translation was used to label these BAC clones inserts with a SpectrumGreen® fluorophore for use as FISH probes (BAC probes) as described in Methods Section 2.3.8.2.

In order to distinguish target locus-specific amplification from aneuploidy, chromosome enumeration probes (CEP probes) for the chromosome on which the target gene lies are commonly included simultaneously for FISH experiments. These probes hybridise to specific repetitive alphoid DNA sequences in the centromeric regions of chromosomes. Chromosome 5 on which the *TRIO* gene locus lies is however known to possess high alphoid DNA sequence homology with chromosomes 1 and 19 (Puechberty, Laurent et al. 1999), which results in cross hybridisation of their CEP probes.

To provide an idea of the overall ploidy status of the genome therefore, an alternative chromosome (chromosome 3) with previously optimised CEP probes was chosen for enumeration. In view of the overall karyotypic complexity of these tumours as demonstrated by preliminary cytogenetic studies shown in Section 5.2.3.4, aneuploidy of chromosome 3 was regarded to be no better and no worse than any other chromosome for which specific centromeric enumeration would be possible. Commercial SpectrumOrange®-labelled (red) CEP3 probes and SpectrumGreen®-labelled (green) BAC probes were therefore simultaneously hybridised to STS tumour material as described in Methods Section 2.3.8.4.

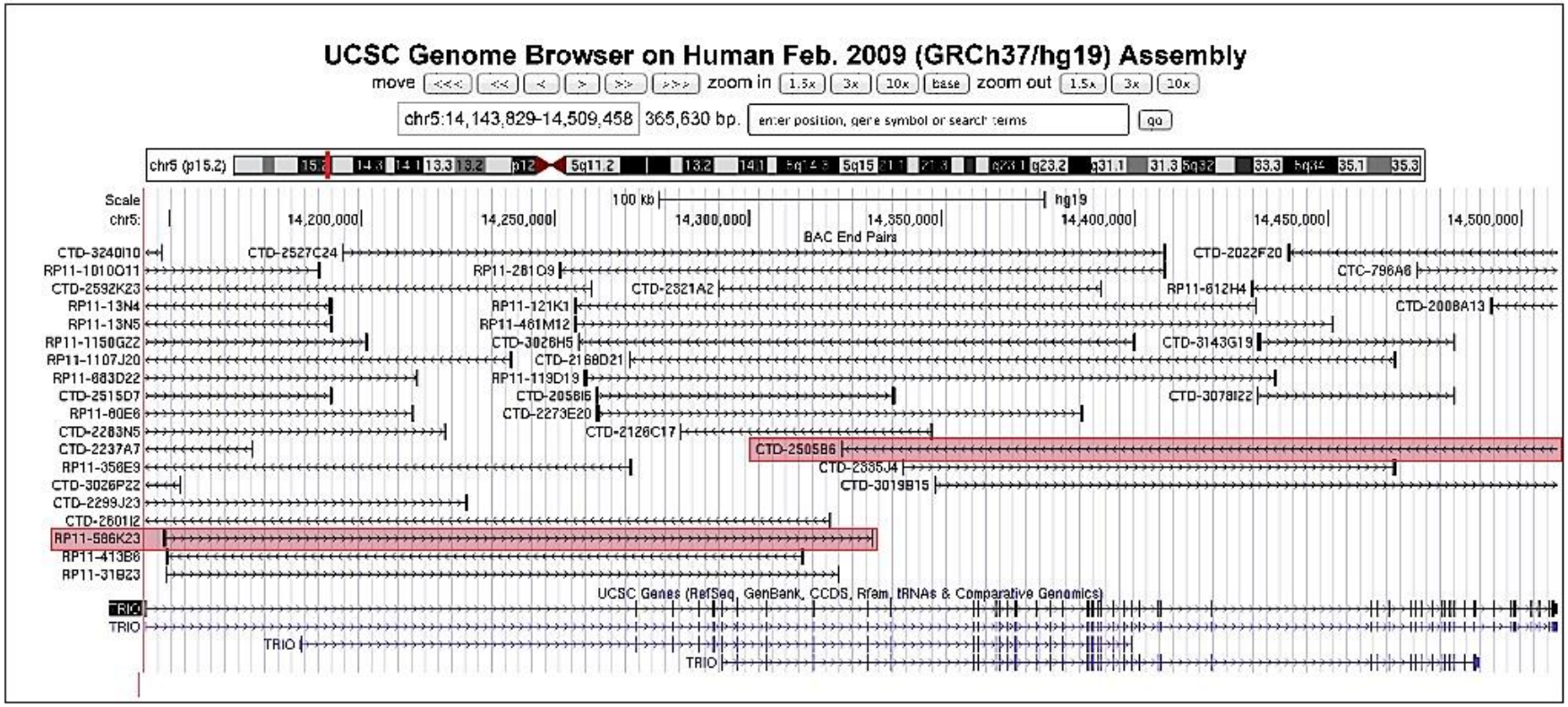

Figure 6.4: UCSC Genome Browser Mapping of BAC clones to *TRIO* gene locus

Top shows the base position of the genome browser as well as its chromosomal location on chromosome 5 (marked in red). Bottom panel shows the 5' to 3' mapping of the *TRIO* gene to that genomic locus (blue arrows) with the blue vertical lines representing the exon positions of known transcripts and blue horizontal lines showing the introns. BAC clones that map to the same locus are shown in the middle as labelled black horizontal lines with arrows that show their 5' to 3' mapping and black horizontal lines showing start and end points. The two BAC clones that were chosen map to the 5' and 3' ends of the locus of the largest known transcript with an overlap that involves exon 11, and are highlighted in red.

6.2.1.1 *Control Experiments*

Initial control experiments were carried out to determine the hybridisation efficiency and chromosomal localisation of the BAC probes. Varying amounts (100 – 300ng) of each BAC probe were hybridised to interphase nuclei and metaphase spreads prepared from normal leucocytes (Figure 6.5). Both showed good signal intensity when 300ng of BAC probe was used. In all 10 metaphase chromosome spreads examined, two sets of paired green signals were observed. They were located on two long sub-metacentric (Denver Group B) chromosomes, just short of the telomeric ends of their short arms, equivalent to the locus of *TRIO* at 5p15.2 (Figure 6.5 A, B and C).

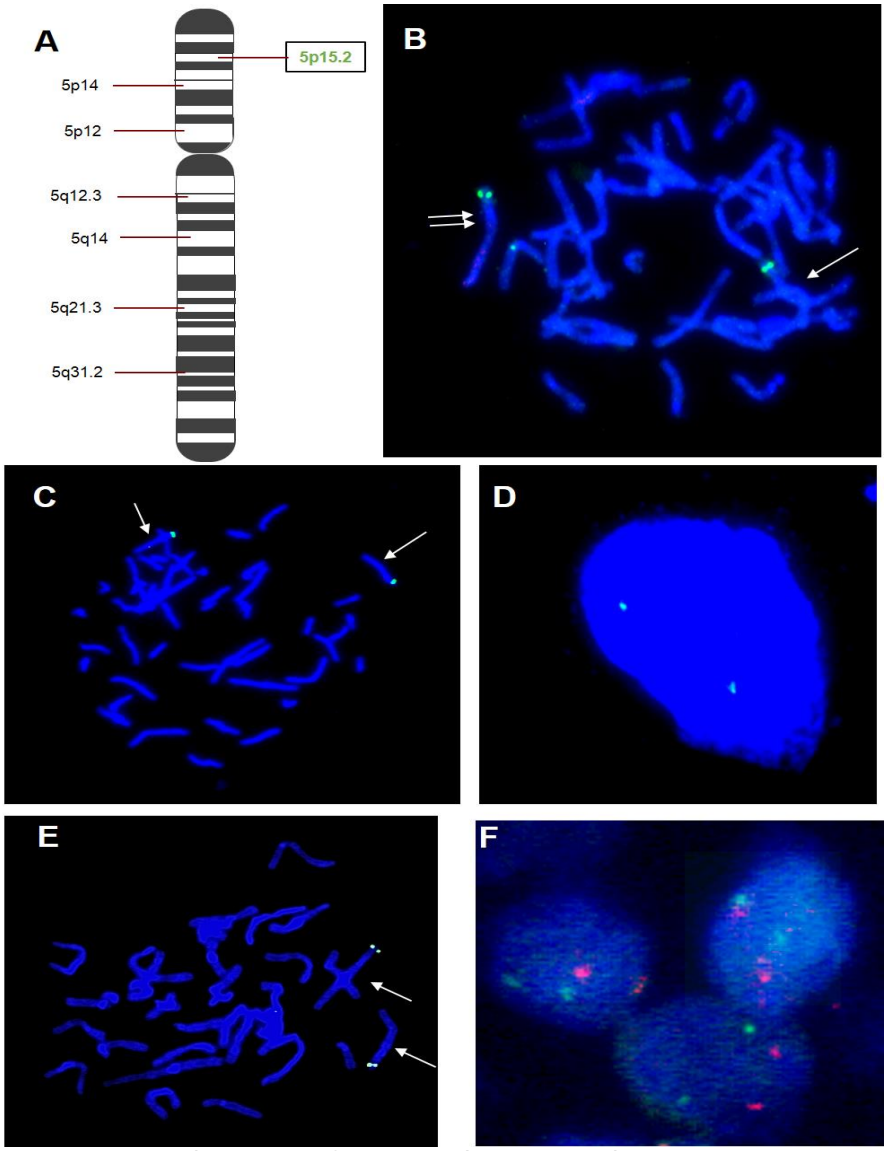

Figure 6.5: Representative Images from control FISH experiments

A: Schematic roughly-scaled representation of the Chromosome 5 banding with 5p15.2 that contains the *TRIO* gene locus highlighted in green. **B** – Metaphase spread from normal leucocyte showing sub-telomeric localisation of paired green signals (RP11586K23 BAC probe) on short arms of two long chromosomes (white arrows). The sub-metacentric structure of these chromosomes identifies them as Denver Group B (centromeres show relatively high blue DAPI staining – double white arrows) **C** – Identical localisation of CTD-2505B6 BAC probe (green signals) on chromosomes (white arrows) in a normal metaphase spread
D – Interphase nucleus from the same experiment (as panel C), showing two distinct green signals. **E** – Combination of both BACs in one probe (green signals) showing identical localisation as single BACs without doubling of signals. **F** – FFPE section of normal tonsil showing two green signals (combined BAC probe) and two red signals (CEP3) in most nuclei (blue DAPI counterstain).

Since FFPE samples provide a relatively difficult target for hybridisation, the possibility of using a combination of the BAC probes in order to increase the likelihood of success was explored. Equal amounts of both BAC probes (200ng each) were therefore co-precipitated and used for FISH experiments on normal metaphase chromosome spreads. The combined BAC clone probe (cBAC) retained the locus specificity of single probes and no increase in green background signal was detected (Figure 6.5 E). Subsequent application of the cBAC probe to 4µm-thick sections of normal tonsil tissue showed good signal intensity.

In addition to background auto-fluorescence, one of the technical challenges faced while performing FISH on FFPE sections is the three-dimensional nature of the tissue architecture. All the probe signals present in one nucleus are rarely on a single horizontal focal plane and fine adjustments of the vertical distance are needed in order to accurately enumerate all the signals present. This makes scoring of nuclei and capturing representative images much more problematic compared with interphase nuclei from cultured cells.

Cut-off values for green (cBAC) and red (CEP3) probes were derived from three control experiments on FFPE sections from normal tonsil tissue. One hundred nuclei each from the three controls were evaluated for green and red signals. The majority of nuclei showed the expected ratio of 2 green and red signals each (Figure 6.5 F and Table 6.1) and cut-off values were calculated as Mean frequency + 2 times standard deviation of signal observed in the three controls (Table 6.1). Amplification of the *TRIO* locus was confirmed when the aberrant signal ratios (cBACs > CEP3) occurred with a frequency above the cut-off value of 18.1% of nuclei.

Table 6.1: Cut-off Values for FISH probes from Interphase Nuclei in Normal Tissue

Sample (100 nuclei)	Signal Ratios			Green (cBAC) signals			Red (CEP3) signals		
	cBACs< CEP3	cBACs = CEP3	cBACs > CEP3	<2	2	>2	<2	2	>2
Tonsil 1	10	88	2	6	92	2	8	90	2
Tonsil 2	8	79	13	6	94	0	13	84	3
Tonsil 3	4	90	6	2	95	3	10	85	5
Mean	*7*	*86*	*7*	*5*	*94*	*2*	*10*	*86*	*3*
S.D	*3.05*	*5.86*	*5.57*	*2.31*	*1.53*	*1.53*	*2.52*	*3.21*	*1.53*
Cut-off (Mean + 2 S.D.)	**16%**	**97.4%**	**18.1%**	**9.3%**	**96.7%**	**4.7%**	**15.4%**	**92.8%**	**6.4%**

Amplification of TRIO gene locus was considered positive if the percentage of cells showing a cBAC > CEP3 signals ≥18.1% (highlighted in red).

6.2.1.2 FFPE Sections and Metaphase Spreads

Interphase FISH was performed on 4µm sections of FFPE tumour samples from three UPS cases – UPS 02, UPS 05 and UPS 06. In all three cases, amplification affecting the *TRIO* locus had been detected by array CGH in fresh tissue samples (Figure 6.3 and Table 6.2) and long-term primary cell cultures were derived (as described in Chapter 5) for use in subsequent functional studies. Both array CGH and FISH analyses were performed on the cultured cells in order to confirm that they retained the relevant targeted amplification. A summary of the results is shown in Table 6.2.

Table 6.2: Comparison of Copy Number Status of *TRIO* gene locus by Array CGH and FISH in STS tissue and Cultured Cells

STS Case	Sample Type	Array CGH Status of *TRIO* locus	% of cells with CBAC > CEP 3 signals (n = 100)	Median (range) of CBAC signals	Median (range) of CEP 3 signals	FISH Status of *TRIO* locus
UPS 02	Tumour Tissue	Amplified	93	5 (2 – 10)	2 (1 – 4)	Amplified
	Cultured Cells (STS 14/10)	Amplified	100	7 (6 – 8)	4 (3 – 5)	Amplified
UPS 05	Tumour Tissue	Highly Amplified	100	+++	3 (2 – 4)	Highly Amplified
	Cultured Cells (STS 06/11)	Highly Amplified	100	>40	16 (13 – 23)	Highly Amplified
UPS 06	Tumour Tissue	Amplified	91	5 (2 – 12)	2 (1 – 6)	Amplified
	Cultured Cells (STS 09/11)	Normal	47	3 (2 – 4)	2 (1 – 4)	Amplified

cBAC – combined BAC probes for TRIO locus
CEP3 – chromosome 3 enumeration probe
***+++** - Green cBAC signals that were so numerous and clumped together that they were indistinct and difficult to enumerate (see figure 6.5A - C)*

Array CGH results suggested that the *TRIO* locus amplification was retained in cultured cells derived from UPS 02 and UPS 05, but not UPS 06 (Table 6.2). Unlike aCGH however, visualisation of this locus using FISH not only confirmed the amplification in FFPE sections of UPS 06 parent tissue (Figure 6.6A), but also in corresponding cultured STS 09/11 cells (Figure 6.6B and C). Of one hundred STS 09/11 cells examined (both interphase nuclei and metaphase chromosome spreads), the majority (approximately 70%) had more than two green *TRIO* locus signals (Figure 6.6B and C).

Even though there was a similar range of green (cBAC) and red (CEP3) signals, around half of the cell population samples showed higher than normal green to red signal ratios (Table 6.2), exceeding the cut-off of 18.1% determined in normal control tissue (Section 6.2.1.1). The overall picture that suggests a mosaic amplification of the *TRIO* locus in STS 09/11 cells at such a low overall magnitude that it could have been missed by array CGH (false negative).

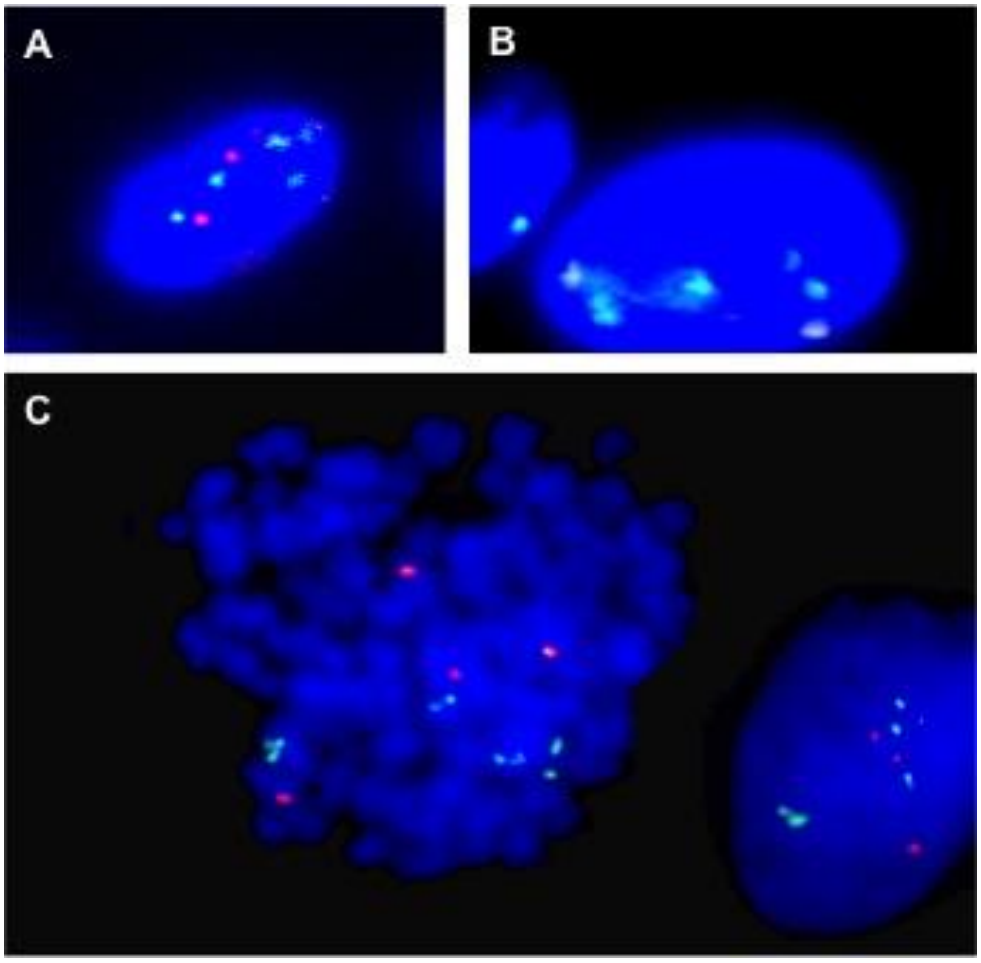

Figure 6.6: Two-colour Fluorescence *in situ* Hybridisation (FISH) for confirmation of *TRIO* amplification on FFPE sections from UPS 06 and corresponding STS 09/11 cells.

The cBAC probe for the *TRIO* locus (green signals) and commercial CEP3 (red signals) were applied simultaneously and the nuclei and chromosomes were counter-stained with DAPI (in blue), except in panel C where CEP3 was not used. **A** - FFPE section of UPS 06 captured in a single focal plane showing 5 green *TRIO* locus signals in an interphase nucleus and 2 red CEP3 signals. **B** – Interphase nucleus of STS 09/11 cells that were derived from long term culture of UPS 06 tissue showing multiple, large green signals (*TRIO amplicons)* as seen in parent tissue (in Panel A). **C** – Interphase nucleus and adjacent metaphase chromosome spread of STS 09/11 cells showing single copy *TRIO* locus amplification of around four green cBAC signals each with 3 and 4 red CEP3 signals, respectively.
All images were captured using x100 objective

Parent tissue from UPS 05 showed a high-level amplification affecting the *TRIO* locus (mean log ratio of 3.1 for probes in the SCNA region) that in theory suggests the presence of around eighteen copies on average of the *TRIO* locus in each tumour cell. In concordance, FISH on UPS 05 FFPE sections showed green (*TRIO* locus) signals that were so numerous that they were indistinct and exact counts were difficult (Figure 6.7A, B and C) while red CEP3 signals however, suggested trisomy 3 in the majority of cells (Table 6.2).

Subsequent experiments on the corresponding cultured cells (STS 06/11) showed that this high level amplification was retained with the green signals more distinct and numbering >40 per interphase nucleus (Figure 6.7D and Table 6.2). Examination of metaphase spreads showed the pattern of amplification as numerous single copy amplifications dispersed across various chromosomes, as well as some clusters present on abnormal-looking chromosomes (Figure 6.5E).

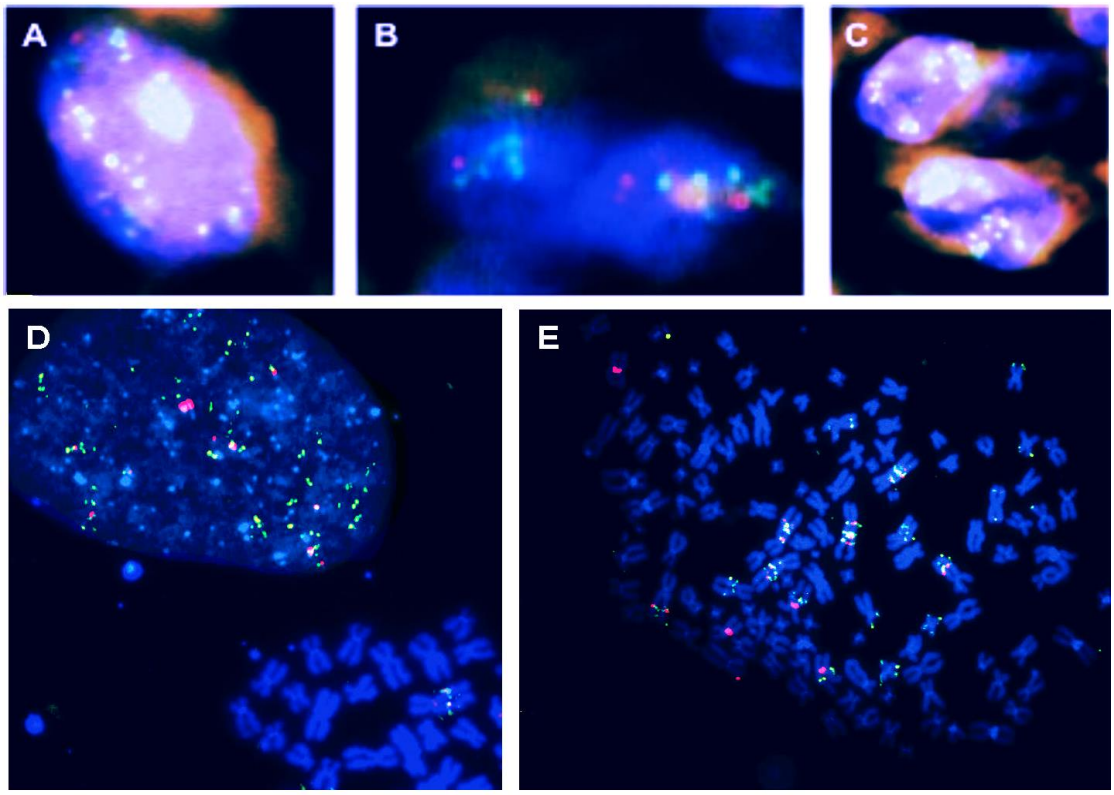

Figure 6.7: Two-colour Fluorescence *in situ* Hybridisation (FISH) for confirmation of *TRIO* amplification on FFPE sections from UPS 05 and corresponding STS 06/11 cells

The cBAC probe for the *TRIO* locus (green signals) and commercial CEP3 (red signals) were applied simultaneously and the nuclei and chromosomes were counter-stained with DAPI (in blue). **A, B and C** – interphase nuclei from FFPE section of UPS 05 captured in a single focal plane showing high number of green *TRIO* locus signals and 2-3 red CEP3 signals (white arrows). Changing the focal plane showed that most nuclei contained similar number of red and green signals. **D** – Interphase nucleus and partial metaphase spread of STS 06/11 cells that were derived from long term culture of UPS 05 tissue showing that they retain the high level *TRIO* amplification but also have numerous copies of CEP3, compared with parent tissue in B. **E** – Metaphase chromosome spread of STS 06/11 cells showing *TRIO* amplification pattern as numerous single copy amplifications dispersed among various chromosomes. Some of the green signals are clustered on abnormal chromosomes, where they are in close proximity to red CEP3 signals (yellow arrows). Only four of the CEP3 signals appear to be centromeric in location (white arrows).
All images were captured with x100 objective

Around sixteen red CEP3 signals were visualised on average in interphase nuclei (Table 6.2). Surprisingly, only four of these were centromeric in location while the remaining were mostly in close proximity to clustered green signals on abnormal chromosomes (Figure 6.5E). In order to rule out non-specific binding from an excess of red CEP3 probe, repeat experiments were carried out using as little as 20% dilution of the probe (based on optimisation by experienced colleagues in the RTRG laboratory). All showed identical distribution of red CEP3 signals on multiple metaphase chromosome spreads.

Similarly, *TRIO* locus amplification detected by array CGH in UPS 02 tissue was visually confirmed by FISH analysis of FFPE sections (Figure 6.8A) as well long-term cultured STS 14/10 cells (Figure

6.8B and C). Metaphase chromosomes derived from these cells showed a near-triploid karyotype with seven copies of the *TRIO* probe on average, only two of which appeared to be on normal-looking Denver Group B chromosomes (Figure 6.8B and C). One of four red CEP3 signals was localised on an acrocentric chromosome (Figure 6.8B and C).

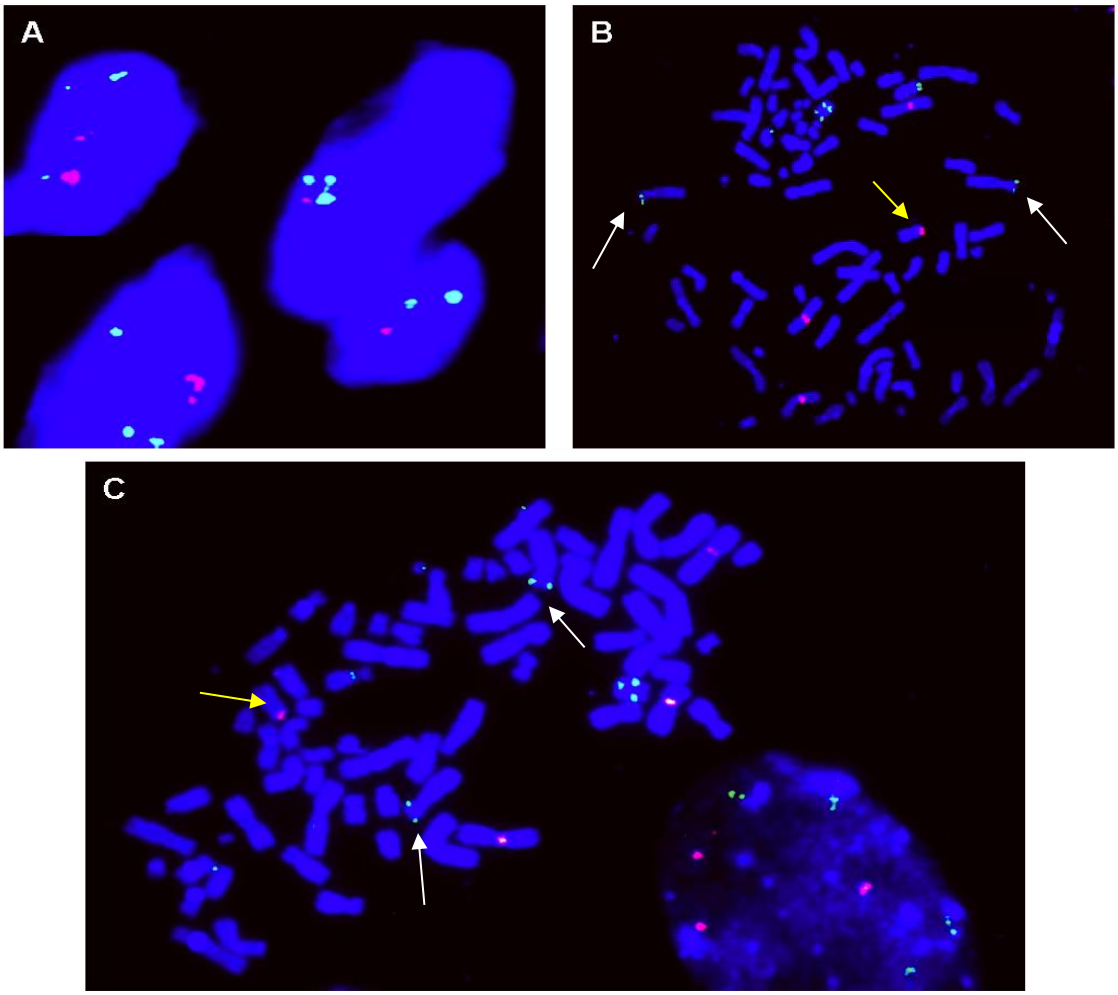

Figure 6.8: Two-colour Fluorescence *in situ* Hybridisation (FISH) for confirmation of *TRIO* amplification on FFPE sections from UPS 02 and corresponding STS 14/10 cells

The cBAC probe for the *TRIO* locus (green signals) and commercial CEP3 (red signals) were applied simultaneously and the nuclei and chromosomes were counter-stained with DAPI (in blue). **A** – Interphase nuclei from FFPE section of UPS 02 captured in a single focal plane showing >2 green cBAC signals in the majority of cells. Adjustment of the focal plane showed that most nuclei contained similar numbers of green and red signals in a ratio of 5:2, respectively. **B and C** – Metaphase chromosome spreads and partial Interphase nucleus of STS 14/10 cells that were derived from long-term culture of UPS 02 tissue showing a near-triploid karyotype and that the tumour cells retain the *TRIO* amplification seen with parent tissue in Panel A with additional red CEP3 signals,. Only two of the green signals appear to be on Denver group B chromosome (white arrows), and one of four red CEP3 signals is located abnormally on an acrocentric chromosome (yellow arrows).
All images were captured with x100 objective

6.2.2 Evaluation of Trio RhoGEF Protein Expression

Immunohistochemistry to confirm Trio protein expression was carried out on 4µm-thick tissue sections that were pre-treated to quench endogenous peroxidase and expose relevant epitopes in 1% H_2O_2 and Tris-EDTA buffer (pH 9.0) respectively, as described in the Methods Section 2.3.9.2. A rabbit polyclonal antibody directed against a recombinant Trio protein epitope signature tag (PrEST) was used. Also referred to as mono-specific antibodies, PrEST antibodies are generated against highly specific 100-150 amino acid fragments of the target protein identified by *in-silico* proteomic analysis. When used for immunohistochemistry, they have been shown to demonstrate specificity comparable to corresponding traditional monoclonal antibodies (Nilsson, Paavilainen et al. 2005).

6.2.2.1 Control Experiments

The DAB colorimetric system was used to detect positive protein expression (brown staining) with haematoxylin as a counter-stain (blue staining). Normal tonsil tissue, which has physiologic Trio expression was used as positive control and the cells showed strong positive cytoplasmic staining (Figure 6.9A). Negative experimental controls in which the primary antibody was omitted were set up simultaneously for every stained section (including positive controls). As shown in the example in Figure 6.9B, they only showed blue haematoxylin counter-stain.

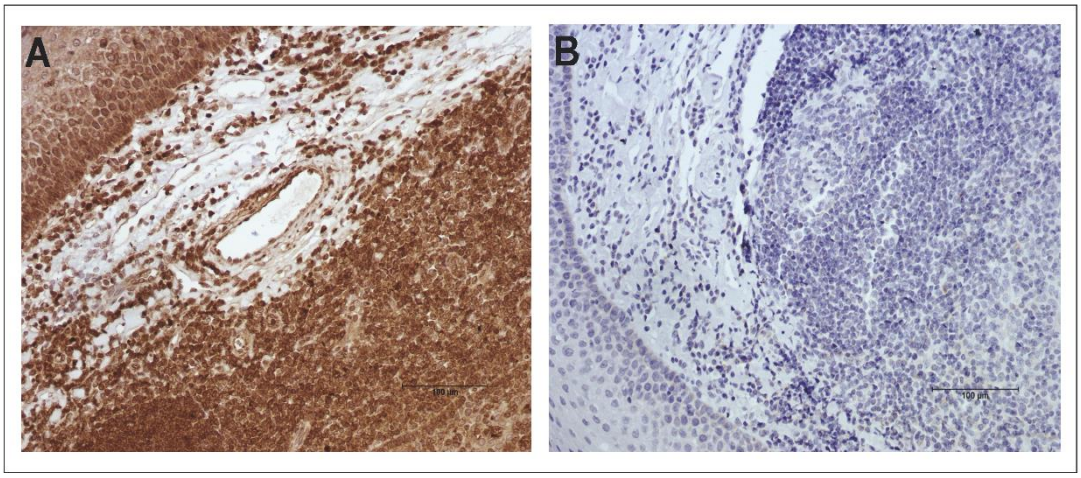

Figure 6.9: Experimental controls for Trio protein detection by Immunohistochemistry

A – Micrograph of FFPE section of normal tonsil tissue immuno-stained with Trio antibody (brown) with haematoxylin counterstain (blue) showing strong positive staining in cells but not acellular areas. **B –** Negative control (primary antibody omitted) of same normal tonsil tissue showing only blue haematoxylin counterstain. *Images were captured at 400x magnification of original*

6.2.2.2 Trio RhoGEF Protein Expression in STS

Evaluation of all immuno-stained sections was carried out at 200x magnification and confirmed by experienced Sarcoma Histopathologist, Dr Malee Fernando. Samples were classified as positive or negative for Trio expression based on cytoplasmic-stain detection in tumour cells. In the positive cases, the stain was detected in both small and bizarre giant tumour cells, which stained fairly uniformly. Positive staining was sub-classified as strong, moderate or weak based on the intensity of staining detected in tumour cells (Figure 6.10).

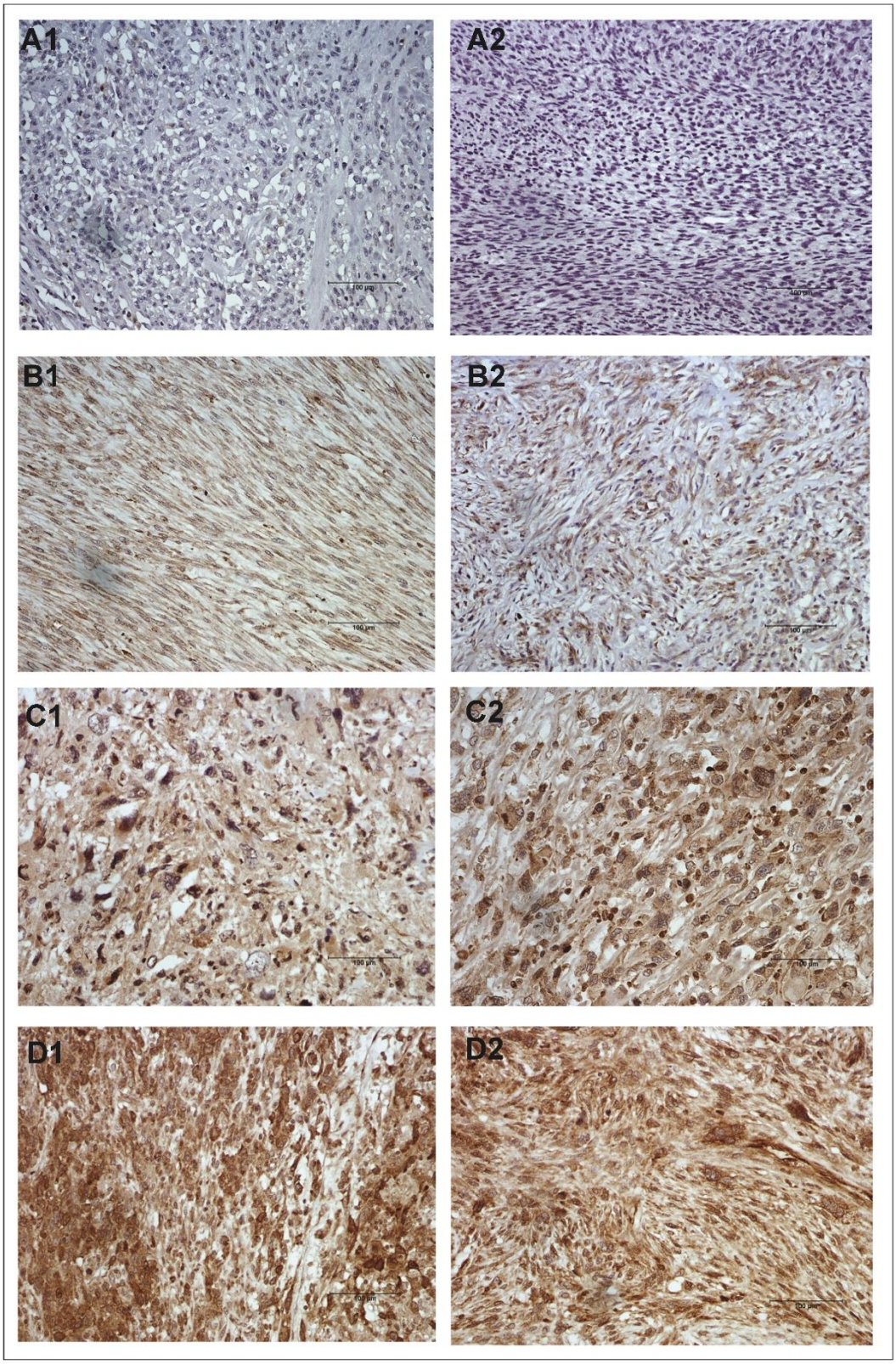

Figure 6.10: Semi-quantitative Analysis of Trio Protein Expression in STS Samples using Immunohistochemistry

Representative micrographs of FFPE sections of STS cases immuno-stained with Trio antibody (brown) and counterstained with haematoxylin (blue) showing the classification of staining intensities of tumour cells. **A1, A2** – negative staining seen in LMS 22 and LMS 14, respectively. **B1, B2** – weak positive staining in LMS 11 and LMS 17, respectively. **C1, C2** – moderate positive staining in UPS 15 and UPS 13, respectively. **D1, D2** – strong positive staining seen in UPS 05 and UPS 06, respectively.
Images were captured at 200x magnification of original and semi-quantitative scoring carried out by M. Fernando (experienced Sarcoma Pathologist).

Cases for immunohistochemical analysis were chosen to include UPS and another subtype (LMS) in which some cases showed *TRIO* amplification by array CGH while others did not. A total of nineteen cases comprising nine (9) UPS and ten (10) LMS tumours were analysed. All but one of these cases (LMS 22) had previously been analysed by array CGH. Results of immunohistochemistry for Trio and corresponding array CGH results are summarised on Table 6.3.

6.2.2.3 Correlation of TRIO Copy Number Status with Protein Expression

All UPS cases analysed were positive for Trio protein expression with at least moderate stain intensity, regardless of array CGH status. All three cases that were confirmed to have *TRIO* amplification (UPS 02, 05 and 06) showed strong staining intensity. Overall, a good correlation (r^2 = 0.7; n = 8) was demonstrated between aCGH status and staining intensity for Trio in UPS when analysed by the Spearman's rank test. Among the LMS cases on the other hand, the majority were either negative for Trio protein expression or only showed weak stain intensity that had poor rank correlation with copy number status (r^2 = 0.0002; n = 9).

Table 6.3: Comparison of Trio protein Expression with Copy Number Status of *TRIO* gene locus among STS cases

Case	aCGH Status of *TRIO* locus	Staining Intensity of Trio RhoGEF Protein
UPS 02	Amplified	Strong
UPS 05	Amplified	Strong
UPS 06	Amplified	Strong
UPS 07	Normal	Strong
UPS 09	Deleted	Moderate
UPS 12*	Normal	Focally Strong
UPS 13	Deleted	Moderate
UPS 14	Normal	Moderate - Strong
UPS 15	Normal	Moderate
LMS 04	Amplified	Moderate
LMS 06	Normal	Weak - Moderate
LMS 11	Amplified	Weak
LMS 13	Normal	Weak
LMS 14	Normal	Negative
LMS 15	Normal	Negative
LMS 16	Normal	Weak- Moderate
LMS 17	Amplified	Weak
LMS 21	Deleted	Moderate
LMS 22*	NP	Negative

***- Sample excluded from Spearman's rank correlation analysis

6.2.4 Effect of Trio RhoGEF Inhibition on Tumour Cell Proliferation

The proliferation of STS 06/11 cells treated with varying concentrations of the Trio GEF1-specific inhibitor, ITX3 was evaluated using the MTT assay as descried in Methods Section 2.3.11. The range of treatment concentrations was based on a review of the literature in which a significant reduction in Rac1 levels was demonstrated in mammalian cells that were treated with ITX3 at 100µM (Bouquier, Vignal et al. 2009). An immortalised retinal epithelial cell line, hTERT-RPE1 was used as non-malignant cell control.

Results were based on five replicate wells and showed that hTERT-RPE1 cells showed no significant reduction in proliferation at any of the treatment doses compared with DMSO vehicle-treated controls. When compared with dose-matched STS 06/11 cells however, two-way ANOVA (with Bonferroni's post hoc test correction) showed that tumour cells treated with ≥ 200µM ITX3 for 72-hours showed significantly reduced proliferation ($p < 0.05$) (Figure 6.11). These results suggest a tumour cell-specific inhibition of cell survival by ITX3.

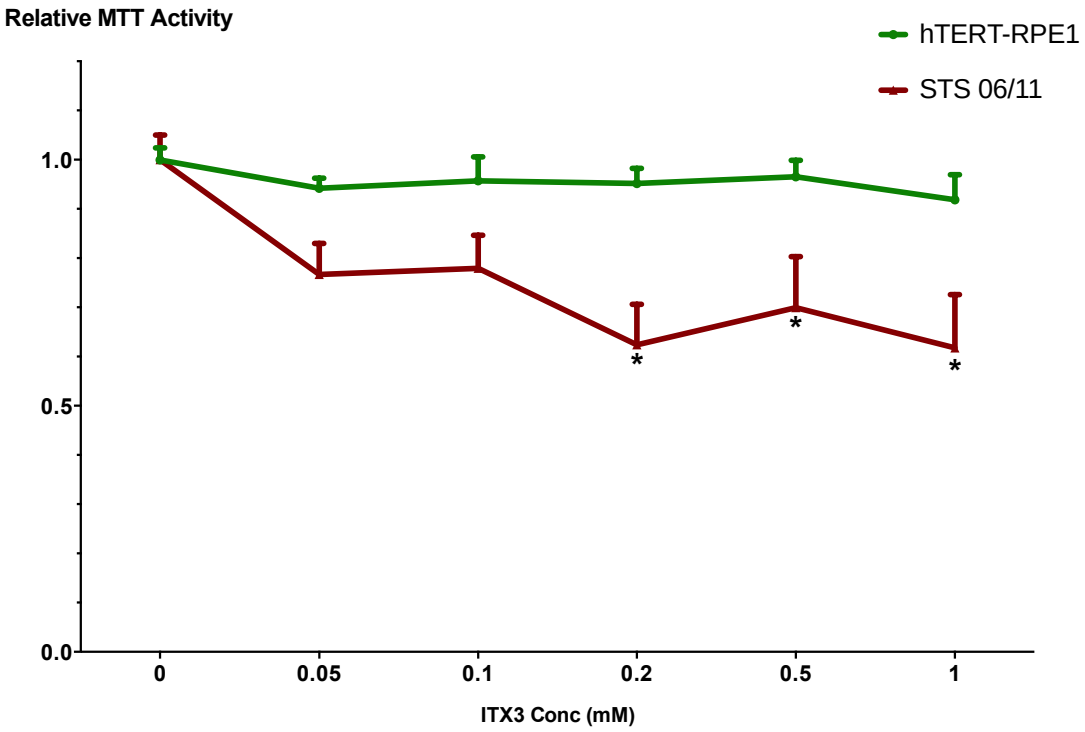

Figure 6.11: Summary of the effects of ITX3 treatment on STS 06/11 cell proliferation in an MTT assay

Relative MTT activities were calculated as the ratio of absorbance in treatment to control wells after 72 hours of treatment and the data is displayed as Mean + SEM of five replicates. Significant reduction in proliferation ($p<0.05$) was seen in STS 06/11 cells treated with ≥ 200µM ITX3 (marked with *) compared with dose-matched hTERT-RPE1 cells as judged by two-way ANOVA with Bonferroni post-hoc correction.

6.2.6 Effect of Trio RhoGEF Inhibition on Tumour Cell Invasion

The Boyden chamber assay was used to assess the invasive ability of STS 06/11 cells through a porous membrane coated with an extracellular matrix surrogate as described in methods section 2.3.12. Treatment doses (concentrations) of ITX3 were chosen as described in the preceding section. Membranes were examined at high power (400x original magnification) in order to distinguish invaded cells present on the underside of the membrane from cells remaining on the top surface as shown in Figure 6.12.

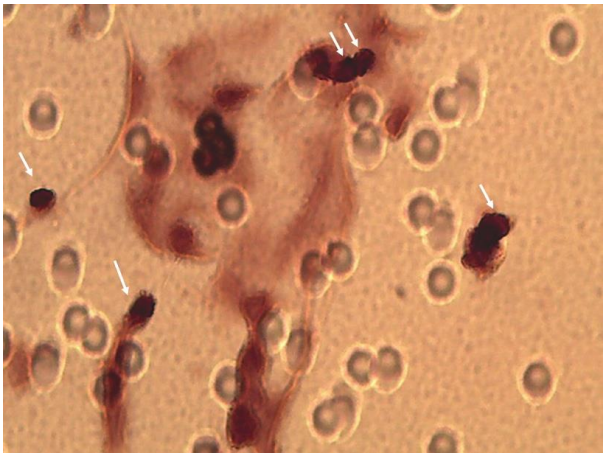

Figure 6.12: Representative micrograph of invasive STS 06/11 cells from Boyden's chamber assay

Cells on the underside of porous polycarbonate membranes were stained with haematoxylin and the membrane mounted flat on a glass slide with the underside facing upwards. Membranes were examined at high power (x400 magnification) at which images were captured. Only cells that came into focus while the 8µm pores were still blurred (white arrows) were considered to be on the underside of the membrane and counted.
Image was captured at 400x original

The average cell counts obtained from five random fields in three replicate wells were plotted against the corresponding dose of Trio inhibitor and the results presented in Figure 6.13. Although ITX3 appeared to have a dose-dependent inhibition on STS 06/11 cell invasion (Figure 6.13), comparison of the mean cell counts by one-way ANOVA with Bonferroni's post-hoc correction (for multiple treatment groups) did not reveal any statistically significant difference in any of the treated groups compared with DMSO controls.

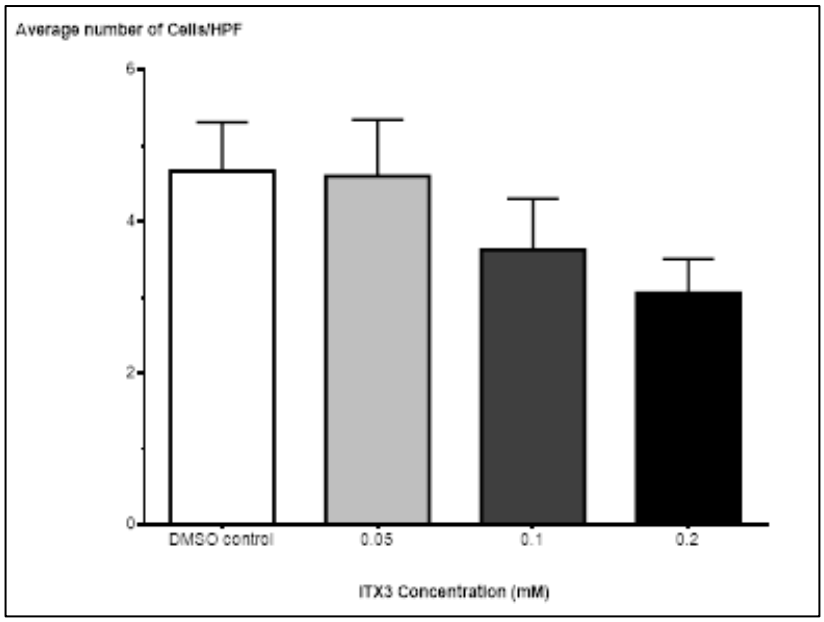

Figure 6.13: Summary of the effects of ITX3 treatment on STS 06/11-tumour cell invasion in the Boyden Chamber Assay.

The data is displayed as Mean + SEM of the number of invaded cells seen in five random microscopy fields at x400 magnification and based on triplicate experiments. No significant reduction ($p<0.05$) was seen in any of the treatment groups compared with DMSO control as judged by one-way ANOVA with Bonferroni's post-hoc correction.

6.3 DISCUSSION

Despite recent developments in molecular classification of STS, the UPS subtype remains largely a diagnosis of exclusion with little known about the molecular characteristics of these tumours apart from genetic instability and karyotypic complexity (Fletcher *et al.*, 2013). On the basis of this relative dearth of specific molecular characteristics, one of the targets identified among UPS using genome-wide array CGH screening (in chapter 4) was chosen for further evaluation. The high frequency of *TRIO* amplification among these tumours, combined with biological evidence for its role in tumour progression as derived from various other cancers made this gene a strong candidate driver and thus the subject of this chapter. The availability of primary cells in stable long–term culture from three separate UPS cases in this PhD project provided a suitable model for the requisite validation studies, the preliminary results of which were presented.

Within this time-limited PhD study, only a few techniques could be optimised for validation studies and for some of these, few cases could be analysed in detail. Relevant FISH probes had to be prepared from BAC DNA and then optimised for use on FFPE samples before actual time-consuming analysis of individual cases as parent tissue and corresponding cultured cells. Similar optimisation had to be carried out with the antibodies for immunohistochemistry, leaving little time for other expression studies and functional *in vitro* experiments. Definite conclusions can therefore not be arrived at based on the results presented here. However, this chapter demonstrates the approach to target validation and all the methods discussed have now been optimised with pilot data that is being used to drive further studies with sufficient numbers and statistical power.

Amplification affecting the *TRIO* locus relative to chromosome 3 centromere was confirmed using FISH on FFPE sections of all three UPS cases and analysis of the corresponding long-term cultured cells showed that the amplification was retained. In the case of UPS 06 where array CGH analysis of the cultured STS 09/11 cells failed to show the *TRIO* amplification that was present in parent tissue, direct, targeted visualisation by FISH identified a low-level amplification that was seen in less than half the cells. These results confirm that the primary cell lines are valid models for further evaluation of *TRIO* amplification as a target and support the hypothesis that this frequent and targeted amplification is important for tumour pathogenesis, since it persists in tumour cells despite long periods of *in vitro* culture. The findings also raise the possibility that the frequency of *TRIO* amplification among UPS tumours may have been underestimated by array CGH, a well-recognised limitation of any genome-wide screening technique. Following the successful development and optimisation of the requisite probes, analysis of more cases including UPS cases that did not show *TRIO* amplification by array CGH as well as other STS subtypes will be carried out as follow-on work from this project. This will also explore the potential diagnostic utility of FISH for UPS in general or identification of tumours that are likely to respond to Trio rhoGEF-targeted therapeutics. Similarly, any other potential genomic aberrations of interest deserve further targeted exploration to confirm (and accurately estimate) their frequency.

While both UPS 02 and UPS 06 showed modest *TRIO* locus amplification, UPS 05 showed very high-level amplification and had a relatively rapid establishment in culture as well as a short doubling time (see Section 5.2.7). Similar variation in the levels of amplification of pathogenetic genes have been described in various other cancers, notable examples including *ERBB2* in breast cancer and *MYCN* in neuroblastoma (Albertson, 2006; Mondello *et al.*, 2010). Further, associations have been described between these different levels of amplification and patient survival and/or response to targeted therapy in *ERBB2*-amplified breast cancer (Arnould, Arveux et al. 2007, Guiu, Gauthier et al. 2010, Gullo, Bettio et al. 2013), raising another important question in relation to *TRIO* amplification in UPS that requires exploration in further study.

Good quality metaphase chromosome spreads from the primary tumour cell lines allowed analysis of the patterns of amplification in UPS 02 and 05. The pattern of single copy amplifications spread over various chromosomes (dispersed insertions) was seen in both cases but UPS 05 showed in addition, clusters of the amplified *TRIO* locus reminiscent of homogenously-stained regions (HSR) on three abnormal-looking chromosomes. A similar pattern of amplification has been described in the *MYCN* gene in neuroblastoma (Schwab, 1999), and more recently in *EGFR* amplifications in gliomas (Lopez-Gines *et al.*, 2010), where it is believed to result from the re-incorporation of extra-chromosomal copies of amplified DNA (dmins) into various genomic regions, where they remain as single copies or undergo further replication to form HSR (Albertson, 2006; Schwab, 1999).

Another interesting feature of UPS 05 metaphase FISH (shown in Figure 6.7) was the proximity of non-centromeric CEP3 signals to the clustered *TRIO* amplicons. This was confirmed by repeat experiments and absent from control experiments suggesting that the pattern was not due to non-specific CEP3 probe binding. One possible explanation however is chromothripsis, a phenomenon recently described in molecular detail by Stephen et al using paired-end analysis (Stephens, Greenman et al. 2011) and that is increasingly being recognised as a mechanism for genomic instability in cancer cells (Korbel and Campbell 2013, Zack, Schumacher et al. 2013). Instead of, or in addition to traditional view of progressive acquisition of mutations or chromosome aberrations by a tumour cell, a single 'cataclysmic' event is believed to result in the shattering of one or few chromosomes. The fragments are then 'stitched' together, by non-homologous end-joining (NHEJ), which in addition to non-uniform cytokinesis gives rise to complex rearrangements on chromosome arms with multiple copy number states on contiguous genomic regions (Stephens, Greenman et al. 2011).

Chromothripsis can also result in the formation of circular double minutes consisting of DNA from different chromosomes as clearly described by Rausch et al in medulloblastoma cells where it targeted two frequently co-amplified genes *MYCN* and *GL12* (Rausch, Jones et al. 2012). Re-incorporation of such double minutes into chromosomal structure can result in the pattern of colocalisation of red CEP 3 and green *TRIO* locus signals seen in STS 06/11 cells. Also interesting is that retrospective examination of the array CGH profile of UPS 05 (and STS 06/11) showed the presence of a high level amplification detected at the centromeric end of 3p, which was seen in

around 60% of UPS cases and involves the locus for *VGLL3,* one of the strong candidate genes identified by this study (Section 4.4.3.7).

Multiple studies have found the frequency of chromothripsis to be between 2 and 5% of cancers in general, but up to 16% in highly genetically-unstable glioblastomas and even 25% among bony mesenchymal tumours such as osteosarcomas and chordomas (Stephens, Greenman et al. 2011, Kim, Xi et al. 2013, Zack, Schumacher et al. 2013) and chromothripsis events have been shown to be more common in regions involved in putative driver genomic aberrations with high level amplifications, most of which were detected using the GISTIC algorithm (Zack, Schumacher et al. 2013). Given the high level of genomic instability of UPS and some of the copy number aberration profiles seen by array CGH in this and other studies, chromothripsis as a contributor to pathogenesis in STS is not unlikely. Although time did not permit within this PhD study, further FISH analysis of the pattern of CEP probes other than chromosome 3 on STS 06/11 metaphases spreads would be a logical first step in discerning the specificity of the co-localisation while PCR and paired-end wise analysis using next generation sequencing would help confirm whether the exact structural variation based on genomic sequences (Hillmer, Yao et al. 2011, Stephens, Greenman et al. 2011).

Evaluation of Trio protein expression by immunohistochemistry, which was carried out on selected UPS and LMS cases also yielded very interesting results. The choice of LMS for subtype comparison was based on the availability of array CGH results on over 20 samples, around a quarter of which showed *TRIO* copy number amplification. Moderate to strong positive staining intensity and good correlation with array CGH copy number was seen among the UPS cases in contrast to LMS where the overall stain intensity was lower, even in cases where array CGH showed increased gene copy number. These results support the hypothesis that the pathological role of *TRIO* amplification is more specific among UPS, where the copy number amplification is more frequent and correlates with higher expression levels, similar to the correlation between *ERBB2* amplification and overexpression in breast cancer (Gullo, Bettio et al. 2013). Further confirmation of Trio expression levels in both tissue samples and the cell line models using techniques that are noted for their quantitative application such as western blotting or quantitative real-time PCR will however be required.

Statistically significant reduction in the proliferation of the primary UPS cell line STS 06/11 treated with specific GEF-inhibitor ITX3 correlated with findings from siRNA–mediated *TRIO* knockdown studies that showed reduced proliferation of glioblastoma cells that were found to be over-expressing Trio (Salhia *et al.*, 2008a). More importantly however, ITX3 inhibition had no effect on the normal epithelial cell line controls, suggesting that its effects are tumour cell-specific - the ‘holy grail’ of cancer therapeutics. Tumour cell invasion appeared to be attenuated by ITX3 treatment, although this was not statistically significant. It is noteworthy, however that the data is limited to a single Boyden chamber experiment with a small range of inhibitor concentrations. Clearly, no definite conclusions are drawn in terms of functional effects of Trio inhibition, but the copy number and expression data remain very promising in light of future work with this exciting potential target.

CHAPTER SEVEN

FINAL DISCUSSION

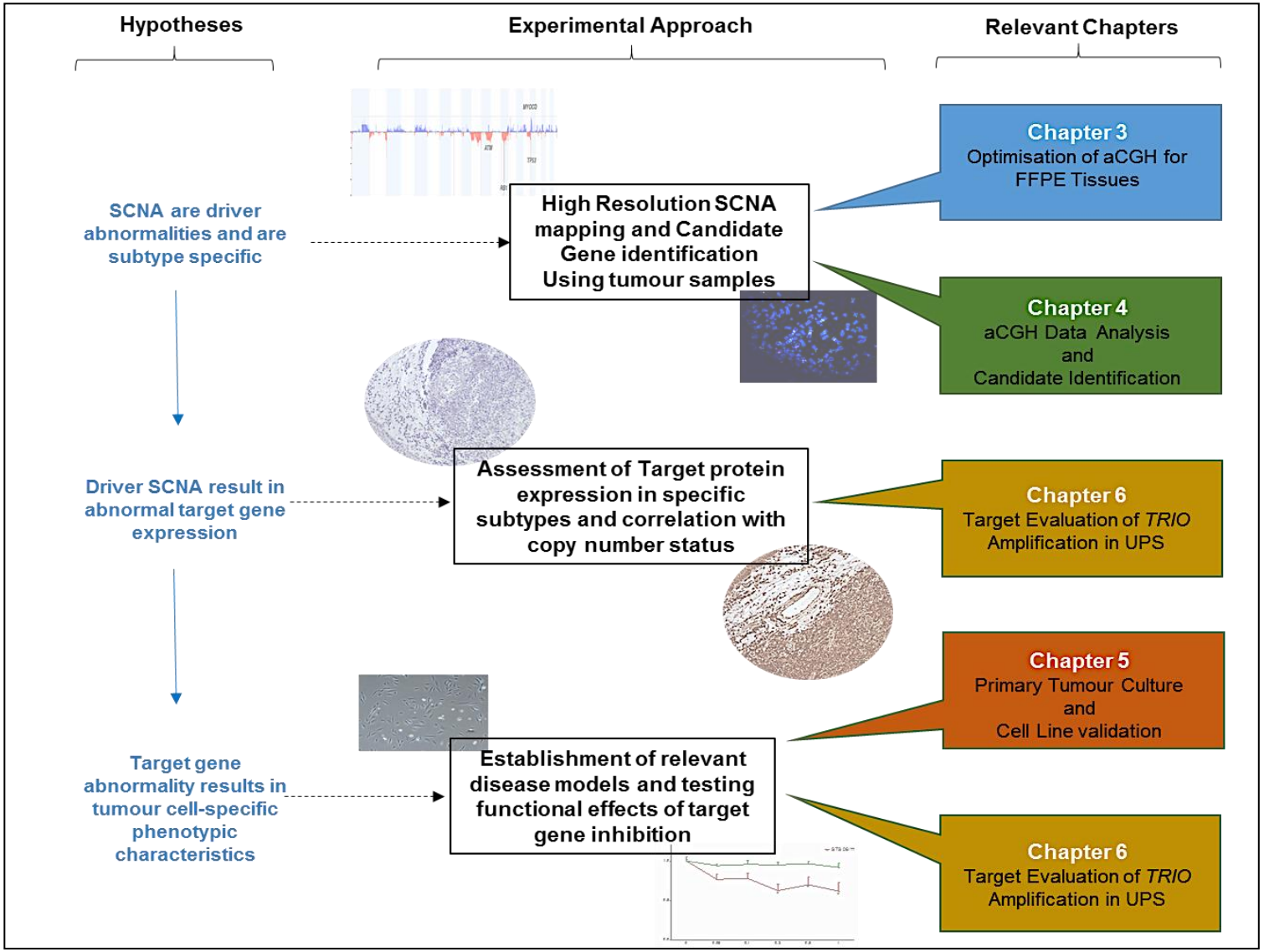

Figure 7.1: Outline of the Approach used in this PhD Study for Pathogenetic Target Identification in Soft Tissue Sarcoma.

7.1 Summary of Approach to Target Identification in STS

In cancers such as STS where they are pervasive, the study of how focal SCNA impact cancer development and evolution remains very important (Zack, Schumacher et al. 2013). With technological advancements in molecular genetic analytic methodologies such as high resolution array CGH and next generation sequencing, the resultant accumulation of larger and more detailed genomic data means that comprehensive approaches are required in addition to statistical analysis to decipher which SCNA are drivers and how best to identify therapeutic targets. A summary of the approach to molecular pathway identification used in this PhD study is outlined in Figure 7.1.

The optimisation of a suitable methodology for analysis of available FFPE tumour samples was addressed in the first results chapter (Chapter 3). This led to accumulation of sufficient tumour sample numbers to permit statistical analysis of the array CGH data with identification of candidate genes and molecular pathways, the results of which are presented in Chapter 4. The next logical steps were the evaluation of karyotypic patterns of focal SCNA in individual tumours and their effects on target gene expression and tumour phenotype, as demonstrated in the final results chapter (Chapter 6) exploring Trio RhoGEF as a potential target in UPS tumours. To permit this evaluation however, suitable material for functional analysis of tumour phenotype in the form of an *in vitro* model was successfully established and validated for a range of STS subtypes, the details of which were presented in Chapter 5.

7.2 Study Limitations

The primary limitation that was identified very early in the study was a perceived unsuitability of FFPE tumour samples for high-resolution analysis of genomic copy number by array CGH. Fresh tissue collection was required for the development of primary cell lines, an essential part of the project but it became clearly evident that this approach would yield insufficient tumour numbers for such a rare tumour as STS. Further, the prospective tumour collection would likely yield a random assortment of STS subtypes, which would preclude meaningful subtype-specific analysis - one of the primary aims of the PhD study. Successful tackling of the technical problem as demonstrated in Chapter 3 by optimising the use of relatively abundant FFPE tumour samples meant that even though fresh tissue from only 7 UPS, 3 LMS and no GIST cases were obtained in the entire course of the study (Table 5.1), reliable data from 16 UPS, 21 LMS and 11 GIST cases among other subtypes was eventually analysed and this methodology led to the first peer-reviewed publication from this PhD project (Salawu, Ul-Hassan et al. 2012).

Another important constraint identified was the limited time available to perform all the required experiments. In order to be pragmatic therefore, different aspects of the project such as primary cell line development, array CGH optimisation and data analysis were undertaken concurrently. Since sufficient time had to be allowed for further evaluation of an identified target, the strongest candidate gene from statistical analysis of aCGH data from one of the most enigmatic STS subtypes, i.e. *TRIO* amplification in UPS was chosen as the subject for Chapter 6. This however meant that candidate

genes that were subsequently identified with even stronger statistical evidence such as *JUN* and *VGLL3* amplifications in UPS could not be explored further within the context of this study.

7.3 STUDY FINDINGS

Subtype-specific candidate genes for LMS, UPS and GIST that were identified have been discussed previously (Section 4.6). The discussion in this chapter is therefore focused on more general observations in the context of STS in general and possible implications for tumour biology, classification and targeted-therapeutic development in LMS and UPS.

Figure 7.2 below summarises some of the most important candidate genes and molecular pathways identified in STS from this study and the likely areas of tumour biology in which they are involved.

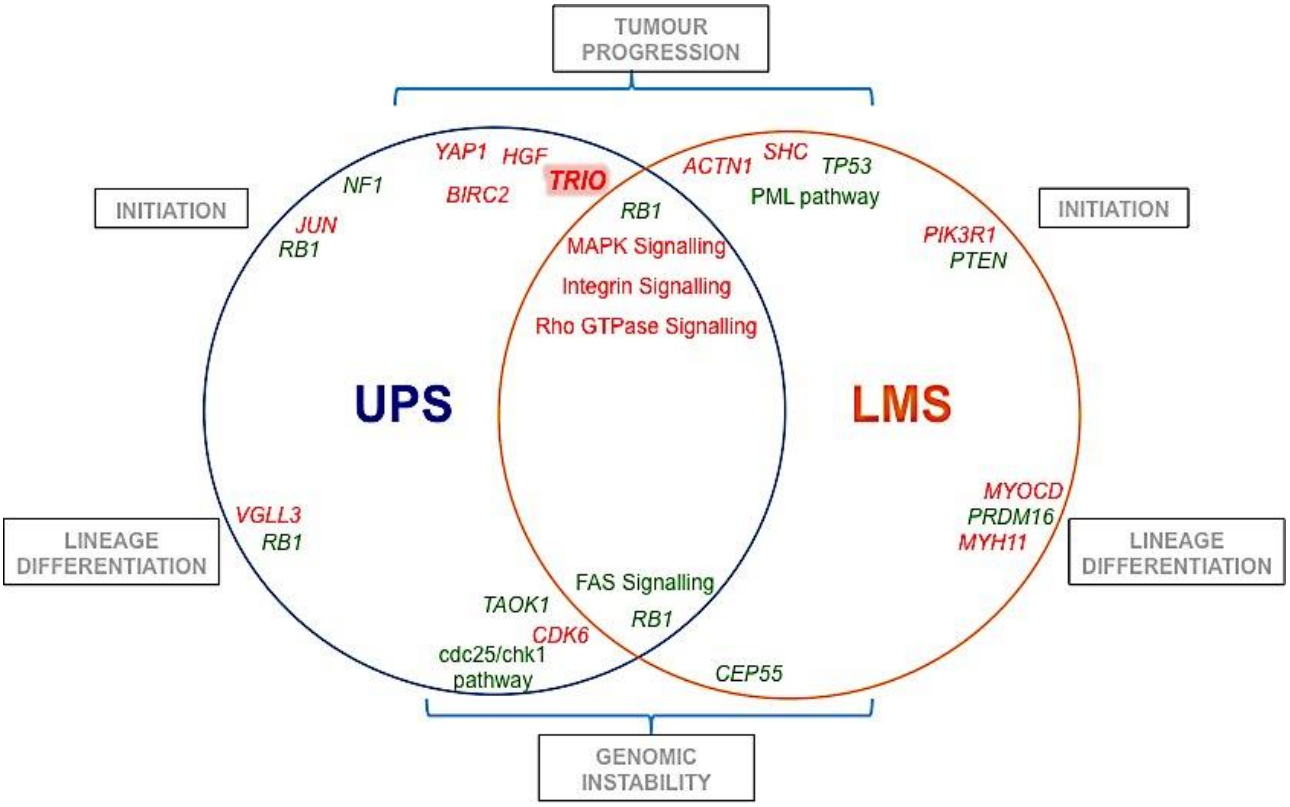

Figure 7.2: An Illustration of the Subtype-specific and Overlapping Candidate Genes and Pathways identified among Leiomyosarcoma (LMS) and Undifferentiated Pleomorphic Sarcoma (UPS) in this study.

Candidate gene identification and likely roles in tumour biology are based on genomic copy number analysis and evidence in literature. Genes and pathways shown in red are involved in amplifications while those in green are involved in deletions. The highlighted gene, *TRIO* has been further evaluated in this study with additional expression and functional basis for a role in tumour progression.

7.3.1 STS subtypes display a wide range of karyotypic complexity

Array CGH profiles from the wide range of STS subtypes analysed generally correlated with the prevailing theory that those with known specific translocations or other genetic characteristics such as synovial sarcoma, alveolar soft part sarcoma and GIST have relatively few unbalanced chromosomal abnormalities and thus simpler array CGH profiles (Mertens, Panagopoulos et al. 2009, Fletcher, Bridge et al. 2013). Among the subtypes with complex karyotypes however, there was significant variation in the degree of karyotypic complexity evidenced by the proportion of the genome that was involved with focal SCNA. UPS showed the most numerous focal SCNA overall with frequent common aberrations affecting up to a third of the genome, compared to only one-fifth of the genome seen in LMS or an average of 10% of the genome as reported for cancers in general (Beroukhim, Mermel et al. 2010). In keeping with this apparently worse magnitude of genomic instability, deletion of several candidate genes and pathways that normally function in genome maintenance including *RB1* and *TAOK1* were identified among UPS in this study many of which have not been previously reported among STS (discussed in Section 4.6).

7.3.2 Are certain focal SCNA and molecular pathway abnormalities common to various STS subtypes?

The most obvious shared frequent aberration between UPS and LMS was the deletion on 13q affecting the *RB1* gene locus. Interestingly, this was one of the most common deletions in both subtypes, affecting 13 out of 16 UPS cases (81%) and 14 out of 21 LMS cases (67%). These results are in keeping with previous studies that reported frequent *RB1* losses among STS with complex karyotypes (Taylor, Barretina et al. 2008, Mertens, Panagopoulos et al. 2009) as well as radiation-related sarcomas that include osteosarcomas, LMS and UPS (Gonin-Laurent, Hadj-Hamou et al. 2007). Further, up to 30% of secondary tumours described in familial retinoblastoma patients (who carry germ-line *RB1* gene loss) are STS and while most of the secondary tumours in these patients appeared to be radiotherapy-related, the majority of secondary LMS occurred outside the radiation field, suggesting a radiation-independent role for the inherited *RB1* loss in sarcomagenesis (Kleinerman, Schonfeld et al. 2012).

One of the most well-studied tumour-suppressor genes, functional disruption of the Rb1 pathway by inactivating mutations, copy number loss or epigenetic mechanisms have been reported in a wide range of cancers (Di Fiore, D'Anneo et al. 2013). Numerous reviews, including a recent one by Henley and Dick highlight the important role *RB1* plays in the G1-S phase cell cycle checkpoint and control of cellular proliferation (Henley and Dick 2012) and evidence suggests that *RB1* disruption in itself may be sufficient for tumour initiation (Liu, Sanchez-Tillo et al. 2013). More recently, objective evidence that inactivation of even one copy of *RB1* promotes telomere dysfunction and genetic instability has been demonstrated (Alessio, Bohn et al. 2013, Gonzalez-Vasconcellos, Anastasov et al. 2013).

The pattern of deletion of the *RB1* locus detected in this study was particularly notable. The majority of tumours had homozygous focal deletion of the *RB1* gene locus, either on a background of hemizygous 13q loss or with normal diploid 13q. This resulted in highest deletion G-scores assigned to this locus by the GISTIC algorithm in both subtypes (Tables 4.7 and 4.22). This deletion pattern

suggests that the *RB1* locus was specifically targeted for copy number loss, a finding that supports a pathogenetic role for *RB1* loss that is statistically valid in itself, but even more exciting in the context of the functional role of the *RB1* gene. Interestingly, although frequent *RB1* loss among STS has been reported in literature, this specific pattern of RB1 deletion is a novel finding in STS or any other cancers for that matter that adds strong evidence this abnormality being a 'driver'. The specific mechanism by which this targeted loss occurs however still remains unidentified.

Further, enrichment analysis of the frequently deleted genes in both LMS and UPS subtypes in this study revealed significant disruption of the PML and cdc25/chk1 pathways respectively (Sections 4.4.1.5 and 4.4.3.5), both of which are positively regulated by the *RB1* gene product with important roles in genome and telomere maintenance (Zhong, Hu et al. 1999, Dai and Grant 2010, Chang, McGhie et al. 2013). Taken together, there is a strong overall suggestion that *RB1* pathway disruption is an important pathogenetic mechanism in STS and may be an early event, particularly in the karyotypically-complex subtypes where *RB1* appears to be specifically targeted copy number loss. Conversely, it is possible that *RB1* disruption may be of less importance among translocation-driven sarcomas, which generally exhibit less genomic instability and have alternative putative initiation mechanisms.

Another common pathway that was highlighted in this study was the Mitogen-activated Protein Kinase (MAPK) pathway, which was detected by enrichment analysis of amplified genes in both UPS and LMS subtypes (Table 4.8 and 4.23). An important molecular signalling pathway for cancer, the MAPK pathway functions via a three-tiered kinase module that links extracellular influences such as growth factors and stress signals to cellular proliferation and other tumour phenotypic characteristics such as cell survival, metastasis and even angiogenesis (Dhillon, Hagan et al. 2007, Yang, Sharrocks et al. 2013). Among the chief regulators of this pathway are members of the Ras or Raf family of proteins that phosphorylate MEK1 and MEK2, which subsequently phosphorylate ERK1 and ERK2, the main mediators of the MAP kinase pathway. Yet other MAPK pathways are known to exist. They are regulated via p38 and Janus-related kinase (JNK) and have significant cross-talk with the Ras/Raf/MEK pathway and different levels of significance in specific tumour types. Detailed reviews of the MAPK pathways and their role in cancer pathogenesis are available in literature (Dhillon, Hagan et al. 2007, Neuzillet, Tijeras-Raballand et al. 2013).

Implicated in up to a third of cancers, the majority of MAPK pathway-activating molecular lesions involve genes that encode regulatory proteins as frequently seen in pancreatic adenocarcinoma (*RAS* mutation in up to 90% of tumours), melanoma (*BRAF* mutation in 60%) and/or overexpression of growth factor receptors such as *ERBB2* in breast cancer (Downward 2003). The frequency of activation of members of the MAPK pathway and/or regulators in cancer therefore make it an attractive target for novel drug development with a number of inhibitors already in clinical use and many others in clinical trials (Neuzillet, Tijeras-Raballand et al. 2013). There is however, a relative paucity of data in literature regarding the role of the MAPK pathway in STS with the majority of available sarcoma data obtained from bone sarcomas.

Silvany et al first demonstrated constitutional activation of the MAPK pathway in cells transformed using the *EWS-FLI* fusion gene (that is characteristic for Ewings' sarcoma) and that treatment with

specific MAPK inhibitors resulted in loss of anchorage-independent proliferation (Silvany, Eliazer et al. 2000). Subsequent studies showed that the MAPK pathway-targeted inhibitors alone or in combination with cytotoxic agents caused a dose-dependent inhibition of proliferation and migration in Ewing sarcoma cell lines (Benini, Manara et al. 2004, Yamamoto, Ohno et al. 2009, Keir, Maris et al. 2010). These and other similar results in osteosarcoma (Huang, Lee et al. 2009, Pignochino, Grignani et al. 2009, Na, Kim et al. 2012) and chondrosarcoma (Papachristou, Papachristou et al. 2005, Lu, Tang et al. 2010) have led to a number of on-going clinical trials (Chandhanayingyong, Kim et al. 2012, Widemann, Kim et al. 2012) with encouraging results seen among patients with inoperable osteosarcoma treated with the *RAF* inhibitor Sorafenib (Grignani, Palmerini et al. 2012).

Sasaki and colleagues who carried out *in vitro* studies in UPS cells to evaluate the role of the MAPK pathway and found increased MEK1 and MEK2 mRNA levels in three cell lines (Sasaki, Hitora et al. 2011), findings which correlate with the enrichment analysis of frequent amplifications seen among UPS cases seen in this study (Table 4.23). In that study, treatment of the UPS cell lines with a MEK-specific inhibitor (U0126) caused reduction in the levels of phosphorylated downstream ERK as well as reduced cell proliferation. Another study based on an *NF1*-deficient transgenic mouse model of MPNST and UPS demonstrated tumour growth inhibition when treated with the MEK inhibitor PD325901, which is accompanied by reduced tumour cell proliferation but not increased apoptosis (Dodd, Mito et al. 2013). It is notable that *NF1* was one of the strong candidate deleted genes identified among UPS in this study, highlighted by multiple statistical approaches including common aberration and pathway enrichment analyses. Yet another study in STS demonstrated decreased proliferation and increased apoptosis in tumour-derived synovial sarcoma cells treated with the Raf-inhibitor, Sorafenib (Peng, Guo et al. 2009). The sum total of evidence therefore suggests that the MAPK pathway is a druggable target that is worthy of further preclinical and maybe even clinical evaluation in UPS and other STS, regardless of karyotypic complexity.

An important consideration in molecular pathway targeting is the 'cross-talk' (bidirectional signalling interaction) that exists among various pathways that have been established as a common mechanism of tumour cell resistance. Two pathways that have close cancer-relevant interactions with the MAPK pathway are the PIK3/PTEN and integrin signalling (focal adhesion) pathways (Giancotti and Ruoslahti 1999, Shimizu, Tolcher et al. 2012). Recently published data established a rationale for and showed that co-inhibition of the MAPK and PIK3/PTEN pathways had a synergistic effect on proliferation in rhabdomyosarcoma cell lines as well as xenograft tumours (Renshaw, Taylor et al. 2013). Similar dual pathway inhibition strategies have shown encouraging results in Ewing's tumours (Yamamoto, Ohno et al. 2009) and Kaposi's sarcoma (Lambert, Shahrier et al. 2007), as well as potential efficacy and tolerance advantages over single pathway inhibition in Phase I clinical trials (Shimizu, Tolcher et al. 2012). This is particularly interesting in the context of relevant amplifications and deletions detected among LMS in this study that implicate the PIK3/PTEN pathway (Section 4.4.1.5).

Also commonly enriched among amplified genes in both LMS and UPS subtypes in this study was the integrin signalling/focal adhesion pathway. A known mechanism of MAPK pathway regulation based on extracellular matrix (ECM) interaction with integrin molecules expressed on the cell

membrane, this pathway is involved in physiologic cell proliferation (Giancotti and Ruoslahti 1999). Dysregulation of this pathway (mainly via increased expression of specific integrins) has been implicated in tumour progression as a mechanism of epithelial-mesenchymal transition (Guo and Giancotti 2004). It is therefore plausible that activation of this pathway in addition to MAPK would be of pathogenetic relevance in such high-grade mesenchymal tumours as UPS and LMS.

7.3.3 Do specific STS Subtypes have different mechanisms for activating a Common Molecular Pathway?

The results from copy number analysis in this study strongly suggest that in the case of STS with complex karyotypes, the MAP kinase pathway is activated by copy number amplification of key genes that result from genomic instability, possibly initiated by *RB1* loss. This study is the first to demonstrate frequent copy number amplification of not only key member of the MAP kinase pathway, but also the related PIK3/PTEN and integrin signalling pathways in these two STS subtypes. Considering the frequency of MAPK pathway activation among other STS subtypes, especially those that do not show a similar degree of genomic copy number imbalance, it is possible that they utilise other alternative molecular mechanisms for MAP kinase pathway activation. In Ewing's sarcoma for instance, MAPK pathway activation is believed to be directly linked to the EWS-FLI1 fusion (Silvany, Eliazer et al. 2000, Rocchi, Manara et al. 2010) and manipulation of the MAPK kinase pathway by Human Herpes Virus 8 (HHV8) is well established and likely pathogenetic in Kaposi's sarcoma development (Lambert, Shahrier et al. 2007). It is therefore proposed that MAPK pathway activation is a pan-sarcoma pathogenetic mechanism that is triggered using subtype-specific mechanisms.

Likely to be involved in tumour progression, particularly in UPS is the amplification and overexpression of the Trio RhoGEF that was confirmed in Chapter 6. Signalling via all three RhoGTPases (Rac1, cdc42 and RhoA) known to be activated by Trio RhoGEF is closely related to both the MAPK and Focal Adhesion pathways (see Figures 4.13 and 4.14). The Rho signalling pathway was significantly enriched among amplified genes in LMS (Table 4.8) suggesting that activation of Rho signalling may be common to both UPS and LMS subtypes, and possibly a wider range of STS as suggested by *in vitro* studies on amoeboid non-human sarcoma cells (Kosla, Pankova et al. 2013). However, the subtype-specificity of Trio RhoGEF overexpression demonstrated in this study (Section 6.2.2.3) leads to the inference that this mechanism of Rho signalling pathway activation is specific to the UPS subtype while LMS may have an alternative mechanism of achieving the same and contribution to tumour progression.

7.3.4 Do LMS and UPS have similar origins?

Considerable overlap between the candidate tumour initiation and progression pathways identified in UPS and LMS in this study (Figure 7.2) are keeping with a similar origin for both tumours as suggested by the differentiation theory for sarcomagenesis. That the two STS subtypes may represent a spectrum of mesenchymal tumours with different degrees of differentiation is further supported by candidates in UPS such as *JUN* and *VGLL3* that promote mesenchymal de-differentiation; and *MYOCD* and *MYH11* in LMS, which promote smooth muscle differentiation. While this preliminary study does not exclude the possibility that the two subtypes are completely distinct in origin and biology, the proposed spectrum could include yet other STS with complex karyotypes and would warrant further exploration within a mesenchymal stem cell context.

7.4 Future Work

This study has identified several potentially important candidate genes, as well as relevant disease models to facilitate even more detailed exploration of pathogenetic abnormalities in different STS subtypes. While data mining can help overcome the limitation of sample size and add statistical power, a further step for identification of even stronger pathogenetic candidates is the application of next-generation exome or whole genome sequencing. Able to provide additional genomic information such as gene mutations and actual SCNA breakpoints, the potential of their application to STS when combined with robust statistical data analysis is likely to give rise to even stronger candidate genes with exciting potential implications for sarcoma biology and therapy. As demonstrated in Chapter 6, actual visualisation of copy number abnormalities aids confirmation of SCNA but in addition shows the patterns of amplification or deletion that gives insights to pathogenesis (e.g. a possible role for chromothripsis) and may be exploited for its diagnostic potential.

It is important to correlate genomic copy number abnormalities to actual gene expression, more so in the context of molecular pathway abnormalities. This is because gene expression is highly complex and regulated by various genetic and epigenetic factors in addition to genomic copy number. Whole transcriptome analysis carried out on the same tumours would be ideal for correlation analysis. However, in spite of recent technological developments such analysis on old FFPE samples remains technically challenging. Again, data mining for expression microarray data on a similar set of STS samples would circumvent this problem. Based on the findings from this study, focused protein expression analysis using pathway-specific protein arrays on cell lines or fresh tissue or immunohistochemistry for specific proteins, as demonstrated in Chapter 6 would help correlate copy number to expression and validate candidate genes and pathway so far highlighted.

Availability of an *in vitro* disease model in the form of long term cell lines for both UPS and LMS subtypes with potential for tumour xenograft development will ultimately make functional analysis by single or multiple pathway inhibition feasible. With some of the methods already optimised and initial pilot data generated for a sample candidate gene, further functional analysis of target genes and pathways can be carried out within a relatively short time. Overall, the results from this PhD study are encouraging and set up exciting potential for future research into sarcoma pathogenesis.

REFERENCES

Abbott, J. J., M. Erickson-Johnson, X. Wang, A. G. Nascimento and A. M. Oliveira (2006). "Gains of COL1A1-PDGFB genomic copies occur in fibrosarcomatous transformation of dermatofibrosarcoma protuberans." Mod Pathol **19**(11): 1512-1518.

Abraham, M. T., M. A. Kuriakose, P. G. Sacks, H. Yee, L. Chiriboga, E. L. Bearer and M. D. Delacure (2001). "Motility-related proteins as markers for head and neck squamous cell cancer." Laryngoscope **111**(7): 1285-1289.

Adamowicz, M., B. Radlwimmer, R. J. Rieker, D. Mertens, M. Schwarzbach, P. Schraml, A. Benner, P. Lichter, G. Mechtersheimer and S. Joos (2006). "Frequent amplifications and abundant expression of TRIO, NKD2, and IRX2 in soft tissue sarcomas." Genes Chromosomes Cancer **45**(9): 829-838.

Albertson, D. G., C. Collins, F. McCormick and J. W. Gray (2003). "Chromosome aberrations in solid tumors." Nat Genet **34**(4): 369-376.

Albini, A., Y. Iwamoto, H. K. Kleinman, G. R. Martin, S. A. Aaronson, J. M. Kozlowski and R. N. McEwan (1987). "A rapid in vitro assay for quantitating the invasive potential of tumor cells." Cancer Res **47**(12): 3239-3245.

Alers, J. C., J. Rochat, P. J. Krijtenburg, H. van Dekken, A. K. Raap and C. Rosenberg (1999). "Universal linkage system: an improved method for labeling archival DNA for comparative genomic hybridization." Genes Chromosomes Cancer **25**(3): 301-305.

Alessio, N., W. Bohn, V. Rauchberger, F. Rizzolio, M. Cipollaro, M. Rosemann, M. Irmler, J. Beckers, A. Giordano and U. Galderisi (2013). "Silencing of RB1 but not of RB2/P130 induces cellular senescence and impairs the differentiation potential of human mesenchymal stem cells." Cell Mol Life Sci **70**(9): 1637-1651.

American Type Culture Collection Standards Development Organization Workgroup, A. S. N. (2010). "Cell line misidentification: the beginning of the end." Nat Rev Cancer **10**(6): 441-448.

Anderson, J., A. Gordon, A. McManus, J. Shipley and K. Pritchard-Jones (1999). "Disruption of imprinted genes at chromosome region 11p15.5 in paediatric rhabdomyosarcoma." Neoplasia **1**(4): 340-348.

Anderson, J., A. Gordon, K. Pritchard-Jones and J. Shipley (1999). "Genes, chromosomes, and rhabdomyosarcoma." Genes Chromosomes Cancer **26**(4): 275-285.

Anderson, K., C. Lutz, F. W. van Delft, C. M. Bateman, Y. Guo, S. M. Colman, H. Kempski, A. V. Moorman, I. Titley, J. Swansbury, L. Kearney, T. Enver and M. Greaves (2011). "Genetic variegation of clonal architecture and propagating cells in leukaemia." Nature **469**(7330): 356-361.

Antonescu, C. R., P. Dal Cin, K. Nafa, L. A. Teot, U. Surti, C. D. Fletcher and M. Ladanyi (2007). "EWSR1-CREB1 is the predominant gene fusion in angiomatoid fibrous histiocytoma." Genes Chromosomes Cancer **46**(12): 1051-1060.

Appella, E. and C. W. Anderson (2001). "Post-translational modifications and activation of p53 by genotoxic stresses." Eur J Biochem **268**(10): 2764-2772.

Armstrong, M. D., D. Von Hoff, B. Barber, L. A. Marlow, C. von Roemeling, S. J. Cooper, P. Travis, E. Campbell, R. Paz-Fumagalli, J. A. Copland and G. Colon-Otero (2011). "An effective personalized approach to a rare tumor: prolonged survival in metastatic pancreatic acinar cell carcinoma based on genetic analysis and cell line development." J Cancer **2**: 142-152.

Arnould, L., P. Arveux, J. Couturier, M. Gelly-Marty, C. Loustalot, F. Ettore, C. Sagan, M. Antoine, F. Penault-Llorca, B. Vasseur, P. Fumoleau and B. P. Coudert (2007). "Pathologic complete response to trastuzumab-based neoadjuvant therapy is related to the level of HER-2 amplification." Clin Cancer Res **13**(21): 6404-6409.

Artandi, S. E., S. Chang, S.-L. Lee, S. Alson, G. J. Gottlieb, L. Chin and R. A. DePinho (2000). "Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice." Nature **406**(6796): 641-645.

Artandi, S. E. and R. A. DePinho (2000). "A critical role for telomeres in suppressing and facilitating carcinogenesis." Current Opinion in Genetics & Development **10**(1): 39-46.

Baldwin, C., C. Garnis, L. Zhang, M. P. Rosin and W. L. Lam (2005). "Multiple microalterations detected at high frequency in oral cancer." Cancer Res **65**(17): 7561-7567.

Barretina, J., G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehar, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jane-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, Jr., M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel and L. A. Garraway (2012). "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity." Nature **483**(7391): 603-607.

Barretina, J., B. S. Taylor, S. Banerji, A. H. Ramos, M. Lagos-Quintana, P. L. Decarolis, K. Shah, N. D. Socci, B. A. Weir, A. Ho, D. Y. Chiang, B. Reva, C. H. Mermel, G. Getz, Y. Antipin, R. Beroukhim, J. E. Major, C. Hatton, R. Nicoletti, M. Hanna, T. Sharpe, T. J. Fennell, K. Cibulskis, R. C. Onofrio, T. Saito, N. Shukla, C. Lau, S. Nelander, S. J. Silver, C. Sougnez, A. Viale, W. Winckler, R. G. Maki, L. A. Garraway, A. Lash, H. Greulich, D. E. Root, W. R. Sellers, G. K. Schwartz, C. R. Antonescu, E. S. Lander, H. E. Varmus, M. Ladanyi, C. Sander, M. Meyerson and S. Singer (2010). "Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy." Nat Genet **42**(8): 715-721.

Barretina, J., B. S. Taylor, S. Banerji, A. H. Ramos, M. Lagos-Quintana, P. L. Decarolis, K. Shah, N. D. Socci, B. A. Weir, A. Ho, D. Y. Chiang, B. Reva, C. H. Mermel, G. Getz, Y. Antipin, R. Beroukhim, J. E. Major, C. Hatton, R. Nicoletti, M. Hanna, T. Sharpe, T. J. Fennell, K. Cibulskis, R. C. Onofrio, T. Saito, N. Shukla, C. Lau, S. Nelander, S. J. Silver, C. Sougnez, A. Viale, W. Winckler, R. G. Maki, L. A. Garraway, A. Lash, H. Greulich, D. E. Root, W. R. Sellers, G. K. Schwartz, C. R. Antonescu, E. S. Lander, H. E. Varmus, M. Ladanyi, C. Sander, M. Meyerson and S. Singer (2010). "Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy." Nat Genet.

Barrett, M. T., A. Scheffer, A. Ben-Dor, N. Sampas, D. Lipson, R. Kincaid, P. Tsang, B. Curry, K. Baird, P. S. Meltzer, Z. Yakhini, L. Bruhn and S. Laderman (2004). "Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA." Proc Natl Acad Sci U S A **101**(51): 17765-17770.

Barrio-Real, L. and M. G. Kazanietz (2012). "Rho GEFs and cancer: linking gene expression and metastatic dissemination." Sci Signal **5**(244): pe43.

Begus-Nahrmann, Y., D. Hartmann, J. Kraus, P. Eshraghi, A. Scheffold, M. Grieb, V. Rasche, P. Schirmacher, H. W. Lee, H. A. Kestler, A. Lechel and K. L. Rudolph (2012). "Transient telomere dysfunction induces chromosomal instability and promotes carcinogenesis." J Clin Invest **122**(6): 2283-2288.

Benini, S., M. C. Manara, V. Cerisano, S. Perdichizzi, R. Strammiello, M. Serra, P. Picci and K. Scotlandi (2004). "Contribution of MEK/MAPK and PI3-K signaling pathway to the malignant behavior of Ewing's sarcoma cells: therapeutic prospects." Int J Cancer **108**(3): 358-366.

Beroukhim, R., G. Getz, L. Nghiemphu, J. Barretina, T. Hsueh, D. Linhart, I. Vivanco, J. C. Lee, J. H. Huang, S. Alexander, J. Du, T. Kau, R. K. Thomas, K. Shah, H. Soto, S. Perner, J. Prensner, R. M. Debiasi, F. Demichelis, C. Hatton, M. A. Rubin, L. A. Garraway, S. F. Nelson, L. Liau, P. S. Mischel, T. F. Cloughesy, M. Meyerson, T. A. Golub, E. S. Lander, I. K. Mellinghoff and W. R. Sellers (2007). "Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma." Proc Natl Acad Sci U S A **104**(50): 20007-20012.

Beroukhim, R., C. H. Mermel, D. Porter, G. Wei, S. Raychaudhuri, J. Donovan, J. Barretina, J. S. Boehm, J. Dobson, M. Urashima, K. T. Mc Henry, R. M. Pinchback, A. H. Ligon, Y. J. Cho, L. Haery, H. Greulich, M. Reich, W. Winckler, M. S. Lawrence, B. A. Weir, K. E. Tanaka, D. Y. Chiang, A. J. Bass, A. Loo, C. Hoffman, J. Prensner, T. Liefeld, Q. Gao, D. Yecies, S. Signoretti, E. Maher, F. J. Kaye, H. Sasaki, J. E. Tepper, J. A. Fletcher, J. Tabernero, J. Baselga, M. S. Tsao, F. Demichelis, M. A. Rubin, P. A. Janne, M. J. Daly, C. Nucera, R. L. Levine, B. L. Ebert, S. Gabriel, A. K. Rustgi, C. R. Antonescu, M. Ladanyi, A. Letai, L. A. Garraway, M. Loda, D. G. Beer, L. D. True, A. Okamoto, S. L. Pomeroy, S. Singer, T. R. Golub, E. S. Lander, G. Getz, W. R. Sellers and M. Meyerson (2010). "The landscape of somatic copy-number alteration across human cancers." Nature **463**(7283): 899-905.

Bhatia, N., J. R. Herter, T. J. Slaga, S. Y. Fuchs and V. S. Spiegelman (2002). "Mouse homologue of HOS (mHOS) is overexpressed in skin tumors and implicated in constitutive activation of NF-kappaB." Oncogene **21**(10): 1501-1509.

Bignell, G. R., C. D. Greenman, H. Davies, A. P. Butler, S. Edkins, J. M. Andrews, G. Buck, L. Chen, D. Beare, C. Latimer, S. Widaa, J. Hinton, C. Fahey, B. Fu, S. Swamy, G. L. Dalgliesh, B. T. Teh, P. Deloukas, F. Yang, P. J. Campbell, P. A. Futreal and M. R. Stratton (2010). "Signatures of mutation and selection in the cancer genome." Nature **463**(7283): 893-898.

Billing, V., F. Mertens, H. A. Domanski and A. Rydholm (2008). "Deep-seated ordinary and atypical lipomas: HISTOPATHOLOGY, CYTOGENETICS, CLINICAL FEATURES, AND OUTCOME IN 215 TUMOURS OF THE EXTREMITY AND TRUNK WALL." Journal of Bone and Joint Surgery - British Volume **90-B**(7): 929-933.

Bjerkehagen, B., M. C. Smastuen, K. S. Hall, S. Skjeldal, S. Smeland and S. D. Fossa (2012). "Why do patients with radiation-induced sarcomas have a poor sarcoma-related survival?" Br J Cancer **106**(2): 297-306.

Blum, J. M., L. Ano, Z. Li, D. Van Mater, B. D. Bennett, M. Sachdeva, I. Lagutina, M. Zhang, J. K. Mito, L. G. Dodd, D. M. Cardona, R. D. Dodd, N. Williams, Y. Ma, C. Lepper, C. M. Linardic, S. Mukherjee, G. C. Grosveld, C. M. Fan and D. G. Kirsch (2013). "Distinct and overlapping sarcoma subtypes initiated from muscle stem and progenitor cells." Cell Rep **5**(4): 933-940.

Boland, C. R. and A. Goel (2010). "Microsatellite instability in colorectal cancer." Gastroenterology **138**(6): 2073-2087 e2073.

Bos, J. L., H. Rehmann and A. Wittinghofer (2007). "GEFs and GAPs: critical elements in the control of small G proteins." Cell **129**(5): 865-877.

Bougeard, G., R. Sesboue, S. Baert-Desurmont, S. Vasseur, C. Martin, J. Tinat, L. Brugieres, A. Chompret, B. B. de Paillerets, D. Stoppa-Lyonnet, C. Bonaiti-Pellie and T. Frebourg (2008). "Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families." J Med Genet **45**(8): 535-538.

Bouquier, N., S. Fromont, J. C. Zeeh, C. Auziol, P. Larrousse, B. Robert, M. Zeghouf, J. Cherfils, A. Debant and S. Schmidt (2009). "Aptamer-derived peptides as potent inhibitors of the oncogenic RhoGEF Tgat." Chem Biol **16**(4): 391-400.

Bouquier, N., E. Vignal, S. Charrasse, M. Weill, S. Schmidt, J. P. Leonetti, A. Blangy and P. Fort (2009). "A cell active chemical GEF inhibitor selectively targets the Trio/RhoG/Rac1 signaling pathway." Chem Biol **16**(6): 657-666.

Bovée, J. and P. Hogendoorn (2009). "Molecular pathology of sarcomas: concepts and clinical implications." Virchows Archiv **456**(2): 193-199.

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

Boveri, T. (2008). "Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris." J Cell Sci **121 Suppl 1**: 1-84.

Brabletz, T. (2012). "To differentiate or not — routes towards metastasis." Nat Rev Cancer **12**(6): 425-436.

Braggio, E., E. R. McPhail, W. Macon, M. B. Lopes, D. Schiff, M. Law, S. Fink, D. Sprau, C. Giannini, A. Dogan, R. Fonseca and B. P. O'Neill (2011). "Primary central nervous system lymphomas: a validation study of array-based comparative genomic hybridization in formalin-fixed paraffin-embedded tumor specimens." Clin Cancer Res **17**(13): 4245-4253.

Brecht, M., A. C. Steenvoorden, J. G. Collard, S. Luf, D. Erz, C. R. Bartram and J. W. Janssen (2005). "Activation of gef-h1, a guanine nucleotide exchange factor for RhoA, by DNA transfection." Int J Cancer **113**(4): 533-540.

Brouquet, A., P. Taleb, A. S. Lot, A. Beauchet, C. Julie, G. Prevost, B. Nordlinger and C. Penna (2011). "A model of primary culture of colorectal cancer and liver metastasis to predict chemosensitivity." J Surg Res **166**(2): 247-254.

Capes-Davis, A., Y. A. Reid, M. C. Kline, D. R. Storts, E. Strauss, W. G. Dirks, H. G. Drexler, R. A. MacLeod, G. Sykes, A. Kohara, Y. Nakamura, E. Elmore, R. W. Nims, C. Alston-Roberts, R. Barallon, G. V. Los, R. M. Nardone, P. J. Price, A. Steuer, J. Thomson, J. R. Masters and L. Kerrigan (2013). "Match criteria for human cell line authentication: where do we draw the line?" Int J Cancer **132**(11): 2510-2519.

Carneiro, A., P. Francis, P.-O. Bendahl, J. Fernebro, M. Akerman, C. Fletcher, A. Rydholm, A. Borg and M. Nilbert (2009). "Indistinguishable genomic profiles and shared prognostic markers in undifferentiated pleomorphic sarcoma and leiomyosarcoma: different sides of a single coin[quest]." Lab Invest **89**(6): 668-675.

Casali, P. G., L. Jost, S. Sleijfer, J. Verweij, J. Y. Blay and E. G. W. Group (2008). "Soft tissue sarcomas: ESMO clinical recommendations for diagnosis, treatment and follow-up." Ann Oncol **19 Suppl 2**: ii89-93.

Chan, A. Y., S. J. Coniglio, Y. Y. Chuang, D. Michaelson, U. G. Knaus, M. R. Philips and M. Symons (2005). "Roles of the Rac1 and Rac3 GTPases in human tumor cell invasion." Oncogene **24**(53): 7821-7829.

Chandhanayingyong, C., Y. Kim, J. R. Staples, C. Hahn and F. Y. Lee (2012). "MAPK/ERK Signaling in Osteosarcomas, Ewing Sarcomas and Chondrosarcomas: Therapeutic Implications and Future Directions." Sarcoma **2012**: 404810.

Chang, F. T., J. D. McGhie, F. L. Chan, M. C. Tang, M. A. Anderson, J. R. Mann, K. H. Andy Choo and L. H. Wong (2013). "PML bodies provide an important platform for the maintenance of telomeric chromatin integrity in embryonic stem cells." Nucleic Acids Res **41**(8): 4447-4458.

Chang, Y., E. Cesarman, M. S. Pessin, F. Lee, J. Culpepper, D. M. Knowles and P. S. Moore (1994). "Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma." Science **266**(5192): 1865-1869.

Chattopadhyay, I., A. Singh, R. Phukan, J. Purkayastha, A. Kataki, J. Mahanta, S. Saxena and S. Kapur (2010). "Genome-wide analysis of chromosomal alterations in patients with esophageal squamous cell carcinoma exposed to tobacco and betel quid from high-risk area in India." Mutation Research-Genetic Toxicology and Environmental Mutagenesis **696**(2): 130-138.

Chen, Z., Z. Liu, W. Li, K. Qu, X. Deng, M. G. Varma, A. Fichera, A. Pigazzi and J. Garcia-Aguilar (2011). "Chromosomal copy number alterations are associated with tumor response to chemoradiation in locally advanced rectal cancer." Genes Chromosomes Cancer **50**(9): 689-699.

Cheung, L. W., B. T. Hennessy, J. Li, S. Yu, A. P. Myers, B. Djordjevic, Y. Lu, K. Stemke-Hale, M. D. Dyer, F. Zhang, Z. Ju, L. C. Cantley, S. E. Scherer, H. Liang, K. H. Lu, R. R. Broaddus and G. B. Mills (2011). "High frequency of PIK3R1 and PIK3R2 mutations in endometrial cancer elucidates a novel mechanism for regulation of PTEN protein stability." Cancer Discov **1**(2): 170-185.

Chitale, D., Y. Gong, B. S. Taylor, S. Broderick, C. Brennan, R. Somwar, B. Golas, L. Wang, N. Motoi, J. Szoke, J. M. Reinersman, J. Major, C. Sander, V. E. Seshan, M. F. Zakowski, V. Rusch, W. Pao, W. Gerald and M. Ladanyi (2009). "An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors." Oncogene **28**(31): 2773-2783.

Clark, J., P. J. Rocques, A. J. Crew, S. Gill, J. Shipley, A. M. Chan, B. A. Gusterson and C. S. Cooper (1994). "Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma." Nat Genet **7**(4): 502-508.

Clark, M. A., C. Fisher, I. Judson and J. M. Thomas (2005). "Soft-Tissue Sarcomas in Adults." New England Journal of Medicine **353**(7): 701-711.

Coe, B. R., L. J. Henderson, C. Garnis, M. S. Tsao, A. F. Gazdar, J. Minna, S. Lam, C. MacAulay and W. L. Lam (2005). "High-resolution chromosome arm 5p array CGH analysis of small cell lung carcinoma cell lines." Genes Chromosomes & Cancer **42**(3): 308-313.

Coindre, J. M. (2003). "Immunohistochemistry in the diagnosis of soft tissue tumours." Histopathology **43**(1): 1-16.

Coindre, J. M. (2006). "Grading of soft tissue sarcomas: review and update." Arch Pathol Lab Med **130**(10): 1448-1453.

Coindre, J. M., O. Mariani, F. Chibon, A. Mairal, N. De Saint Aubain Somerhausen, E. Favre-Guillevin, N. B. Bui, E. Stoeckle, I. Hostein and A. Aurias (2003). "Most malignant fibrous histiocytomas developed in the retroperitoneum are dedifferentiated liposarcomas: a review of 25 cases initially diagnosed as malignant fibrous histiocytoma." Mod Pathol **16**(3): 256-262.

Coindre, J. M., B. B. Nguyen, F. Bonichon, I. de Mascarel and M. Trojani (1988). "Histopathologic grading in spindle cell soft tissue sarcomas." Cancer **61**(11): 2305-2309.

Coindre, J. M., F. Pedeutour and A. Aurias (2009). "Well-differentiated and dedifferentiated liposarcomas." Virchows Arch **456**(2): 167-179.

Coindre, J. M., P. Terrier, N. B. Bui, F. Bonichon, F. Collin, V. Le Doussal, A. M. Mandard, M. O. Vilain, J. Jacquemier, H. Duplay, X. Sastre, C. Barlier, M. Henry-Amar, J. Macé-Lesech and G. Contesso (1996). "Prognostic factors in adult patients with locally controlled soft tissue sarcoma. A study of 546 patients from the French Federation of Cancer Centers Sarcoma Group." Journal of Clinical Oncology **14**(3): 869-877.

Conrad, D. F., D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T. D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C. H. Ihm, K. Kristiansson, D. G. Macarthur, J. R. Macdonald, I. Onyiah, A. W. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, C. Wellcome Trust Case Control, C. Tyler-Smith, N. P. Carter, C. Lee, S. W. Scherer and M. E. Hurles (2010). "Origins and functional impact of copy number variation in the human genome." Nature **464**(7289): 704-712.

Cormier, J. N. and R. E. Pollock (2004). "Soft Tissue Sarcomas." CA: A Cancer Journal for Clinicians **54**(2): 94-109.

Cree, I. A., S. Glaysher and A. L. Harvey (2010). "Efficacy of anti-cancer agents in cell lines versus human primary tumour tissue." Current Opinion in Pharmacology **10**(4): 375-379.

Cully, M., H. You, A. J. Levine and T. W. Mak (2006). "Beyond PTEN mutations: the PI3K pathway as an integrator of multiple inputs during tumorigenesis." Nat Rev Cancer **6**(3): 184-192.

Dai, Y. and S. Grant (2010). "New insights into checkpoint kinase 1 in the DNA damage response signaling network." Clin Cancer Res **16**(2): 376-383.

Daley, G. Q., R. A. Van Etten and D. Baltimore (1990). "Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome." Science **247**(4944): 824-830.

Dangles-Marie, V., M. Pocard, S. Richon, L. B. Weiswald, F. Assayag, P. Saulnier, J. G. Judde, J. L. Janneau, N. Auger, P. Validire, B. Dutrillaux, F. Praz, D. Bellet and M. F. Poupon (2007). "Establishment of human colon cancer cell lines from fresh tumors versus xenografts: comparison of success rate and cell line features." Cancer Res **67**(1): 398-407.

Davoli, T. and T. de Lange (2011). "The Causes and Consequences of Polyploidy in Normal Development and Cancer." Annual Review of Cell and Developmental Biology **27**(1): 585-610.

de Bruijn, D. R., J. P. Nap and A. G. van Kessel (2007). "The (epi)genetics of human synovial sarcoma." Genes Chromosomes Cancer **46**(2): 107-117.

de Jong, D., S. Verbeke, D. Meijer, P. C. Hogendoorn, J. V. Bovee and K. Szuhai (2011). "Opening the archives for state of the art tumour genetic research: sample processing for array-CGH using decalcified, formalin-fixed, paraffin-embedded tissue-derived DNA samples." BMC Res Notes **4**: 1.

de Leeuw, B., M. Balemans, D. Olde Weghuis and A. Geurts van Kessel (1995). "Identification of two alternative fusion genes, SYT-SSX1 and SYT-SSX2, in t(X;18)(p11.2;q11.2)-positive synovial sarcomas." Hum Mol Genet **4**(6): 1097-1099.

de Vreeze, R. S. A., D. de Jong, R. L. Haas, F. Stewart and F. van Coevorden (2008). "Effectiveness of Radiotherapy in Myxoid Sarcomas Is Associated With a Dense Vascular Pattern." International journal of radiation oncology, biology, physics **72**(5): 1480-1487.

Deininger, M., E. Buchdunger and B. J. Druker (2005). "The development of imatinib as a therapeutic agent for chronic myeloid leukemia." Blood **105**(7): 2640-2653.

Derre, J., R. Lagace, A. Nicolas, A. Mairal, F. Chibon, J. M. Coindre, P. Terrier, X. Sastre and A. Aurias (2001). "Leiomyosarcomas and most malignant fibrous histiocytomas share very similar comparative genomic hybridization imbalances: an analysis of a series of 27 leiomyosarcomas." Lab Invest **81**(2): 211-215.

Dhillon, A. S., S. Hagan, O. Rath and W. Kolch (2007). "MAP kinase signalling pathways in cancer." Oncogene **26**(22): 3279-3290.

Di Fiore, R., A. D'Anneo, G. Tesoriere and R. Vento (2013). "RB1 in cancer: Different mechanisms of RB1 inactivation and alterations of pRb pathway in tumorigenesis." Journal of Cellular Physiology **228**(8): 1676-1687.

Diskin, S. J., T. Eck, J. Greshock, Y. P. Mosse, T. Naylor, C. J. Stoeckert, Jr., B. L. Weber, J. M. Maris and G. R. Grant (2006). "STAC: A method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments." Genome Res **16**(9): 1149-1158.

Dobin, S. M., J. A. Diaz, M. T. Silva, L. M. Truss and L. R. Donner (1999). "Translocation (17;22)(q22;q13) in a case of subcutaneous dermatofibrosarcoma protuberans in an adult." Cancer Genet Cytogenet **109**(1): 86-87.

Dodd, R. D., J. K. Mito, W. C. Eward, R. Chitalia, M. Sachdeva, Y. Ma, J. Barretina, L. Dodd and D. G. Kirsch (2013). "NF1 deletion generates multiple subtypes of soft-tissue sarcoma that respond to MEK inhibition." Mol Cancer Ther **12**(9): 1906-1917.

Downward, J. (2003). "Targeting RAS signalling pathways in cancer therapy." Nat Rev Cancer **3**(1): 11-22.

Draviam, V. M., F. Stegmeier, G. Nalepa, M. E. Sowa, J. Chen, A. Liang, G. J. Hannon, P. K. Sorger, J. W. Harper and S. J. Elledge (2007). "A functional genomic screen identifies a role for TAO1 kinase in spindle-checkpoint signalling." Nat Cell Biol **9**(5): 556-564.

Eder, A. M., X. Sui, D. G. Rosen, L. K. Nolden, K. W. Cheng, J. P. Lahad, M. Kango-Singh, K. H. Lu, C. L. Warneke, E. N. Atkinson, I. Bedrosian, K. Keyomarsi, W.-l. Kuo, J. W. Gray, J. C. P. Yin, J. Liu, G. Halder and G. B. Mills (2005). "Atypical PKCι contributes to poor prognosis through loss of apical-

basal polarity and Cyclin E overexpression in ovarian cancer." Proceedings of the National Academy of Sciences of the United States of America **102**(35): 12519-12524.

Fabbro, M., B. B. Zhou, M. Takahashi, B. Sarcevic, P. Lal, M. E. Graham, B. G. Gabrielli, P. J. Robinson, E. A. Nigg, Y. Ono and K. K. Khanna (2005). "Cdk1/Erk2- and Plk1-dependent phosphorylation of a centrosome protein, Cep55, is required for its recruitment to midbody and cytokinesis." Dev Cell **9**(4): 477-488.

Fingerhut, M. A., W. E. Halperin, D. A. Marlow, L. A. Piacitelli, P. A. Honchar, M. H. Sweeney, A. L. Greife, P. A. Dill, K. Steenland and A. J. Suruda (1991). "Cancer mortality in workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin." N Engl J Med **324**(4): 212-218.

Fisher, C. (2011). "Immunohistochemistry in diagnosis of soft tissue tumours." Histopathology **58**(7): 1001-1012.

Fletcher, C. D. (2014). "The evolving classification of soft tissue tumours - an update based on the new 2013 WHO classification." Histopathology **64**(1): 2-11.

Fletcher, C. D. M., J. J. Berman, C. Corless, F. Gorstein, J. Lasota, B. J. Longley, M. Miettinen, T. J. O'Leary, H. Remotti, B. P. Rubin, B. Shmookler, L. H. Sobin and S. W. Weiss (2002). "Diagnosis of gastrointestinal stromal tumors: A consensus approach." Human Pathology **33**(5): 459-465.

Fletcher, C. D. M., J. A. Bridge, P. C. W. Hogendoorn and F. Mertens (2013). Classification of Tumours of Soft Tissue and Bone Classification of Tumours, World Health Organization.

Fletcher, C. D. M., K. K. Unni, F. Mertens and (Eds.) (2002). World Health Organization Classification of Tumours: Pathology and genetics of tumours of soft tissue and bone. Lyon, IARC Press.

Fong, Y., D. G. Coit, J. M. Woodruff and M. F. Brennan (1993). "Lymph node metastasis from soft tissue sarcoma in adults. Analysis of data from a prospective database of 1772 sarcoma patients." Ann Surg **217**(1): 72-77.

Forbes, S. A., N. Bindal, S. Bamford, C. Cole, C. Y. Kok, D. Beare, M. Jia, R. Shepherd, K. Leung, A. Menzies, J. W. Teague, P. J. Campbell, M. R. Stratton and P. A. Futreal (2011). "COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer." Nucleic Acids Res **39**(Database issue): D945-950.

Francis, M., N. Dennis, J. Charman, G. Lawrence and R. Grimer (2013) "Bone and Soft Tissue Sarcomas. UK Incidence and Survival: 1996 to 2010."

Fritz, G., C. Brachetti, F. Bahlmann, M. Schmidt and B. Kaina (2002). "Rho GTPases in human breast tumours: expression and mutation analyses and correlation with clinical parameters." Br J Cancer **87**(6): 635-644.

Fritz, G., I. Just and B. Kaina (1999). "Rho GTPases are over-expressed in human tumors." International Journal of Cancer **81**(5): 682-687.

Frolov, A., S. Chahwan, M. Ochs, J. P. Arnoletti, Z.-Z. Pan, O. Favorova, J. Fletcher, M. von Mehren, B. Eisenberg and A. K. Godwin (2003). "Response Markers and the Molecular Mechanisms of Action of Gleevec in Gastrointestinal Stromal Tumors1." Molecular Cancer Therapeutics **2**(8): 699-709.

Gabriel, B., D. C. Fischer, M. Orlowska-Volk, A. zur Hausen, R. Schule, J. M. Muller and A. Hasenburg (2006). "Expression of the transcriptional coregulator FHL2 in human breast cancer: a clinicopathologic study." J Soc Gynecol Investig **13**(1): 69-75.

Gabriel, B., S. Mildenberger, C. W. Weisser, E. Metzger, G. Gitsch, R. Schule and J. M. Muller (2004). "Focal adhesion kinase interacts with the transcriptional coactivator FHL2 and both are overexpressed in epithelial ovarian cancer." Anticancer Res **24**(2B): 921-927.

Garnett, M. J., E. J. Edelman, S. J. Heidorn, C. D. Greenman, A. Dastur, K. W. Lau, P. Greninger, I. R. Thompson, X. Luo, J. Soares, Q. Liu, F. Iorio, D. Surdez, L. Chen, R. J. Milano, G. R. Bignell, A. T. Tam, H. Davies, J. A. Stevenson, S. Barthorpe, S. R. Lutz, F. Kogera, K. Lawrence, A. McLaren-Douglas, X. Mitropoulos, T. Mironenko, H. Thi, L. Richardson, W. Zhou, F. Jewitt, T. Zhang, P. O'Brien, J. L. Boisvert, S. Price, W. Hur, W. Yang, X. Deng, A. Butler, H. G. Choi, J. W. Chang, J. Baselga, I. Stamenkovic, J. A. Engelman, S. V. Sharma, O. Delattre, J. Saez-Rodriguez, N. S. Gray, J. Settleman, P. A. Futreal, D. A. Haber, M. R. Stratton, S. Ramaswamy, U. McDermott and C. H. Benes (2012). "Systematic identification of genomic markers of drug sensitivity in cancer cells." Nature **483**(7391): 570-575.

Gazdar, A. F., B. Gao and J. D. Minna (2010). "Lung cancer cell lines: Useless artifacts or invaluable tools for medical science?" Lung Cancer **68**(3): 309-318.

Gazdar, A. F., V. Kurvari, A. Virmani, L. Gollahon, M. Sakaguchi, M. Westerfield, D. Kodagoda, V. Stasny, H. T. Cunningham, Wistuba, II, G. Tomlinson, V. Tonk, R. Ashfaq, A. M. Leitch, J. D. Minna and J. W. Shay (1998). "Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer." Int J Cancer **78**(6): 766-774.

Genini, M., P. Schwalbe, F. A. Scholl, A. Remppis, M. G. Mattei and B. W. Schafer (1997). "Subtractive cloning and characterization of DRAL, a novel LIM-domain protein down-regulated in rhabdomyosarcoma." DNA Cell Biol **16**(4): 433-442.

Giancotti, F. G. and E. Ruoslahti (1999). "Integrin signaling." Science **285**(5430): 1028-1032.

Gillet, J. P., S. Varma and M. M. Gottesman (2013). "The clinical relevance of cancer cell lines." J Natl Cancer Inst **105**(7): 452-458.

Gisselsson, D., L. Pettersson, M. Hoglund, M. Heidenblad, L. Gorunova, J. Wiegant, F. Mertens, P. Dal Cin, F. Mitelman and N. Mandahl (2000). "Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity." Proc Natl Acad Sci U S A **97**(10): 5357-5362.

Gladdy, R. A., L. X. Qin, N. Moraco, M. A. Edgar, C. R. Antonescu, K. M. Alektiar, M. F. Brennan and S. Singer (2010). "Do radiation-associated soft tissue sarcomas have the same prognosis as sporadic soft tissue sarcomas?" J Clin Oncol **28**(12): 2064-2069.

Gondek, L. P., R. Tiu, C. L. O'Keefe, M. A. Sekeres, K. S. Theil and J. P. Maciejewski (2008). "Chromosomal lesions and uniparental disomy detected by SNP arrays in MDS, MDS/MPD, and MDS-derived AML." Blood **111**(3): 1534-1542.

Gonin-Laurent, N., N. S. Hadj-Hamou, N. Vogt, C. Houdayer, M. Gauthiers-Villars, C. Dehainault, X. Sastre-Garau, S. Chevillard and B. Malfoy (2007). "RB1 and TP53 pathways in radiation-induced sarcomas." Oncogene **26**(41): 6106-6112.

Gonzalez-Vasconcellos, I., N. Anastasov, B. Sanli-Bonazzi, O. Klymenko, M. J. Atkinson and M. Rosemann (2013). "Rb1 haploinsufficiency promotes telomere attrition and radiation-induced genomic instability." Cancer Res **73**(14): 4247-4255.

Graham, C., S. Chilton-MacNeill, M. Zielenska and G. R. Somers (2012). "The CIC-DUX4 fusion transcript is present in a subgroup of pediatric primitive round cell sarcomas." Human Pathology **43**(2): 180-189.

Greaves, M. and C. C. Maley (2012). "Clonal evolution in cancer." Nature **481**(7381): 306-313.

Grignani, G., E. Palmerini, P. Dileo, S. D. Asaftei, L. D'Ambrosio, Y. Pignochino, M. Mercuri, P. Picci, F. Fagioli, P. G. Casali, S. Ferrari and M. Aglietta (2012). "A phase II trial of sorafenib in relapsed and unresectable high-grade osteosarcoma after failure of standard multimodal therapy: an Italian Sarcoma Group study." Ann Oncol **23**(2): 508-516.

Grimer, R., I. Judson, D. Peake and B. Seddon (2010). "Guidelines for the management of soft tissue sarcomas." Sarcoma **2010**: 506182.

Grobmyer, S. R. and M. F. Brennan (2003). "Predictive variables detailing the recurrence rate of soft tissue sarcomas." Curr Opin Oncol **15**(4): 319-326.

Grossel, M. J. and P. W. Hinds (2006). "From cell cycle to differentiation: an expanding role for cdk6." Cell Cycle **5**(3): 266-270.

Guillou, L. and A. Aurias (2009). "Soft tissue sarcomas with complex genomic profiles." Virchows Arch.

Guiu, S., M. Gauthier, B. Coudert, F. Bonnetain, L. Favier, S. Ladoire, H. Tixier, B. Guiu, F. Penault-Llorca, F. Ettore, P. Fumoleau and L. Arnould (2010). "Pathological complete response and survival according to the level of HER-2 amplification after trastuzumab-based neoadjuvant therapy for breast cancer." Br J Cancer **103**(9): 1335-1342.

Gullo, G., D. Bettio, M. Zuradelli, G. Masci, L. Giordano, C. Bareggi, M. Tomirotti, P. Salvini, L. Runza, N. La Verde and A. Santoro (2013). "Level of HER2/neu amplification in primary tumours and metastases in HER2-positive breast cancer and survival after trastuzumab therapy." Breast **22**(2): 190-193.

Guo, W. and F. G. Giancotti (2004). "Integrin signalling during tumour progression." Nat Rev Mol Cell Biol **5**(10): 816-826.

Gupta, G. P. and J. Massague (2006). "Cancer metastasis: building a framework." Cell **127**(4): 679-695.

Halazonetis, T. D., V. G. Gorgoulis and J. Bartek (2008). "An oncogene-induced DNA damage model for cancer development." Science **319**(5868): 1352-1355.

Halperin, D. S., C. Pan, A. J. Lusis and P. Tontonoz (2013). "Vestigial-like 3 is an inhibitor of adipocyte differentiation." Journal of Lipid Research **54**(2): 473-481.

Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of cancer: the next generation." Cell **144**(5): 646-674.

Hart, M. J., A. Eva, T. Evans, S. A. Aaronson and R. A. Cerione (1991). "Catalysis of Guanine-Nucleotide Exchange on the Cdc42hs Protein by the Dbl Oncogene Product." Nature **354**(6351): 311-314.

Hart, M. J., A. Eva, D. Zangrilli, S. A. Aaronson, T. Evans, R. A. Cerione and Y. Zheng (1994). "Cellular transformation and guanine nucleotide exchange activity are catalyzed by a common domain on the dbl oncogene product." J Biol Chem **269**(1): 62-65.

Heaphy, C. M., A. P. Subhawong, S. M. Hong, M. G. Goggins, E. A. Montgomery, E. Gabrielson, G. J. Netto, J. I. Epstein, T. L. Lotan, W. H. Westra, M. Shih Ie, C. A. Iacobuzio-Donahue, A. Maitra, Q. K. Li, C. G. Eberhart, J. M. Taube, D. Rakheja, R. J. Kurman, T. C. Wu, R. B. Roden, P. Argani, A. M. De Marzo, L. Terracciano, M. Torbenson and A. K. Meeker (2011). "Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes." Am J Pathol **179**(4): 1608-1615.

Heemers, H. V., K. M. Regan, S. M. Dehm and D. J. Tindall (2007). "Androgen induction of the androgen receptor coactivator four and a half LIM domain protein-2: evidence for a role for serum response factor in prostate cancer." Cancer Res **67**(21): 10592-10599.

Heinrich, M. C., C. L. Corless, A. Duensing, L. McGreevey, C.-J. Chen, N. Joseph, S. Singer, D. J. Griffith, A. Haley, A. Town, G. D. Demetri, C. D. M. Fletcher and J. A. Fletcher (2003). "PDGFRA Activating Mutations in Gastrointestinal Stromal Tumors." Science **299**(5607): 708-710.

Heisterkamp, N., G. Jenster, J. ten Hoeve, D. Zovich, P. K. Pattengale and J. Groffen (1990). "Acute leukaemia in bcr/abl transgenic mice." Nature **344**(6263): 251-253.

Helias-Rodzewicz, Z., G. Perot, F. Chibon, C. Ferreira, P. Lagarde, P. Terrier, J. M. Coindre and A. Aurias (2010). "YAP1 and VGLL3, encoding two cofactors of TEAD transcription factors, are amplified and overexpressed in a subset of soft tissue sarcomas." Genes Chromosomes Cancer **49**(12): 1161-1171.

Helman, L. J. and P. Meltzer (2003). "Mechanisms of sarcoma development." Nat Rev Cancer **3**(9): 685-694.

Henley, S. A. and F. A. Dick (2012). "The retinoblastoma family of proteins and their regulatory functions in the mammalian cell division cycle." Cell Div **7**(1): 10.

Hernando, E., E. Charytonowicz, M. E. Dudas, S. Menendez, I. Matushansky, J. Mills, N. D. Socci, N. Behrendt, L. Ma, R. G. Maki, P. P. Pandolfi and C. Cordon-Cardo (2007). "The AKT-mTOR pathway plays a critical role in the development of leiomyosarcomas." Nat Med **13**(6): 748-753.

Hidalgo, M., E. Bruckheimer, N. V. Rajeshkumar, I. Garrido-Laguna, E. De Oliveira, B. Rubio-Viqueira, S. Strawn, M. J. Wick, J. Martell and D. Sidransky (2011). "A pilot clinical study of treatment guided by personalized tumorgrafts in patients with advanced cancer." Mol Cancer Ther **10**(8): 1311-1316.

Hillmer, A. M., F. Yao, K. Inaki, W. H. Lee, P. N. Ariyaratne, A. S. Teo, X. Y. Woo, Z. Zhang, H. Zhao, L. Ukil, J. P. Chen, F. Zhu, J. B. So, M. Salto-Tellez, W. T. Poh, K. F. Zawack, N. Nagarajan, S. Gao, G. Li, V. Kumar, H. P. Lim, Y. Y. Sia, C. S. Chan, S. T. Leong, S. C. Neo, P. S. Choi, H. Thoreau, P. B. Tan, A. Shahab, X. Ruan, J. Bergh, P. Hall, V. Cacheux-Rataboul, C. L. Wei, K. G. Yeoh, W. K. Sung, G. Bourque, E. T. Liu and Y. Ruan (2011). "Comprehensive long-span paired-end-tag mapping reveals characteristic patterns of structural variations in epithelial cancer genomes." Genome Res **21**(5): 665-675.

Hirota, S., K. Isozaki, Y. Moriyama, K. Hashimoto, T. Nishida, S. Ishiguro, K. Kawano, M. Hanada, A. Kurata, M. Takeda, G. Muhammad Tunio, Y. Matsuzawa, Y. Kanakura, Y. Shinomura and Y. Kitamura (1998). "Gain-of-Function Mutations of c-kit in Human Gastrointestinal Stromal Tumors." Science **279**(5350): 577-580.

Hirsch, D., J. Camps, S. Varma, R. Kemmerling, M. Stapleton, T. Ried and T. Gaiser (2012). "A new whole genome amplification method for studying clonal evolution patterns in malignant colorectal polyps." Genes Chromosomes Cancer **51**(5): 490-500.

Hirsch, D., R. Kemmerling, S. Davis, J. Camps, P. S. Meltzer, T. Ried and T. Gaiser (2013). "Chromothripsis and focal copy number alterations determine poor outcome in malignant melanoma." Cancer Res **73**(5): 1454-1460.

Hoppin, J. A., P. E. Tolbert, R. F. Herrick, D. S. Freedman, B. D. Ragsdale, K. R. Horvat and E. A. Brann (1998). "Occupational chlorophenol exposure and soft tissue sarcoma risk among men aged 30-60 years." Am J Epidemiol **148**(7): 693-703.

Hostetter, G., S. Y. Kim, S. Savage, G. C. Gooden, M. Barrett, J. Zhang, L. Alla, A. Watanabe, J. Einspahr, A. Prasad, B. J. Nickoloff, J. Carpten, J. Trent, D. Alberts and M. Bittner (2010). "Random DNA fragmentation allows detection of single-copy, single-exon alterations of copy number by oligonucleotide array CGH in clinical FFPE samples." Nucleic Acids Res **38**(2): e9.

Hsu, S. M., L. Raine and H. Fanger (1981). "Use of avidin-biotin-peroxidase complex (ABC) in immunoperoxidase techniques: a comparison between ABC and unlabeled antibody (PAP) procedures." J Histochem Cytochem **29**(4): 577-580.

Hsu, T. C. and C. M. Pomerat (1953). "MAMMALIAN CHROMOSOMES IN VITRO: II. A Method for Spreading the Chromosomes of Cells in Tissue Culture." Journal of Heredity **44**(1): 23-30.

Hu, J., U. N. Rao, S. Jasani, V. Khanna, K. Yaw and U. Surti (2005). "Loss of DNA copy number of 10q is associated with aggressive behavior of leiomyosarcomas: a comparative genomic hybridization study." Cancer Genet Cytogenet **161**(1): 20-27.

Huang, C. Y., C. Y. Lee, M. Y. Chen, W. H. Yang, Y. H. Chen, C. H. Chang, H. C. Hsu, Y. C. Fong and C. H. Tang (2009). "Stromal cell-derived factor-1/CXCR4 enhanced motility of human osteosarcoma cells involves MEK1/2, ERK and NF-kappaB-dependent pathways." J Cell Physiol **221**(1): 204-212.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009). "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists." Nucleic Acids Res **37**(1): 1-13.

Huang da, W., B. T. Sherman and R. A. Lempicki (2009). "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources." Nat Protoc **4**(1): 44-57.

Huret, J. L., M. Ahmad, M. Arsaban, A. Bernheim, J. Cigna, F. Desangles, J. C. Guignard, M. C. Jacquemot-Perbal, M. Labarussias, V. Leberre, A. Malo, C. Morel-Pair, H. Mossafa, J. C. Potier, G. Texier, F. Viguie, S. Yau Chun Wan-Senon, A. Zasadzinski and P. Dessen (2013). "Atlas of genetics and cytogenetics in oncology and haematology in 2013." Nucleic Acids Res **41**(Database issue): D920-924.

Iafrate, A. J., L. Feuk, M. N. Rivera, M. L. Listewnik, P. K. Donahoe, Y. Qi, S. W. Scherer and C. Lee (2004). "Detection of large-scale variation in the human genome." Nat Genet **36**(9): 949-951.

Idbaih, A., J.-M. Coindre, J. Derre, O. Mariani, P. Terrier, D. Ranchere, A. Mairal and A. Aurias (2004). "Myxoid malignant fibrous histiocytoma and pleomorphic liposarcoma share very similar genomic imbalances." Lab Invest **85**(2): 176-181.

Italiano, A., L. Bianchini, E. Gjernes, F. d. r. Keslair, D. Ranchere-Vince, J.-M. Dumollard, J. Haudebourg, A. s. Leroux, C. MainguenÃ©, P. Terrier, F. d. r. Chibon, J.-M. Coindre and F. Pedeutour (2009). "Clinical and Biological Significance of CDK4 Amplification in Well-Differentiated and Dedifferentiated Liposarcomas." Clinical Cancer Research **15**(18): 5696-5703.

Italiano, A., C.-L. Chen, R. Thomas, M. Breen, F. Bonnet, N. Sevenet, M. Longy, R. G. Maki, J.-M. Coindre and C. R. Antonescu (2012). "Alterations of the p53 and PIK3CA/AKT/mTOR pathways in angiosarcomas." Cancer **118**(23): 5878-5887.

Iwasaki, H., K. Nabeshima, J. Nishio, S. Jimi, M. Aoki, K. Koga, M. Hamasaki, H. Hayashi and A. Mogi (2009). "Pathology of soft-tissue tumors: daily diagnosis, molecular cytogenetics and experimental approach." Pathol Int **59**(8): 501-521.

Jaffe, A. B. and A. Hall (2005). "Rho GTPases: biochemistry and biology." Annu Rev Cell Dev Biol **21**: 247-269.

Jain, S., R. Xu, V. G. Prieto and P. Lee (2010). "Molecular classification of soft tissue sarcomas and its clinical applications." Int J Clin Exp Pathol **3**(4): 416-428.

Jefford, C. E. and I. Irminger-Finger (2006). "Mechanisms of chromosome instability in cancers." Crit Rev Oncol Hematol **59**(1): 1-14.

Johnson, J. E., E. J. Gettings, J. Schwalm, J. Pei, J. R. Testa, S. Litwin, M. von Mehren and D. Broccoli (2007). "Whole-genome profiling in liposarcomas reveals genetic alterations common to specific telomere maintenance mechanisms." Cancer Res **67**(19): 9221-9228.

Jonigk, D., F. Laenger, L. Maegel, N. Izykowski, J. Rische, C. Tiede, C. Klein, B. Maecker-Kolhoff, H. Kreipe and K. Hussein (2012). "Molecular and clinicopathological analysis of Epstein-Barr virus-associated posttransplant smooth muscle tumors." Am J Transplant **12**(7): 1908-1917.

Kahl, P., L. Gullotti, L. C. Heukamp, S. Wolf, N. Friedrichs, R. Vorreuther, G. Solleder, P. J. Bastian, J. Ellinger, E. Metzger, R. Schule and R. Buettner (2006). "Androgen receptor coactivators lysine-specific histone demethylase 1 and four and a half LIM domain protein 2 predict risk of prostate cancer recurrence." Cancer Res **66**(23): 11341-11347.

Kaick, G. v., A. Dalheimer, S. Hornik, A. Kaul, D. Liebermann, H. LÃ¼hrs, A. Spiethoff, K. Wegener and H. Wesch (1999). "The German Thorotrast Study: Recent Results and Assessment of Risks." Radiation Research **152**(6): S64-S71.

Kamai, T., T. Yamanishi, H. Shirataki, K. Takagi, H. Asami, Y. Ito and K.-I. Yoshida (2004). "Overexpression of RhoA, Rac1, and Cdc42 GTPases Is Associated with Progression in Testicular Cancer." Clinical Cancer Research **10**(14): 4799-4805.

Kamb, A. (2010). "At a crossroads in oncology." Curr Opin Pharmacol **10**(4): 356-361.

Kamiyama, H., S. Rauenzahn, J. S. Shim, C. A. Karikari, G. Feldmann, L. Hua, M. Kamiyama, F. W. Schuler, M. T. Lin, R. M. Beaty, B. Karanam, H. Liang, M. E. Mullendore, G. Mo, M. Hidalgo, E. Jaffee, R. H. Hruban, H. A. Jinnah, R. B. Roden, A. Jimeno, J. O. Liu, A. Maitra and J. R. Eshleman (2013). "Personalized chemotherapy profiling using cancer cell lines from selectable mice." Clin Cancer Res **19**(5): 1139-1146.

Kato, S., N. Espinoza, S. Lange, M. Villalon, M. Cuello and G. I. Owen (2008). "Characterization and phenotypic variation with passage number of cultured human endometrial adenocarcinoma cells." Tissue Cell **40**(2): 95-102.

Keely, P. J., J. K. Westwick, I. P. Whitehead, C. J. Der and L. V. Parise (1997). "Cdc42 and Rac1 induce integrin-mediated cell motility and invasiveness through PI(3)K." Nature **390**(6660): 632-636.

Keir, S. T., J. M. Maris, R. Lock, E. A. Kolb, R. Gorlick, H. Carol, C. L. Morton, C. P. Reynolds, M. H. Kang, A. Watkins, P. J. Houghton and M. A. Smith (2010). "Initial testing (stage 1) of the multi-targeted kinase inhibitor sorafenib by the pediatric preclinical testing program." Pediatr Blood Cancer **55**(6): 1126-1133.

Kim, T. M., R. Xi, L. J. Luquette, R. W. Park, M. D. Johnson and P. J. Park (2013). "Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes." Genome Res **23**(2): 217-227.

King, A. A., M. R. Debaun, V. M. Riccardi and D. H. Gutmann (2000). "Malignant peripheral nerve sheath tumors in neurofibromatosis 1." Am J Med Genet **93**(5): 388-392.

Kiuru-Kuhlefelt, S., W. El-Rifai, J. Fanburg-Smith, J. Kere, M. Miettinen and S. Knuutila (2001). "Concomitant DNA copy number amplification at 17q and 22q in dermatofibrosarcoma protuberans." Cytogenetic and Genome Research **92**(3-4): 192-195.

Kleiber, K., K. Strebhardt and B. T. Martin (2007). "The biological relevance of FHL2 in tumour cells and its role as a putative cancer target." Anticancer Res **27**(1A): 55-61.

Kleinerman, R. A., S. J. Schonfeld and M. A. Tucker (2012). "Sarcomas in hereditary retinoblastoma." Clin Sarcoma Res **2**(1): 15.

Kloosterman, W. P., M. Hoogstraat, O. Paling, M. Tavakoli-Yaraki, I. Renkens, J. S. Vermaat, M. J. van Roosmalen, S. van Lieshout, I. J. Nijman, W. Roessingh, R. van 't Slot, J. van de Belt, V. Guryev, M. Koudijs, E. Voest and E. Cuppen (2011). "Chromothripsis is a common mechanism driving genomic rearrangements in primary and metastatic colorectal cancer." Genome Biol **12**(10): R103.

Kloth, J. N., J. Oosting, T. van Wezel, K. Szuhai, J. Knijnenburg, A. Gorter, G. G. Kenter, G. J. Fleuren and E. S. Jordanova (2007). "Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer." Bmc Genomics **8**.

Knezevich, S. R., D. E. McFadden, W. Tao, J. F. Lim and P. H. Sorensen (1998). "A novel ETV6-NTRK3 gene fusion in congenital fibrosarcoma." Nat Genet **18**(2): 184-187.

Knowles, M. and P. Selby (2006). Introduction to the cellular and molecular biology of cancer. Oxford, Oxford university press**:** 1-24.

Knudson, A. G. (1971). "Mutation and Cancer - Statistical Study of Retinoblastoma " Proceedings of the National Academy of Sciences of the United States of America **68**(4): 820-&.

Koch, A., A. Waha, W. Hartmann, A. Hrychyk, U. Schüller, A. Waha, K. A. Wharton, S. Y. Fuchs, D. v. Schweinitz and T. Pietsch (2005). "Elevated Expression of Wnt Antagonists Is a Common Event in Hepatoblastomas." Clinical Cancer Research **11**(12): 4295-4304.

Korbel, J. O. and P. J. Campbell (2013). "Criteria for inference of chromothripsis in cancer genomes." Cell **152**(6): 1226-1236.

Korf, B. R. (2000). "Malignancy in neurofibromatosis type 1." Oncologist **5**(6): 477-485.

Kosla, J., D. Pankova, J. Plachy, O. Tolde, K. Bicanova, M. Dvorak, D. Rosel and J. Brabek (2013). "Metastasis of aggressive amoeboid sarcoma cells is dependent on Rho/ROCK/MLC signaling." Cell Commun Signal **11**: 51.

Kozma, R., S. Ahmed, A. Best and L. Lim (1995). "The Ras-related protein Cdc42Hs and bradykinin promote formation of peripheral actin microspikes and filopodia in Swiss 3T3 fibroblasts." Mol Cell Biol **15**(4): 1942-1952.

Krishnan, B., G. Khanna and D. Clohisy (2008). "Gene Translocations in Musculoskeletal Neoplasms." Clinical Orthopaedics and Related Research® **466**(9): 2131-2146.

Krontiris, T. G. and G. M. Cooper (1981). "Transforming activity of human tumor DNAs." Proc Natl Acad Sci U S A **78**(2): 1181-1184.

Kurpinski, K., H. Lam, J. Chu, A. Wang, A. Kim, E. Tsay, S. Agrawal, D. V. Schaffer and S. Li (2010). "Transforming growth factor-beta and notch signaling mediate stem cell differentiation into smooth muscle cells." Stem Cells **28**(4): 734-742.

Kwaepila, N., G. Burns and A. S. Leong (2006). "Immunohistological localisation of human FAT1 (hFAT) protein in 326 breast cancers. Does this adhesion molecule have a role in pathogenesis?" Pathology **38**(2): 125-131.

Kwiatkowska, A., S. Didier, S. Fortin, Y. Y. Chuang, T. White, M. E. Berens, E. Rushing, J. Eschbacher, N. L. Tran, A. Chan and M. Symons (2012). "The small GTPase RhoG mediates glioblastoma cell invasion." Molecular Cancer **11**.

Lahortiga, I., K. De Keersmaecker, P. Van Vlierberghe, C. Graux, B. Cauwelier, F. Lambert, N. Mentens, H. B. Beverloo, R. Pieters, F. Speleman, M. D. Odero, M. Bauters, G. Froyen, P. Marynen, P. Vandenberghe, I. Wlodarska, J. P. P. Meijerink and J. Cools (2007). "Duplication of the MYB oncogene in T cell acute lymphoblastic leukemia." Nat Genet **39**(5): 593-595.

Lallemand-Breitenbach, V. and H. de Thé (2010). "PML Nuclear Bodies." Cold Spring Harbor Perspectives in Biology **2**(5).

Lambert, P. J., A. Z. Shahrier, A. G. Whitman, O. F. Dyson, A. J. Reber, J. A. McCubrey and S. M. Akula (2007). "Targeting the PI3K and MAPK pathways to treat Kaposi's-sarcoma-associated herpes virus infection and pathogenesis." Expert Opin Ther Targets **11**(5): 589-599.

Lane, J., T. A. Martin, R. E. Mansel and W. G. Jiang (2008). "The expression and prognostic value of the guanine nucleotide exchange factors (GEFs) Trio, Vav1 and TIAM-1 in human breast cancer." Int Semin Surg Oncol **5**: 23.

Larramendy, M. L., M. Gentile, S. Soloneski, S. Knuutila and T. Böhling (2008). "Does comparative genomic hybridization reveal distinct differences in DNA copy number sequence patterns between leiomyosarcoma and malignant fibrous histiocytoma?" Cancer Genetics and Cytogenetics **187**(1): 1-11.

Larramendy, M. L., S. Kaur, C. Svarvar, T. Bohling and S. Knuutila (2006). "Gene copy number profiling of soft-tissue leiomyosarcomas by array-comparative genomic hybridization." Cancer Genet Cytogenet **169**(2): 94-101.

Launay, E., C. Pangault, P. Bertrand, F. Jardin, T. Lamy, H. Tilly, K. Tarte, C. Bastard and T. Fest (2012). "High rate of TNFRSF14 gene alterations related to 1p36 region in de novo follicular lymphoma and impact on prognosis." Leukemia **26**(3): 559-562.

Lawrence, B., A. Perez-Atayde, M. K. Hibbard, B. P. Rubin, P. Dal Cin, J. L. Pinkus, G. S. Pinkus, S. Xiao, E. S. Yi, C. D. M. Fletcher and J. A. Fletcher (2000). TPM3-ALK and TPM4-ALK Oncogenes in Inflammatory Myofibroblastic Tumors. **157:** 377-384.

Lazer, G. and S. Katzav (2011). "Guanine nucleotide exchange factors for RhoGTPases: good therapeutic targets for cancer therapy?" Cell Signal **23**(6): 969-979.

Lengauer, C., K. W. Kinzler and B. Vogelstein (1998). "Genetic instabilities in human cancers." Nature **396**(6712): 643-649.

Li, F. P. and J. F. Fraumeni, Jr. (1982). Prospective Study of a Family Cancer Syndrome. **247:** 2692-2694.

Li, M., J. Wang, S. S. Ng, C. Y. Chan, A. C. Chen, H. P. Xia, D. T. Yew, B. C. Wong, Z. Chen, H. F. Kung and M. C. Lin (2008). "The four-and-a-half-LIM protein 2 (FHL2) is overexpressed in gliomas and associated with oncogenic activities." Glia **56**(12): 1328-1338.

Lieber, M. R. (2010). "The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway." Annu Rev Biochem **79**: 181-211.

Liegl-Atzwanger, B., J. Fletcher and C. Fletcher (2010). "Gastrointestinal stromal tumors." Virchows Archiv **456**(2): 111-127.

Lin, P. P., Y. Wang and G. Lozano (2011). "Mesenchymal Stem Cells and the Origin of Ewing's Sarcoma." Sarcoma **2011**.

Liu, Y., E. Sanchez-Tillo, X. Lu, B. Clem, S. Telang, A. B. Jenson, M. Cuatrecasas, J. Chesney, A. Postigo and D. C. Dean (2013). "Rb1 family mutation is sufficient for sarcoma initiation." Nat Commun **4**: 2650.

Lord, C. J. and A. Ashworth (2012). "The DNA damage response and cancer therapy." Nature **481**(7381): 287-294.

Lu, J., L. Chan, H. D. Fiji, R. Dahl, O. Kwon and F. Tamanoi (2009). "In vivo antitumor effect of a novel inhibitor of protein geranylgeranyltransferase-I." Mol Cancer Ther **8**(5): 1218-1226.

Lu, X., X. Tang, W. Guo, T. Ren and H. Zhao (2010). "Sorafenib induces growth inhibition and apoptosis of human chondrosarcoma cells by blocking the RAF/ERK/MEK pathway." J Surg Oncol **102**(7): 821-826.

Luca, T., G. Privitera, M. Lo Monaco, C. Prezzavento and S. Castorina (2007). "Validation study of a cell culture model of colorectal cancer." Ital J Anat Embryol **112**(2): 81-92.

Macieira-Coelho, A. (1998). "Markers of 'cell senescence'." Mechanisms of Ageing and Development **103**(1): 105-109.

Makk, L., J. L. Creech, J. G. Whelan, Jr. and M. N. Johnson (1974). "Liver Damage and Angiosarcoma in Vinyl Chloride Workers: A Systematic Detection Program." JAMA: The Journal of the American Medical Association **230**(1): 64-68.

Malkin, D., F. P. Li, L. C. Strong, C. E. Nelson, D. H. Kim and et al. (1990). "Germ Line p53 Mutations in a Familial Syndrome of Breast Cancer, Sarcomas, and Other Neoplasms." Science **250**(4985): 1233.

Mandahl, N., F. Mertens, I. Panagopoulos and S. Knuutila (2004). "Genetic characterization of bone and soft tissue tumors." Acta Orthop Scand Suppl **75**(311): 21-28.

Mankoo, P. K., R. Shen, N. Schultz, D. A. Levine and C. Sander (2011). "Time to recurrence and survival in serous ovarian tumors predicted from integrated genomic profiles." PLoS One **6**(11): e24709.

Marees, T., A. C. Moll, S. M. Imhof, M. R. de Boer, P. J. Ringens and F. E. van Leeuwen (2008). "Risk of Second Malignancies in Survivors of Retinoblastoma: More Than 40 Years of Follow-up." Journal of the National Cancer Institute **100**(24): 1771-1779.

Mariani, O., C. Brennetot, J.-M. Coindre, N. Gruel, C. Ganem, O. Delattre, M.-H. Stern and A. Aurias (2007). "JUN Oncogene Amplification and Overexpression Block Adipocytic Differentiation in Highly Aggressive Sarcomas." Cancer Cell **11**(4): 361-374.

Martin, B. T., K. Kleiber, V. Wixler, M. Raab, B. Zimmer, M. Kaufmann and K. Strebhardt (2007). "FHL2 regulates cell cycle-dependent and doxorubicin-induced p21Cip1/Waf1 expression in breast cancer cells." Cell Cycle **6**(14): 1779-1788.

Martin, S., W. Grear Hurt, L. Hedges, M. Butler and H. Schwartz (1998). "Microsatellite instability in sarcomas." Annals of Surgical Oncology **5**(4): 356-360.

Mc Sherry, E. A., A. Mc Goldrick, E. W. Kay, A. M. Hopkins, W. M. Gallagher and P. A. Dervan (2007). "Formalin-fixed paraffin-embedded clinical tissues show spurious copy number changes in array-CGH profiles." Clin Genet **72**(5): 441-447.

McArthur, G. A., G. D. Demetri, A. van Oosterom, M. C. Heinrich, M. Debiec-Rychter, C. L. Corless, Z. Nikolova, S. Dimitrijevic and J. A. Fletcher (2005). "Molecular and Clinical Analysis of Locally Advanced Dermatofibrosarcoma Protuberans Treated With Imatinib: Imatinib Target Exploration Consortium Study B2225." **23**(4): 866-873.

McBain, J. A., J. L. Weese, L. F. Meisner, W. H. Wolberg and J. K. Willson (1984). "Establishment and characterization of human colorectal cancer cell lines." Cancer Res **44**(12 Pt 1): 5813-5821.

McClain, K. L., C. T. Leach, H. B. Jenson, V. V. Joshi, B. H. Pollock, R. T. Parmley, F. J. DiCarlo, E. G. Chadwick and S. B. Murphy (1995). Association of Epstein-Barr Virus with Leiomyosarcomas in Young People with AIDS. **332:** 12-18.

Mertens, F., I. Panagopoulos and N. Mandahl (2009). "Genomic characteristics of soft tissue sarcomas." Virchows Arch **456**(2): 129-139.

Miki, T., C. L. Smith, J. E. Long, A. Eva and T. P. Fleming (1993). "Oncogene Ect2 Is Related to Regulators of Small Gtp-Binding Proteins." Nature **362**(6419): 462-465.

Minoletti, F., M. Miozzo, F. Pedeutour, L. Sard, S. Pilotti, A. Azzarelli, C. Turc-Carel, M. A. Pierotti and G. Sozzi (1995). "Involvement of chromosomes 17 and 22 in dermatofibrosarcoma protuberans." Genes Chromosomes Cancer **13**(1): 62-65.

Mitelman, F., B. Johansson and F. Mertens (2007). "The impact of translocations and gene fusions on cancer causation." Nat Rev Cancer **7**(4): 233-245.

Mitelman, F., B. Johansson, F. Mertens and P. Cancer Genome Anatomy (2003). Mitelman database of chromosome aberrations in cancer, Cancer Genome Anatomy Project.

Mitra, A., L. Mishra and S. Li (2013). "Technologies for deriving primary tumor cells for use in personalized cancer therapy." Trends Biotechnol.

Miyagawa, Y., H. Okita, H. Nakaijima, Y. Horiuchi, B. Sato, T. Taguchi, M. Toyoda, Y. U. Katagiri, J. Fujimoto, J. Hata, A. Umezawa and N. Kiyokawa (2008). "Inducible expression of chimeric EWS/ETS proteins confers Ewing's family tumor-like phenotypes to human mesenchymal progenitor cells." Mol Cell Biol **28**(7): 2125-2137.

Mizuarai, S., K. Yamanaka and H. Kotani (2006). "Mutant p53 Induces the GEF-H1 Oncogene, a Guanine Nucleotide Exchange Factor-H1 for RhoA, Resulting in Accelerated Cell Proliferation in Tumor Cells." Cancer Research **66**(12): 6319-6326.

Mondal, G., M. Rowley, L. Guidugli, J. Wu, Vernon S. Pankratz and Fergus J. Couch (2012). "BRCA2 Localization to the Midbody by Filamin A Regulates CEP55 Signaling and Completion of Cytokinesis." Developmental Cell **23**(1): 137-152.

Montgomery, E., P. Argani, J. L. Hicks, A. M. DeMarzo and A. K. Meeker (2004). "Telomere lengths of translocation-associated and nontranslocation-associated sarcomas differ dramatically." Am J Pathol **164**(5): 1523-1529.

Monument, M. J., S. L. Lessnick, J. D. Schiffman and R. T. Randall (2012). "Microsatellite instability in sarcoma: fact or fiction?" ISRN Oncol **2012**: 473146.

Müerköster, S., A. Arlt, B. Sipos, M. Witt, M. Großmann, G. Klöppel, H. Kalthoff, U. R. Fölsch and H. Schäfer (2005). "Increased Expression of the E3-Ubiquitin Ligase Receptor Subunit βTRCP1 Relates to Constitutive Nuclear Factor-κB Activation and Chemoresistance in Pancreatic Carcinoma Cells." Cancer Res **65**(4): 1316-1324.

Mullighan, C. G., S. Goorha, I. Radtke, C. B. Miller, E. Coustan-Smith, J. D. Dalton, K. Girtman, S. Mathew, J. Ma, S. B. Pounds, X. Su, C.-H. Pui, M. V. Relling, W. E. Evans, S. A. Shurtleff and J. R. Downing (2007). "Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia." Nature **446**(7137): 758-764.

Murnane, J. P. (2012). "Telomere dysfunction and chromosome instability." Mutat Res **730**(1-2): 28-36.

Myllykangas, S., T. Böhling and S. Knuutila (2007). "Specificity, selection and significance of gene amplifications in cancer." Seminars in Cancer Biology **17**(1): 42-55.

Na, K. Y., Y. W. Kim and Y. K. Park (2012). "Mitogen-activated protein kinase pathway in osteosarcoma." Pathology **44**(6): 540-546.

Naeem, R., M. L. Lux, S. F. Huang, S. P. Naber, J. M. Corson and J. A. Fletcher (1995). "Ring chromosomes in dermatofibrosarcoma protuberans are composed of interspersed sequences from chromosomes 17 and 22." Am J Pathol **147**(6): 1553-1558.

Nakachi, S., T. Nakazato, C. Ishikawa, R. Kimura, D. A. Mann, M. Senba, H. Masuzaki and N. Mori (2011). "Human T-cell leukemia virus type 1 Tax transactivates the matrix metalloproteinase 7 gene via JunD/AP-1 signaling." Biochimica et Biophysica Acta (BBA) - Molecular Cell Research **1813**(5): 731-741.

Nayak, S. K., S. Kakati, S. R. Harvey, C. C. Malone, A. N. Cornforth and R. O. Dillman (2000). "Characterization of cancer cell lines established from two human metastatic breast cancers." In Vitro Cell Dev Biol Anim **36**(3): 188-193.

Negrini, S., V. G. Gorgoulis and T. D. Halazonetis (2010). "Genomic instability--an evolving hallmark of cancer." Nat Rev Mol Cell Biol **11**(3): 220-228.

Neill, N. J., B. S. Torchia, B. A. Bejjani, L. G. Shaffer and B. C. Ballif (2010). "Comparative analysis of copy number detection by whole-genome BAC and oligonucleotide array CGH." Mol Cytogenet **3**: 11.

Neuzillet, C., A. Tijeras-Raballand, L. de Mestier, J. Cros, S. Faivre and E. Raymond (2013). "MEK in cancer and cancer therapy." Pharmacol Ther.

Ng, G., D. Winder, B. Muralichar, E. Gooding, I. Roberts, M. Pett, G. Mukherjee, J. Huang and N. Coleman (2007). "Gain and overexpression of the oncostatin M receptor occur frequently in cervical squamous cell carcinoma and are associated with adverse clinical outcome." Journal of Pathology **212**(3): 325-334.

Nilsson, P., L. Paavilainen, K. Larsson, J. Odling, M. Sundberg, A. C. Andersson, C. Kampf, A. Persson, C. Al-Khalili Szigyarto, J. Ottosson, E. Bjorling, S. Hober, H. Wernerus, K. Wester, F. Ponten and M. Uhlen (2005). "Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling." Proteomics **5**(17): 4327-4337.

Novais, E. N., B. Demiralp, J. Alderete, M. C. Larson, P. S. Rose and F. H. Sim (2010). "Do surgical margin and local recurrence influence survival in soft tissue sarcomas?" Clin Orthop Relat Res **468**(11): 3003-3011.

Oikawa, M., K. Yoshiura, H. Kondo, S. Miura, T. Nagayasu and M. Nakashima (2011). "Significance of genomic instability in breast cancer in atomic bomb survivors: analysis of microarray-comparative genomic hybridization." Radiat Oncol **6**: 168.

Ordonez, J. L., D. Osuna, D. J. Garcia-Dominguez, A. T. Amaral, A. P. Otero-Motta, C. Mackintosh, M. V. Sevillano, M. V. Barbado, T. Hernandez and E. de Alava (2010). "The clinical relevance of molecular genetics in soft tissue sarcomas." Adv Anat Pathol **17**(3): 162-181.

Ougolkov, A., B. Zhang, K. Yamashita, V. Bilim, M. Mai, S. Y. Fuchs and T. Minamoto (2004). "Associations Among β-TrCP, an E3 Ubiquitin Ligase Receptor, β-Catenin, and NF-κB in Colorectal Cancer." J Natl Cancer Inst **96**(15): 1161-1170.

Pan, C., C. Kumar, S. Bohl, U. Klingmueller and M. Mann (2009). "Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions." Mol Cell Proteomics **8**(3): 443-450.

Pan, Y., F. Bi, N. Liu, Y. Xue, X. Yao, Y. Zheng and D. Fan (2004). "Expression of seven main Rho family members in gastric carcinoma." Biochem Biophys Res Commun **315**(3): 686-691.

Papachristou, D. J., G. I. Papachristou, O. A. Papaefthimiou, N. J. Agnantis, E. K. Basdra and A. G. Papavassiliou (2005). "The MAPK-AP-1/-Runx2 signalling axes are implicated in chondrosarcoma pathobiology either independently or via up-regulation of VEGF." Histopathology **47**(6): 565-574.

Patel, A. S., K. M. Murphy, A. L. Hawkins, J. S. Cohen, P. P. Long, E. J. Perlman and C. A. Griffin (2007). "RANBP2 and CLTC are involved in ALK rearrangements in inflammatory myofibroblastic tumors." Cancer Genetics and Cytogenetics **176**(2): 107-114.

Pedeutour, F., J. P. Lacour, C. Perrin, K. Huffermann, M. P. Simon, N. Ayraud and C. Turc-Carel (1996). "Another case of t(17;22)(q22;q13) in an infantile dermatofibrosarcoma protuberans." Cancer Genetics and Cytogenetics **89**(2): 175-176.

Pedeutour, F., R. F. Suijkerbuijk, A. Forus, J. Vangaal, W. Vandeklundert, J. M. Coindre, G. Nicolo, F. Collin, U. Vanhaelst, K. Huffermann and C. Turccarel (1994). "Complex Composition and Coamplification of Sas and Mdm2 in Ring and Giant Rod Marker Chromosomes in Well-Differentiated Liposarcoma." Genes Chromosomes & Cancer **10**(2): 85-94.

Peng, C. L., W. Guo, T. Ji, T. Ren, Y. Yang, D. S. Li, H. Y. Qu, X. Li, S. Tang, T. Q. Yan and X. D. Tang (2009). "Sorafenib induces growth inhibition and apoptosis in human synovial sarcoma cells via inhibiting the RAF/MEK/ERK signaling pathway." Cancer Biol Ther **8**(18): 1729-1736.

Perez-Losada, J., B. Pintado, A. Gutierrez-Adan, T. Flores, B. Banares-Gonzalez, J. C. del Campo, J. F. Martin-Martin, E. Battaner and I. Sanchez-Garcia (2000). "The chimeric FUS/TLS-CHOP fusion protein specifically induces liposarcomas in transgenic mice." Oncogene **19**(20): 2413-2422.

Perot, G., J. Derre, J. M. Coindre, F. Tirode, C. Lucchesi, O. Mariani, L. Gibault, L. Guillou, P. Terrier and A. Aurias (2009). "Strong smooth muscle differentiation is dependent on myocardin gene amplification in most human retroperitoneal leiomyosarcomas." Cancer Res **69**(6): 2269-2278.

Pignochino, Y., G. Grignani, G. Cavalloni, M. Motta, M. Tapparo, S. Bruno, A. Bottos, L. Gammaitoni, G. Migliardi, G. Camussi, M. Alberghini, B. Torchio, S. Ferrari, F. Bussolino, F. Fagioli, P. Picci and M. Aglietta (2009). "Sorafenib blocks tumour growth, angiogenesis and metastatic potential in preclinical models of osteosarcoma through a mechanism potentially involving the inhibition of ERK1/2, MCL-1 and ezrin pathways." Mol Cancer **8**: 118.

Puechberty, J., A. M. Laurent, S. Gimenez, A. Billault, M. E. Brun-Laurent, A. Calenda, B. Marcais, C. Prades, P. Ioannou, Y. Yurov and G. Roizes (1999). "Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: recombination across 5cen." Genomics **56**(3): 274-287.

Quick, Q. and O. Skalli (2010). "α-Actinin 1 and α-actinin 4: Contrasting roles in the survival, motility, and RhoA signaling of astrocytoma cells." Experimental Cell Research **316**(7): 1137-1147.

Rajagopalan, H. and C. Lengauer (2004). "Aneuploidy and cancer." Nature **432**(7015): 338-341.

Raman, M., S. Earnest, K. Zhang, Y. Zhao and M. H. Cobb (2007). "TAO kinases mediate activation of p38 in response to DNA damage." EMBO J **26**(8): 2005-2014.

Raphel, L., A. Talasila, C. Cheung and S. Sinha (2012). "Myocardin overexpression is sufficient for promoting the development of a mature smooth muscle cell-like phenotype from human embryonic stem cells." PLoS One **7**(8): e44052.

Rausch, T., D. T. Jones, M. Zapatka, A. M. Stutz, T. Zichner, J. Weischenfeldt, N. Jager, M. Remke, D. Shih, P. A. Northcott, E. Pfaff, J. Tica, Q. Wang, L. Massimi, H. Witt, S. Bender, S. Pleier, H. Cin, C. Hawkins, C. Beck, A. von Deimling, V. Hans, B. Brors, R. Eils, W. Scheurlen, J. Blake, V. Benes, A. E. Kulozik, O. Witt, D. Martin, C. Zhang, R. Porat, D. M. Merino, J. Wasserman, N. Jabado, A. Fontebasso, L. Bullinger, F. G. Rucker, K. Dohner, H. Dohner, J. Koster, J. J. Molenaar, R. Versteeg, M. Kool, U. Tabori, D. Malkin, A. Korshunov, M. D. Taylor, P. Lichter, S. M. Pfister and J. O. Korbel (2012). "Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations." Cell **148**(1-2): 59-71.

Renshaw, J., K. R. Taylor, R. Bishop, M. Valenti, A. De Haven Brandon, S. Gowan, S. A. Eccles, R. R. Ruddle, L. D. Johnson, F. I. Raynaud, J. L. Selfe, K. Thway, T. Pietsch, A. D. Pearson and J. Shipley (2013). "Dual blockade of the PI3K/AKT/mTOR (AZD8055) and RAS/MEK/ERK (AZD6244) pathways synergistically inhibits rhabdomyosarcoma cell growth in vitro and in vivo." Clin Cancer Res **19**(21): 5940-5951.

Rhodes, D. R., J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey and A. M. Chinnaiyan (2004). "ONCOMINE: a cancer microarray database and integrated data-mining platform." Neoplasia **6**(1): 1-6.

Ridley, A. J. and A. Hall (1992). "The small GTP-binding protein rho regulates the assembly of focal adhesions and actin stress fibers in response to growth factors." Cell **70**(3): 389-399.

Ridley, A. J., H. F. Paterson, C. L. Johnston, D. Diekmann and A. Hall (1992). "The small GTP-binding protein rac regulates growth factor-induced membrane ruffling." Cell **70**(3): 401-410.

Rigby, P. W., M. Dieckmann, C. Rhodes and P. Berg (1977). "Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I." J Mol Biol **113**(1): 237-251.

Robinson, E., A. I. Neugut and P. Wylie (1988). "Clinical Aspects of Postirradiation Sarcomas1." JNCI Journal of the National Cancer Institute **80**(4): 233-240.

Rocchi, A., M. C. Manara, M. Sciandra, D. Zambelli, F. Nardi, G. Nicoletti, C. Garofalo, S. Meschini, A. Astolfi, M. P. Colombo, S. L. Lessnick, P. Picci and K. Scotlandi (2010). "CD99 inhibits neural

differentiation of human Ewing sarcoma cells and thereby contributes to oncogenesis." J Clin Invest **120**(3): 668-680.

Rossi, E., C. Klersy, R. Manca, O. Zuffardi and E. Solcia (2011). "Correlation between genomic alterations assessed by array comparative genomic hybridization, prognostically informative histologic subtype, stage, and patient survival in gastric cancer." Hum Pathol **42**(12): 1937-1945.

Routhier, A., M. Astuccio, D. Lahey, N. Monfredo, A. Johnson, W. Callahan, A. Partington, K. Fellows, L. Ouellette, S. Zhidro, C. Goodrow, A. Smith, K. Sullivan, P. Simone, L. Le, B. Vezuli, M. Zohni, E. West, D. Gleason and B. Bryan (2010). "Pharmacological inhibition of Rho-kinase signaling with Y-27632 blocks melanoma tumor growth." Oncology Reports **23**(3): 861-867.

Rueda, O. M. and R. Diaz-Uriarte (2010). "Finding recurrent copy number alteration regions: A review of methods." Curr. Bioinform. Current Bioinformatics **5**(1): 1-17.

Sahai, E. and C. J. Marshall (2002). "RHO-GTPases and cancer." Nat Rev Cancer **2**(2): 133-142.

Salawu, A., A. Ul-Hassan, D. Hammond, M. Fernando, M. Reed and K. Sisley (2012). "High Quality Genomic Copy Number Data from Archival Formalin-Fixed Paraffin-Embedded Leiomyosarcoma: Optimisation of Universal Linkage System Labelling." Plos One **7**(11).

Salhia, B., N. L. Tran, A. Chan, A. Wolf, M. Nakada, F. Rutka, M. Ennis, W. S. McDonough, M. E. Berens, M. Symons and J. T. Rutka (2008). "The Guanine Nucleotide Exchange Factors Trio, Ect2, and Vav3 Mediate the Invasive Behavior of Glioblastoma." American Journal of Pathology **173**(6): 1828-1838.

Salhia, B., N. L. Tran, A. Chan, A. Wolf, M. Nakada, F. Rutka, M. Ennis, W. S. McDonough, M. E. Berens, M. Symons and J. T. Rutka (2008). "The guanine nucleotide exchange factors trio, Ect2, and Vav3 mediate the invasive behavior of glioblastoma." Am J Pathol **173**(6): 1828-1838.

Sanz-Moreno, V., G. Gadea, J. Ahn, H. Paterson, P. Marra, S. Pinner, E. Sahai and C. J. Marshall (2008). "Rac activation and inactivation control plasticity of tumor cell movement." Cell **135**(3): 510-523.

Sarlomo-Rikala, M., A. J. Kovatich, A. Barusevicius and M. Miettinen (1998). "CD117: a sensitive marker for gastrointestinal stromal tumors that is more specific than CD34." Mod Pathol **11**(8): 728-734.

Sasaki, K., T. Hitora, O. Nakamura, R. Kono and T. Yamamoto (2011). "The role of MAPK pathway in bone and soft tissue tumors." Anticancer Res **31**(2): 549-553.

Schmidt, S., S. Diriong, J. Mery, E. Fabbrizio and A. Debant (2002). "Identification of the first Rho-GEF inhibitor, TRIPalpha, which targets the RhoA-specific GEF domain of Trio." FEBS Lett **523**(1-3): 35-42.

Schreiber, H., F. M. Barry, W. C. Russell, W. L. I. V. Macon, J. L. Ponsky and W. J. Pories (1979). "Stewart-Treves Syndrome: A Lethal Complication of Postmastectomy Lymphedema and Regional Immune Deficiency." Arch Surg **114**(1): 82-85.

Seale, P., B. Bjork, W. Yang, S. Kajimura, S. Chin, S. Kuang, A. Scime, S. Devarakonda, H. M. Conroe, H. Erdjument-Bromage, P. Tempst, M. A. Rudnicki, D. R. Beier and B. M. Spiegelman (2008). "PRDM16 controls a brown fat/skeletal muscle switch." Nature **454**(7207): 961-967.

Seipel, K., Q. G. Medley, N. L. Kedersha, X. A. Zhang, S. P. O'Brien, C. Serra-Pages, M. E. Hemler and M. Streuli (1999). "Trio amino-terminal guanine nucleotide exchange factor domain expression promotes actin cytoskeleton reorganization, cell migration and anchorage-independent cell growth." Journal of Cell Science **112**(12): 1825-1834.

Shaulian, E. (2010). "AP-1--The Jun proteins: Oncogenes or tumor suppressors in disguise?" Cell Signal **22**(6): 894-899.

Sheppard, D. G. and H. I. Libshitz (2001). "Post-radiation Sarcomas: A Review of the Clinical and Imaging Features in 63 Cases." Clinical Radiology **56**(1): 22-29.

Shimada, H., M. Kuboshima, T. Shiratori, Y. Nabeya, A. Takeuchi, H. Takagi, F. Nomura, M. Takiguchi, T. Ochiai and T. Hiwasa (2007). "Serum anti-myomegalin antibodies in patients with esophageal squamous cell carcinoma." Int J Oncol **30**(1): 97-103.

Shimizu, T., A. W. Tolcher, K. P. Papadopoulos, M. Beeram, D. W. Rasco, L. S. Smith, S. Gunn, L. Smetzer, T. A. Mays, B. Kaiser, M. J. Wick, C. Alvarez, A. Cavazos, G. L. Mangold and A. Patnaik (2012). "The clinical effect of the dual-targeting strategy involving PI3K/AKT/mTOR and RAS/MEK/ERK pathways in patients with advanced cancer." Clin Cancer Res **18**(8): 2316-2325.

Shoemaker, R. H. (2006). "The NCI60 human tumour cell line anticancer drug screen." Nat Rev Cancer **6**(10): 813-823.

Silvany, R. E., S. Eliazer, N. C. Wolff and R. L. Ilaria, Jr. (2000). "Interference with the constitutive activation of ERK1 and ERK2 impairs EWS/FLI-1-dependent transformation." Oncogene **19**(39): 4523-4530.

Simon, M. P., F. Pedeutour, N. Sirvent, J. Grosgeorge, F. Minoletti, J. M. Coindre, M. J. Terrier-Lacombe, N. Mandahl, R. D. Craver, N. Blin, G. Sozzi, C. Turc-Carel, K. P. O'Brien, D. Kedra, I. Fransson, C. Guilbaud and J. P. Dumanski (1997). "Deregulation of the platelet-derived growth factor B-chain gene via fusion with collagen gene COL1A1 in dermatofibrosarcoma protuberans and giant-cell fibroblastoma." Nat Genet **15**(1): 95-98.

Skytting, B., G. Nilsson, B. Brodin, Y. Xie, J. Lundeberg, M. Uhlen and O. Larsson (1999). "A Novel Fusion Gene, SYT-SSX4, in Synovial Sarcoma." J. Natl. Cancer Inst. **91**(11): 974-975.

Slupianek, A., S. K. Jozwiakowski, E. Gurdek and T. Skorski (2009). "BCR/ABL kinase interacts with and phosphorylates the RAD51 paralog, RAD51B." Leukemia **23**(12): 2308-2310.

Smith, J., B. J. Stewart, S. Glaysher, K. Peregrin, L. A. Knight, D. J. Weber and I. A. Cree (2010). "The effect of pentamidine on melanoma ex vivo." Anticancer Drugs **21**(2): 181-185.

Sordillo, P. P., R. Chapman, S. I. Hajdu, G. B. Magill and R. B. Golbey (1981). "Lymphangiosarcoma." Cancer **48**(7): 1674-1679.

Srinivasan, M., D. Sedmak and S. Jewell (2002). "Effect of fixatives and tissue processing on the content and integrity of nucleic acids." Am J Pathol **161**(6): 1961-1971.

Steeg, P. S. (2006). "Tumor metastasis: mechanistic insights and clinical challenges." Nat Med **12**(8): 895-904.

Stephens, P. J., C. D. Greenman, B. Fu, F. Yang, G. R. Bignell, L. J. Mudie, E. D. Pleasance, K. W. Lau, D. Beare, L. A. Stebbings, S. McLaren, M. L. Lin, D. J. McBride, I. Varela, S. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, M. A. Quail, J. Burton, H. Swerdlow, N. P. Carter, L. A. Morsberger, C. Iacobuzio-Donahue, G. A. Follows, A. R. Green, A. M. Flanagan, M. R. Stratton, P. A. Futreal and P. J. Campbell (2011). "Massive genomic rearrangement acquired in a single catastrophic event during cancer development." Cell **144**(1): 27-40.

Stewart, F. W. and N. Treves (1948). "Lymphangiosarcoma in postmastectomy lymphedema. A report of six cases in elephantiasis chirurgica." Cancer **1**(1): 64-81.

Stratton, M. R., P. J. Campbell and P. A. Futreal (2009). "The cancer genome." Nature **458**(7239): 719-724.

Szponar, A., M. V. Yusenko and G. Kovacs (2010). "High-resolution array CGH of metanephric adenomas: lack of DNA copy number changes." Histopathology **56**(2): 212-216.

Tamama, K., C. K. Sen and A. Wells (2008). "Differentiation of bone marrow mesenchymal stem cells into the smooth muscle lineage by blocking ERK/MAPK signaling pathway." Stem Cells Dev **17**(5): 897-908.

Tan, D. S., M. B. Lambros, R. Natrajan and J. S. Reis-Filho (2007). "Getting it right: designing microarray (and not 'microawry') comparative genomic hybridization studies for cancer research." Lab Invest **87**(8): 737-754.

Taniguchi, C. M., J. Winnay, T. Kondo, R. T. Bronson, A. R. Guimaraes, J. O. Aleman, J. Luo, G. Stephanopoulos, R. Weissleder, L. C. Cantley and C. R. Kahn (2010). "The phosphoinositide 3-kinase regulatory subunit p85alpha can exert tumor suppressor properties through negative regulation of growth factor signaling." Cancer Res **70**(13): 5305-5315.

Taylor, B. S., J. Barretina, R. G. Maki, C. R. Antonescu, S. Singer and M. Ladanyi (2011). "Advances in sarcoma genomics and new therapeutic targets." Nat Rev Cancer **11**(8): 541-557.

Taylor, B. S., J. Barretina, N. D. Socci, P. DeCarolis, M. Ladanyi, M. Meyerson, S. Singer and C. Sander (2008). "Functional Copy-Number Alterations in Cancer." PLoS ONE **3**(9): e3179.

The Cancer Genome Atlas Research Network (2008). "Comprehensive genomic characterization defines human glioblastoma genes and core pathways." Nature **455**(7216): 1061-1068.

The Cancer Genome Atlas Research Network (2011). "Integrated genomic analyses of ovarian carcinoma." Nature **474**(7353): 609-615.

Thomas, M., A. Gessner, H.-P. Vornlocher, P. Hadwiger, J. Greil and O. Heidenreich (2005). "Targeting MLL-AF4 with short interfering RNAs inhibits clonogenicity and engraftment of t(4;11)-positive human leukemic cells." Blood **106**(10): 3559-3566.

Thomas, M., J. Greil and O. Heidenreich (2006). "Targeting leukemic fusion proteins with small interfering RNAs: recent advances and therapeutic potentials." Acta Pharmacol Sin **27**(3): 273-281.

Thway, K. (2009). "Pathology of soft tissue sarcomas." Clin Oncol (R Coll Radiol) **21**(9): 695-705.

Toguchida, J. and T. Nakayama (2009). "Molecular genetics of sarcomas: applications to diagnoses and therapy." Cancer Sci **100**(9): 1573-1580.

Toro, J. R., L. B. Travis, H. J. Wu, K. Zhu, C. D. M. Fletcher and S. S. Devesa (2006). "Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978-2001: An analysis of 26,758 cases." International Journal of Cancer **119**(12): 2922-2930.

Trojani, M., G. Contesso, J. M. Coindre, J. Rouesse, N. B. Bui, A. de Mascarel, J. F. Goussot, M. David, F. Bonichon and C. Lagarde (1984). "Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system." Int J Cancer **33**(1): 37-42.

Ul-Hassan, A., K. Sisley, D. Hughes, D. W. Hammond, M. H. Robinson and M. W. Reed (2009). "Common genetic changes in leiomyosarcoma and gastrointestinal stromal tumour: implication for ataxia telangiectasia mutated involvement." Int J Exp Pathol **90**(5): 549-557.

Ul-Hassan, A., K. Sisley, D. Hughes, D. W. Hammond, M. H. Robinson and M. W. R. Reed (2009). "Common genetic changes in leiomyosarcoma and gastrointestinal stromal tumour: implication for ataxia telangiectasia mutated involvement." International Journal of Experimental Pathology **90**(5): 549-557.

Valsesia, A., A. Mace, S. Jacquemont, J. S. Beckmann and Z. Kutalik (2013). "The Growing Importance of CNVs: New Insights for Detection and Clinical Interpretation." Front Genet **4**: 92.

van Beers, E. H., S. A. Joosse, M. J. Ligtenberg, R. Fles, F. B. Hogervorst, S. Verhoef and P. M. Nederlof (2006). "A multiplex PCR predictor for aCGH success of FFPE samples." Br J Cancer **94**(2): 333-337.

van Rijssel, J., M. Hoogenboezem, L. Wester, P. L. Hordijk and J. D. Van Buul (2012). "The N-Terminal DH-PH Domain of Trio Induces Cell Spreading and Migration by Regulating Lamellipodia Dynamics in a Rac1-Dependent Fashion." PLoS ONE **7**(1): e29912.

van Tuyn, J., S. Knaän-Shanzer, M. J. M. van de Watering, M. de Graaf, A. van der Laarse, M. J. Schalij, E. E. van der Wall, A. A. F. de Vries and D. E. Atsma (2005). "Activation of cardiac and smooth muscle-specific genes in primary human cells after forced expression of human myocardin." Cardiovascular Research **67**(2): 245-255.

Wagner, E. F. and A. R. Nebreda (2009). "Signal integration by JNK and p38 MAPK pathways in cancer development." Nat Rev Cancer **9**(8): 537-549.

Wang, J., Y. Yang, H. H. Xia, Q. Gu, M. C. Lin, B. Jiang, Y. Peng, G. Li, X. An, Y. Zhang, Z. Zhuang, Z. Zhang, H. F. Kung and B. C. Wong (2007). "Suppression of FHL2 expression induces cell differentiation and inhibits gastric and colon carcinogenesis." Gastroenterology **132**(3): 1066-1076.

Wang, Z., D.-Z. Wang, G. C. T. Pipes and E. N. Olson (2003). "Myocardin is a master regulator of smooth muscle gene expression." Proceedings of the National Academy of Sciences **100**(12): 7129-7134.

Warren, S. and G. N. J. J. Sommer (1936). "FIBROSARCOMA OF THE SOFT PARTS: WITH SPECIAL REFERENCE TO RECURRENCE AND METASTASIS." Archives of Surgery **33**(3): 425-450.

Weir, B., X. Zhao and M. Meyerson (2004). "Somatic alterations in the human cancer genome." Cancer Cell **6**(5): 433-438.

Weir, B. A., M. S. Woo, G. Getz, S. Perner, L. Ding, R. Beroukhim, W. M. Lin, M. A. Province, A. Kraja, L. A. Johnson, K. Shah, M. Sato, R. K. Thomas, J. A. Barletta, I. B. Borecki, S. Broderick, A. C. Chang, D. Y. Chiang, L. R. Chirieac, J. Cho, Y. Fujii, A. F. Gazdar, T. Giordano, H. Greulich, M. Hanna, B. E. Johnson, M. G. Kris, A. Lash, L. Lin, N. Lindeman, E. R. Mardis, J. D. McPherson, J. D. Minna, M. B. Morgan, M. Nadel, M. B. Orringer, J. R. Osborne, B. Ozenberger, A. H. Ramos, J. Robinson, J. A. Roth, V. Rusch, H. Sasaki, F. Shepherd, C. Sougnez, M. R. Spitz, M.-S. Tsao, D. Twomey, R. G. W. Verhaak, G. M. Weinstock, D. A. Wheeler, W. Winckler, A. Yoshizawa, S. Yu, M. F. Zakowski, Q. Zhang, D. G. Beer, I. I. Wistuba, M. A. Watson, L. A. Garraway, M. Ladanyi, W. D. Travis, W. Pao, M. A. Rubin, S. B. Gabriel, R. A. Gibbs, H. E. Varmus, R. K. Wilson, E. S. Lander and M. Meyerson (2007). "Characterizing the cancer genome in lung adenocarcinoma." Nature **450**(7171): 893-898.

Weiss, R. A., D. Whitby, S. Talbot, P. Kellam and C. Boshoff (1998). "Human Herpesvirus Type 8 and Kaposi's Sarcoma." J Natl Cancer Inst Monogr **1998**(23): 51-54.

Widemann, B. C., A. Kim, E. Fox, S. Baruchel, P. C. Adamson, A. M. Ingle, J. Glade Bender, M. Burke, B. Weigel, D. Stempak, F. M. Balis and S. M. Blaney (2012). "A phase I trial and pharmacokinetic study of sorafenib in children with refractory solid tumors or leukemias: a Children's Oncology Group Phase I Consortium report." Clin Cancer Res **18**(21): 6011-6022.

Wiedemeyer, R., C. Brennan, T. P. Heffernan, Y. Xiao, J. Mahoney, A. Protopopov, H. Zheng, G. Bignell, F. Furnari, W. K. Cavenee, W. C. Hahn, K. Ichimura, V. P. Collins, G. C. Chu, M. R. Stratton, K. L. Ligon, P. A. Futreal and L. Chin (2008). "Feedback Circuit among INK4 Tumor Suppressors Constrains Human Glioblastoma Development." Cancer Cell **13**(4): 355-364.

Wilkinson, K., E. R. P. Velloso, L. F. Lopes, C. Lee, J. C. Aster, M. A. Shipp and R. C. T. Aguiar (2003). "Cloning of the t(1;5)(q23;q33) in a myeloproliferative disorder associated with eosinophilia: involvement of PDGFRB and response to imatinib." Blood **102**(12): 4187-4190.

Willems, L., J. Tamburini, N. Chapuis, C. Lacombe, P. Mayeux and D. Bouscary (2012). "PI3K and mTOR signaling pathways in cancer: new data on targeted therapies." Curr Oncol Rep **14**(2): 129-138.

Winokur, S. T., Y. W. Chen, P. S. Masny, J. H. Martin, J. T. Ehmsen, S. J. Tapscott, S. M. van der Maarel, Y. Hayashi and K. M. Flanigan (2003). "Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation." Hum Mol Genet **12**(22): 2895-2907.

Wong, F. L., J. D. Boice, Jr., D. H. Abramson, R. E. Tarone, R. A. Kleinerman, M. Stovall, M. B. Goldman, J. M. Seddon, N. Tarbell, J. F. Fraumeni, Jr. and F. P. Li (1997). Cancer Incidence After Retinoblastoma: Radiation Dose and Sarcoma Risk. **278:** 1262-1267.

Xia, S. J., J. G. Pressey and F. G. Barr (2002). "Molecular pathogenesis of rhabdomyosarcoma." Cancer Biol Ther **1**(2): 97-104.

Xiao, W., A. B. Mohseny, P. C. Hogendoorn and A. M. Cleton-Jansen (2013). "Mesenchymal stem cell transformation and sarcoma genesis." Clin Sarcoma Res **3**(1): 10.

Xu, L., S. Li and B. A. Stohr (2013). "The role of telomere biology in cancer." Annu Rev Pathol **8**: 49-78.

Yamamoto, T., T. Ohno, K. Wakahara, A. Nagano, G. Kawai, M. Saitou, I. Takigami, A. Matsuhashi, K. Yamada and K. Shimizu (2009). "Simultaneous inhibition of mitogen-activated protein kinase and phosphatidylinositol 3-kinase pathways augment the sensitivity to actinomycin D in Ewing sarcoma." J Cancer Res Clin Oncol **135**(8): 1125-1136.

Yamanaka, R., R. Blumenthal, M. V. Lorenzi, T. Tatsumoto and T. Miki (2001). "Ostip2, a novel oncoprotein that associates with the Rho exchange factor Ost." DNA and Cell Biology **20**(7): 383-390.

Yamori, T. (2003). "Panel of human cancer cell lines provides valuable database for drug discovery and bioinformatics." Cancer Chemother Pharmacol **52 Suppl 1**: S74-79.

Yang, S.-H., A. D. Sharrocks and A. J. Whitmarsh (2013). "MAP kinase signalling cascades and transcriptional regulation." Gene **513**(1): 1-13.

Yau, C., D. Mouradov, R. N. Jorissen, S. Colella, G. Mirza, G. Steers, A. Harris, J. Ragoussis, O. Sieber and C. C. Holmes (2010). "A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data." Genome Biology **11**(9).

Ylipaa, A., K. K. Hunt, J. Yang, A. J. Lazar, K. E. Torres, D. C. Lev, M. Nykter, R. E. Pollock, J. Trent and W. Zhang (2011). "Integrative genomic characterization and a genomic staging system for gastrointestinal stromal tumors." Cancer **117**(2): 380-389.

Yoshizuka, N., R. Moriuchi, T. Mori, K. Yamada, S. Hasegawa, T. Maeda, T. Shimada, Y. Yamada, S. Kamihira, M. Tomonaga and S. Katamine (2004). "An alternative transcript derived from the trio locus encodes a guanosine nucleotide exchange factor with mouse cell-transforming potential." Journal of Biological Chemistry **279**(42): 43998-44004.

Zack, T. I., S. E. Schumacher, S. L. Carter, A. D. Cherniack, G. Saksena, B. Tabak, M. S. Lawrence, C. Z. Zhang, J. Wala, C. H. Mermel, C. Sougnez, S. B. Gabriel, B. Hernandez, H. Shen, P. W. Laird, G. Getz, M. Meyerson and R. Beroukhim (2013). "Pan-cancer patterns of somatic copy number alteration." Nat Genet **45**(10): 1134-1140.

Zeilstra, J., S. P. Joosten, F. M. Wensveen, M. C. Dessing, D. M. Schutze, E. Eldering, M. Spaargaren and S. T. Pals (2011). "WNT signaling controls expression of pro-apoptotic BOK and BAX in intestinal cancer." Biochem Biophys Res Commun **406**(1): 1-6.

Zender, L., M. S. Spector, W. Xue, P. Flemming, C. Cordon-Cardo, J. Silke, S.-T. Fan, J. M. Luk, M. Wigler, G. J. Hannon, D. Mu, R. Lucito, S. Powers and S. W. Lowe (2006). "Identification and

Validation of Oncogenes in Liver Cancer Using an Integrative Oncogenomic Approach." Cell **125**(7): 1253-1267.

Zerbini, L. F., J. F. de Vasconcellos, A. Czibere, Y. Wang, J. D. Paccez, X. Gu, J.-R. Zhou and T. A. Libermann (2011). "JunD-mediated repression of GADD45α and γ regulates escape from cell death in prostate cancer." Cell Cycle **10**(15): 2583-2591.

Zhang, P., K. S. Bhakta, P. L. Puri, R. O. Newbury, J. R. Feramisco and J. Y. Wang (2003). "Association of ataxia telangiectasia mutated (ATM) gene mutation/deletion with rhabdomyosarcoma." Cancer Biol Ther **2**(1): 87-91.

Zheng, M., R. Simon, M. Mirlacher, T. Forster, P. A. Diener, M. J. Mihatsch, G. Sauter and P. Schraml (2004). "TRIO amplification and abundant mRNA expression is associated with invasive tumor growth and rapid tumor cell proliferation in urinary bladder cancer." American Journal of Pathology **165**(1): 63-69.

Zhong, S., P. Hu, T. Z. Ye, R. Stan, N. A. Ellis and P. P. Pandolfi (1999). "A role for PML and the nuclear body in genomic stability." Oncogene **18**(56): 7941-7947.

APPENDICES

APPENDIX 1

Details of All STS Cases used in this Study

STS Subtype	Subtype No.	Gender	Age (years)	Site	Size	Grade	Neoadjuvant Therapy	Sample Type	aCGH	Tissue Culture	Culture Designation	Culture outcome
Alveolar Soft Part Sarcoma (ASPS)	01	Female	30	Lower Limb	92	-	-	FRESH	Yes	Yes	STS 01/10	p5
	02	Male	18	Trunk	38	-	-	FRESH	Yes	Yes	STS 15/10	p9
Angiosarcoma (ANGIO)	01	Female	68	Trunk	70	3	-	FRESH	Yes	Yes	STS 05/10	p5
	02	Male	48	Trunk	120	3	-	FRESH	Yes	Yes	STS 10/11	p5
	03	Male	69	Lower Limb	60	3	-	FRESH	Yes	Yes	STS 21/10	p9
	04	Female	41	Trunk	35	3	-	FRESH	No	Yes	STS 03/12	Early Senescence
	05	Female	76	Trunk	27	3	-	FRESH	No	Yes	STS 06/12	Early Senescence
	06	Female	59	Trunk	40	3	-	FRESH	No	Yes	STS 08/12	Early Senescence
De-differentiated Liposarcoma (DDLPS)	01	Female	68	Retroperitoneum	300	3	-	FRESH	Yes	Yes	STS 09/10	Long Term
	02	Male	65	Retroperitoneum	80	3	-	FRESH	Yes	Yes	STS 17/11	Early Senescence
	03	Female	69	Retroperitoneum	170	3	-	FRESH	Yes	Yes	STS 19/11	Nil
	04	Female	64	Retroperitoneum	210	3	-	FRESH	Yes	Yes	STS18/10	Nil
	05	Female	77	Lower Limb	340	3	-	FRESH	No	Yes	STS 13/10	Nil
	06	Female	70	Lower Limb	170	3	Radiotherapy	FRESH	Yes	Yes	STS 20/11	Long Term
	07	Female	57	Retroperitoneum	200	3	-	FRESH	No	Yes	STS 01/12	Early Senescence
Ewings' Sarcoma (EWING)	01	Male	23	Pelvis	40	-	Chemotherapy	FRESH	Yes	Yes	STS 13/11	p9
Extraskeletal Myxoid Chondrosarcoma (EMCS)	01	Male	30	Lower Limb	103	2	-	FRESH	Yes	Yes	STS 02/10	p5
Gastrointestinal Stromal Tumour (GIST)	01	Male	75	Stomach	62	High	-	FFPE	Yes	No	-	-
	02	Male	70	Small bowel	100	Intermediate	-	FFPE	Yes	No	-	-
	03	Male	71	Small bowel	35	High	-	FFPE	Yes	No	-	-
	04	Male	59	Stomach	39	Intermediate	-	FFPE	Yes	No	-	-
	05	Male	76	Stomach	70	Low	-	FFPE	Yes	No	-	-
	06	Male	61	Stomach	27	Low	-	FFPE	Yes	No	-	-
	07	Male	27	Small Bowel	95	Intermediate	-	FFPE	Yes	No	-	-
	08	Male	58	Liver Met	67	High	-	FFPE	Yes	No	-	-
	09	Male	74	Stomach	50	Intermediate	-	FFPE	Yes	No	-	-
	10	Female	73	Stomach	54	Low	-	FFPE	Yes	No	-	-
	11	Male	79	Stomach	30	-	-	FFPE	Yes	No	-	-

APPENDIX 1

Details of All STS Cases used in this Study

STS Subtype	Subtype No.	Gender	Age (years)	Site	Size	Grade	Neoadjuvant Therapy	Sample Type	aCGH	Tissue Culture	Culture Designation	Culture outcome
Leiomyosarcoma (LMS)	01	Female	63	Intraabdominal	210	3	-	FFPE	Yes	No	-	-
	02	Female	38	Lower Limb	60	3	-	FFPE	Yes	No	-	-
	03	Female	47	Lower Limb	50	3	-	FFPE	Yes	No	-	-
	04	Male	67	Pelvis	85	3	-	FFPE	Yes	No	-	-
	05	Female	63	Intraabdominal	50	2	-	FFPE	Yes	No	-	-
	06	Female	82	Lower Limb	60	3	-	FFPE	Yes	No	-	-
	07	Male	37	Pelvis	50	3	-	FFPE	Yes	No	-	-
	08	Female	63	Pelvis	50	3	-	BOTH	Yes	Yes	STS 02/11	Long Term
	09	Male	39	Intraabdominal	95	3	-	BOTH	Yes	Yes	STS 11/11	Nil
	10	Female	49	Pelvis	85	1	-	BOTH	Yes	Yes	STS 18/11	p9
	11	Female	80	Intraabdominal	60	3	-	FFPE	Yes	No	-	-
	12	Female	58	Pelvis	85	3	-	FFPE	Yes	No	-	-
	13	Female	76	Intraabdominal	40	3	-	FFPE	Yes	No	-	-
	14	Male	77	Pelvis	130	3	-	FFPE	Yes	No	-	-
	15	Male	54	Pelvis	160	3	-	FFPE	Yes	No	-	-
	16	Female	69	Head and Neck	50	2	-	FFPE	Yes	No	-	-
	17	Female	54	Pelvis	15	3	-	FFPE	Yes	No	-	-
	18	Female	49	Intraabdominal	180	3	-	FFPE	Yes	No	-	-
	19	Female	52	Intraabdominal	79	3	-	FFPE	Yes	No	-	-
	20	Female	69	Intraabdominal	82	3	-	FFPE	Yes	No	-	-
	21	Female	49	Lower Limb	35	1	-	FFPE	Yes	No	-	-
Low Grade Myofibroblastic Sarcoma (LGMFS)	01	Female	37	Lower Limb	80	1	-	FRESH	No	Yes	STS 07/12	P5
Malignant Peripheral Nerve sheath Tumour (MPNST)	01	Female	46	Lower Limb	88	3	-	FRESH	Yes	Yes	STS 01/11	p5
Myxofibrosarcoma (MFS)	01	Male	65	Lower Limb	105	3	-	FRESH	Yes	Yes	STS 08/11	Early Senescence
	02	Male	43	Lower Limb	35	3	-	FRESH	Yes	Yes	STS 16/10	p9
	03	Male	73	Upper Limb	50	3	-	FRESH	Yes	Yes	STS 21/11	Long Term
	04	Female	43	Lower Limb	80	1	-	FRESH	No	Yes	STS 22/11	Early Senescence
	05	Male	82	Lower Limb	210	3	-	FRESH	No	Yes	STS 04/12	Early Senescence

APPENDIX 1

Details of All STS Cases used in this Study

STS Subtype	Subtype No.	Gender	Age (years)	Site	Size	Grade	Neoadjuvant Therapy	Sample Type	aCGH	Tissue Culture	Culture Designation	Culture outcome
Myxoid Liposarcoma (MLPS)	01	Female	46	Lower Limb	112	-	-	FRESH	No	Yes	STS 10/10	Nil
Pleomorphic Liposarcoma (PLPS)	01	Male	83	Lower Limb	190	-	Radiotherapy	FRESH	Yes	Yes	STS 07/10	Nil
	02	Male	77	Lower Limb	60	3	-	FRESH	Yes	Yes	STS 03/11	p5
	03	Male	65	Lower Limb	100	3	-	FFPE	Yes	No	-	-
Pleomorphic Rhabdomyosarcoma (PRMS)	01	Female	31	Lower Limb	50	-	Chemotherapy	FRESH	Yes	Yes	STS 12/10	p9
Undifferentiated Pleomorphic Sarcoma (UPS)	01	Male	65	Lower Limb	120	3	-	FFPE	Yes	No	-	-
	02	Female	54	Lower Limb	230	3	-	BOTH	Yes	Yes	STS 14/10	Long Term
	03	Male	44	Trunk	108	3	-	FFPE	Yes	No	-	-
	04	Female	36	Lower Limb	72	2	-	FRESH	Yes	Yes	STS 04/11	p9
	05	Male	76	Lower Limb	175	3	-	FRESH	Yes	Yes	STS 06/11	Long Term
	06	Female	67	Lower Limb	115	3	-	BOTH	Yes	Yes	STS 09/11	Long Term
	07	Female	74	Upper Limb	90	3	-	FRESH	Yes	Yes	STS 15/11	Early Senescence
	08	Female	75	Lower Limb	102	3	-	FFPE	Yes	No	-	-
	09	Female	34	Intraabdominal	270	3	-	FFPE	Yes	No	-	-
	10	Male	82		93	3	-	FFPE	Yes	No	-	-
	11	Male	64	Intraabdominal	100	3	-	FFPE	Yes	No	-	-
	12	Male	65	Lower Limb	113	3	-	FFPE	Yes	No	-	-
	13	Male	82	Trunk	87	3	-	FFPE	Yes	No	-	-
	14	Male	66	Lower Limb	48	3	-	FFPE	Yes	No	-	-
	15	Female	67	Lower Limb	72	2	-	FFPE	Yes	No	-	-
	16	Male	80	Lower Limb	55	3	-	FFPE	Yes	No	-	-
	17	Male	72	Trunk	165	-	Radiotherapy	FRESH	No	Yes	STS 16/11	Nil
	18	Female	73	Lower Limb	230	-	Radiotherapy	FRESH	No	Yes	STS 05/12	Nil
Malignant Solitary Fibrous Tumour (SFT)	01	Female	63	Lower Limb	100	2	-	FRESH	Yes	Yes	STS 17/10	p5

APPENDIX 1

Details of All STS Cases used in this Study

STS Subtype	Subtype No.	Gender	Age (years)	Site	Size	Grade	Neoadjuvant Therapy	Sample Type	aCGH	Tissue Culture	Culture Designation	Culture outcome
Synovial Sarcoma (SYN)	01	Female	19	Trunk	40	3	-	FRESH	Yes	Yes	STS 04/10	p5
Well-differentiated Liposarcoma (WDLPS)	01	Male	82	Trunk	180	1	-	FRESH	Yes	Yes	STS 06/10	Nil
	02	Male	53	Trunk	62	1	-	FRESH	Yes	Yes	STS 11/10	Nil
	03	Male	62	Retroperitoneal	195	1	-	FRESH	Yes	Yes	STS 14/11	Early Senescence
	04	Female	50	Limb	105	1	-	FRESH	Yes	Yes	STS 19/10	Nil
	05	Male	52	Lower Limb	240	1	-	FRESH	No	Yes	STS 20/10	Nil
	06	Male	63	Neck	87	1	-	FRESH	No	Yes	STS 05/11	Nil
	07	Male	81	Lower Limb	115	1	-	FRESH	No	Yes	STS 02/12	Nil

APPENDIX 2

Table A: General Laboratory Reagents

Reagent	Supplier
Agarose	Fischer® Scientific, UK
Bromophenol blue	Sigma-Aldrich®, UK
Deionised water	
DePex	vWR® International, UK
EDTA	Fischer® Scientific, UK
Ethanol	Fischer® Scientific, UK
Ethidium bromide	Fischer® Scientific, UK
Gill's haematoxylin	Fischer® Scientific, UK
Glacial acetic acid	Fischer® Scientific, UK
Hydrochloric acid	Fischer® Scientific, UK
Hydrogen peroxide	
Magnesium sulphate	Fischer® Scientific, UK
Methanol	Fischer® Scientific, UK
Nail varnish	
Rubber solution	
Sodium acetate	Fischer® Scientific, UK
Sodium bicarbonate	Fischer® Scientific, UK
Sodium citrate	Fischer® Scientific, UK
Sodium hydroxide	Fischer® Scientific, UK
Sterile phosphate buffered saline (PBS)	vWR® International, UK
Sucrose	Fischer® Scientific, UK
Tris base	Fischer® Scientific, UK
Tween 20	Sigma-Aldrich®, UK
Xylene	Fischer® Scientific, UK

APPENDIX 2

Table B: Laboratory Consumables

Consumable	Supplier
Sterile plastic syringes	Becton Dickinson®, UK
Sterile needles (21G, 23G)	Becton Dickinson®, UK
Sterile scalpels	Swann Morton®, UK
10, 20, 200 and 1000µl pipette tips	StarLab®, UK
10, 20, 200 and 1000µl filter pipette tips	StarLab® UK
0.22um sterile filters	Millipore®, UK
22x 22/32/50mm coverslips	vWR® International, UK
Plastic disposable pipettes	Scientific Laboratory Supplies®, UK
Plastic universal tubes (25ml)	Starstedt®, UK
Centrifuge tubes (15ml, 25ml, 50ml)	Starstedt® UK
Eppendorf Microfuge tubes (0.2, 0.5, 1.5 and 2ml)	Starstedt®, UK
Latex examination gloves	Schottlander® UK
10ml and 25ml Stripettes	
Tissue Culture Plates (6, 12, 24 and 96- well)	Corning®, UK
Sterile petri dishes	Corning®, UK
Tissue Culture flasks (T25, T75)	Corning®, UK

APPENDIX 3

Table A1: Significant Common Focal Amplifications in all 74 STS cases (STAC)

	Cytoband	Region	Length (Mb)	Freq. %	P-Value	Gene Count	Genes	% of CNV Overlap
Chr 1	p36.33	2,270,056-2,285,830	0.016	28	2.00E-02	2	*LOC100129534, MORN1*	100
	p36.32	2,930,401-3,075,672	0.145	28	2.00E-02	4	*ACTRT2, FLJ42875, MIR4251, PRDM16*	14
	p36.32	3,142,248-3,304,483	0.162	28	2.00E-02	1	*PRDM16*	100
	p32.1	59,247,642-59,722,944	0.475	32	0.00E+00	3	*HSD52, JUN, LOC100131060*	35
	p32.1	59,878,157-59,889,297	0.011	30	2.00E-03	1	*FGGY*	0
	p31.3	66,723,046-66,828,021	0.105	28	2.00E-02	1	*PDE4B*	0
	q21.1	144,991,860-144,995,110	0.003	41	0.00E+00	1	*PDE4DIP*	100
	q21.1	145,112,508-145,166,867	0.054	43	0.00E+00	1	*SEC22B*	100
	q21.2	149,216,636-149,301,125	0.084	34	0.00E+00	1	*LOC388692*	100
	q21.2	149,818,426-149,865,193	0.047	34	0.00E+00	10	*HIST2H2AA3, HIST2H2AA4, HIST2H2AB, HIST2H2AC, HIST2H2BC, HIST2H2BE, HIST2H3A, HIST2H3C, HIST2H4A, HIST2H4B*	100
	q21.3	152,615,218-152,784,882	0.170	30	1.70E-02	12	*C1orf68, KPRP, LCE1B, LCE1C, LCE1D, LCE1E, LCE1F, LCE2A, LCE2B, LCE2C, LCE2D, LCE4A*	100
	q21.3	153,479,293-153,560,344	0.081	31	4.00E-03	5	*S100A2, S100A3, S100A4, S100A5, S100A6*	0
	q21.3	154,413,442-154,457,437	0.044	31	4.00E-03	2	*IL6R, SHE*	0
	q21.3	154,789,066-154,804,436	0.015	35	0.00E+00	1	*KCNN3*	0
	q21.3	154,899,255-154,958,577	0.059	38	0.00E+00	7	*CKS1B, FLAD1, MIR4258, PBXIP1, PMVK, PYGO2, SHC1*	0
	q22	155,177,125-155,182,550	0.005	34	0.00E+00	2	*MTX1, THBS3*	100
	q22	155,995,099-156,045,700	0.051	32	1.00E-03	4	*LAMTOR2, MEX3A, RAB25, UBQLN4*	2
	q23.3	164,765,823-164,846,101	0.080	34	0.00E+00	1	*PBX1*	0
	q23.3	165,117,251-165,195,584	0.078	32	1.00E-03	1	*LMX1A*	54
	q24.2	168,235,779-168,258,696	0.023	30	1.70E-02	1	*TBX19*	100
	q24.2	168,460,188-168,614,561	0.154	32	1.00E-03	2	*XCL1, XCL2*	2
	q24.3	172,098,669-172,150,518	0.052	34	0.00E+00	5	*DNM3, DNM3OS, MIR199A2, MIR214, MIR3120*	0
	q24.3	172,340,818-172,372,191	0.031	35	0.00E+00	1	*DNM3*	0
	q25.1	173,144,090-173,262,290	0.118	31	4.00E-03	2	*LOC100506023, TNFSF4*	0
	q25.1	175,470,159-175,529,973	0.060	30	1.70E-02	1	*TNR*	100
Chr 2	p11.1	91,693,395-91,906,643	0.213	34	0.00E+00	1	*LOC654342*	100
	q32.2	189,837,546-189,953,549	0.116	20	0.00E+00	5	*COL3A1, COL5A2, MIR1245A, MIR1245B, MIR3606*	0
Chr 3	p25.3	10,447,023-10,596,307	0.149	26	1.00E-02	1	*ATP2B2*	0
	p25.3	10,653,365-10,961,465	0.308	26	1.00E-02	2	*LINC00606, SLC6A11*	0
	p21.31	50,387,910-50,392,620	0.005	28	0.00E+00	3	*CYB561D2, NPRL2, TMEM115*	100
	q13.2	111,674,780-111,690,254	0.015	23	2.80E-02	1	*PHLDB2*	0
	q13.31	114,115,098-114,383,728	0.269	23	2.80E-02	1	*ZBTB20*	5
	q21.1 - q21.2	123,632,424-123,883,936	0.252	23	2.80E-02	4	*CCDC14, KALRN, MIR5002, ROPN1*	8
	q21.3	126,728,262-126,749,827	0.022	23	2.80E-02	1	*PLXNA1*	0
	q23	141,076,875-141,137,481	0.061	27	0.00E+00	1	*ZBTB38*	0
	q25.1	151,066,762-151,172,224	0.105	23	2.80E-02	3	*IGSF10, MED12L, P2RY12*	16
	q25.31	156,147,410-156,404,495	0.257	23	2.80E-02	5	*KCNAB1, KCNAB1-AS1, SSR3, TIPARP, TIPARP-AS1*	67
	q26.31	170,940,571-171,081,942	0.141	23	2.80E-02	1	*TNIK*	0
	q29	195,435,600-195,450,340	0.015	24	3.00E-03	1	*MUC20*	100
Chr 4	p16.3	3,292,884-3,360,378	0.067	24	4.80E-02	1	*RGS12*	14
	p16.1	6,360,943-6,472,783	0.112	28	0.00E+00	1	*PPP2R2C*	23
	p16.1	7,199,456-7,232,868	0.033	28	0.00E+00	1	*SORCS2*	100
	p16.1	7,616,190-7,705,764	0.090	31	0.00E+00	1	*SORCS2*	100
	p16.1	7,791,491-7,824,654	0.033	26	7.00E-03	1	*AFAP1*	100
	p16.1	8,250,485-8,362,583	0.112	26	7.00E-03	1	*HTRA3*	100
	p16.1	10,099,082-10,209,246	0.110	27	0.00E+00	1	*WDR1*	100
	p15.32	15,335,720-15,462,229	0.127	26	7.00E-03	1	*C1QTNF7*	15
	p15.32	16,404,273-16,706,812	0.303	24	4.80E-02	2	*LDB2, MIR548AX*	5
Chr 5	p15.33	1,056,507-1,117,314	0.061	41	0.00E+00	2	*MIR4635, SLC12A7*	100
	p15.33	1,198,260-1,242,452	0.044	41	0.00E+00	2	*SLC6A18, SLC6A19*	21
	p15.31	9,428,015-9,548,941	0.121	38	1.30E-02	3	*LOC100505806, SEMA5A, SNORD123*	1
	p15.2	14,207,189-14,318,260	0.111	42	0.00E+00	1	*TRIO*	3
	q31.2	136,536,367-136,685,175	0.149	24	4.40E-02	1	*SPOCK1*	0
	q33.3	158,139,982-158,293,641	0.154	28	0.00E+00	1	*EBF1*	3

APPENDIX 3

Table A1: Significant Common Focal Amplifications in all 74 STS cases (STAC)

	Cytoband	Region	Length (Mb)	Freq. %	P-Value	Gene Count	Genes	% of CNV Overlap
Chr 6	p22.2	26,124,423-26,191,035	0.067	27	0.00E+00	5	*HIST1H1E, HIST1H2AC, HIST1H2BD, HIST1H2BE, HIST1H4D*	64
	q21	112,072,112-112,177,222	0.105	24	0.00E+00	1	*FYN*	3
	q21	112,442,154-112,458,332	0.016	22	1.50E-02	1	*LAMA4*	0
Chr 7	p22.3	1,076,058-1,084,756	0.009	35	8.00E-03	1	*C7orf50*	1
	p22.3	1,277,550-1,314,845	0.037	35	8.00E-03	0		100
	p22.3	2,119,407-2,221,398	0.102	36	1.00E-03	1	*MAD1L1*	56
	p15.2	27,165,134-27,196,532	0.031	41	0.00E+00	6	*HOXA-AS3, HOXA3, HOXA4, HOXA5, HOXA6, HOXA7*	100
	p14.1	38,291,860-38,364,472	0.073	57	0.00E+00	1	*TARP*	100
	q21.11	81,241,202-81,429,130	0.188	28	3.40E-02	1	*HGF*	1
	q31.2	115,843,542-115,869,356	0.026	28	3.40E-02	1	*TES*	64
	q32.3	130,570,363-130,648,215	0.078	28	3.40E-02	2	*FLJ43663, LOC646329*	0
	q32.3	131,792,652-132,166,465	0.374	32	0.00E+00	1	*PLXNA4*	35
	q33	134,373,878-134,558,594	0.185	31	1.00E-03	1	*CALD1*	1
	q34	142,348,015-142,384,712	0.037	30	6.00E-03	1	*MTRNR2L6*	100
	q36.1	150,918,904-150,949,877	0.031	30	6.00E-03	4	*ABCF2, CHPF2, MIR671, SMARCD3*	23
	q36.3	155,267,123-155,298,608	0.031	31	1.00E-03	1	*CNPY1*	100
	q36.3	157,624,156-157,657,275	0.033	31	1.00E-03	2	*LOC100506585, PTPRN2*	100
Chr 8	p23.1	7,722,742-7,943,546	0.221	24	2.00E-03	8	*DEFB103A, DEFB103B, DEFB109P1B, DEFB4A, FAM66E, USP17L3, USP17L8, ZNF705B*	100
	p21.3	22,516,604-22,551,427	0.035	23	1.60E-02	2	*BIN3, EGR3*	100
	p11.22	39,267,045-39,354,153	0.087	30	0.00E+00	1	*ADAM3A*	100
	q24.3	142,250,663-142,369,023	0.118	30	5.00E-03	2	*GPR20, LOC731779*	62
	q24.3	142,440,081-142,461,553	0.021	30	5.00E-03	2	*MROH5, PTP4A3*	100
	q24.3	143,453,630-143,656,045	0.202	30	5.00E-03	2	*BAI1, TSNARE1*	96
	q24.3	144,877,684-144,985,070	0.107	30	5.00E-03	5	*EPPK1, MIR937, NRBP2, PUF60, SCRIB*	59
	q24.3	145,022,535-145,036,930	0.014	32	0.00E+00	1	*PLEC*	100
	q24.3	145,662,057-145,715,477	0.053	30	5.00E-03	4	*CYHR1, FOXH1, KIFC2, TONSL*	100
Chr 9	p11.2	43,686,924-44,259,464	0.573	31	0.00E+00	1	*CNTNAP3B*	100
	q13 - q21.11	68,369,294-69,194,128	0.825	32	0.00E+00	4	*LOC100132352, LOC440896, LOC642236, PGM5P2*	100
	q22.31	93,924,918-93,992,140	0.067	27	5.00E-03	1	*AUH*	0
	q22.32	97,847,752-97,893,329	0.046	27	5.00E-03	5	*C9orf3, FANCC, MIR24-1, MIR27B, MIR3074*	0
	q34.3	139,480,157-139,577,038	0.097	26	2.80E-02	3	*AGPAT2, EGFL7, MIR126*	100
Chr 11	p15.5	2,722,900-2,748,031	0.025	26	0.00E+00	1	*KCNQ1*	0
Chr 12	p13.33	2,280,958-2,506,203	0.225	22	2.50E-02	3	*CACNA1C, CACNA1C-AS4, CACNA1C-IT3*	6
	q13.13	52,770,362-52,844,548	0.074	22	1.00E-03	4	*KRT6B, KRT75, KRT82, KRT84*	100
	q13.13	53,035,044-53,091,207	0.056	22	1.00E-03	3	*KRT1, KRT2, KRT77*	25
	q13.13	54,382,251-54,441,012	0.059	22	1.00E-03	8	*HOXC10, HOXC4, HOXC5, HOXC6, HOXC8, HOXC9, MIR196A2, MIR615*	100
	q24.31	124,951,417-124,978,160	0.027	22	1.00E-03	1	*NCOR2*	100
Chr 13	q34	110,918,305-111,046,505	0.128	23	0.00E+00	2	*COL4A1, COL4A2*	8
Chr 14	q13.1	33,983,246-34,106,522	0.123	27	2.30E-02	1	*NPAS3*	100
	q22.1	51,983,513-52,116,762	0.133	30	1.00E-03	3	*FRMD6, FRMD6-AS1, FRMD6-AS2*	6
	q24.1	68,917,581-69,163,933	0.246	34	0.00E+00	1	*RAD51B*	61
Chr 15	q22.2	63,009,973-63,140,185	0.130	26	3.60E-02	2	*MIR190A, TLN2*	0
	q22.2	63,328,861-63,370,712	0.042	26	3.60E-02	1	*TPM1*	0
	q22.31	63,793,316-63,891,739	0.098	26	3.60E-02	3	*FBXL22, LOC100130855, USP3*	13
	q26.2	96,861,171-96,901,074	0.040	26	3.60E-02	2	*MIR1469, NR2F2*	100
Chr 16	p13.3	704,825-751,634	0.047	30	0.00E+00	7	*FBXL16, JMJD8, RHBDL1, RHOT2, STUB1, WDR24, WDR90*	100
	p13.3	1,107,394-1,230,023	0.123	30	0.00E+00	4	*C1QTNF8, CACNA1H, SSTR5, SSTR5-AS1*	100
	p11.2	33,549,833-33,604,960	0.055	35	0.00E+00	1	*RNU6-76*	100
Chr 17	p11.2	17,725,378-17,784,405	0.059	31	0.00E+00	2	*SREBF1, TOM1L2*	100
	q23.1	57,816,707-57,935,524	0.119	22	3.90E-02	2	*MIR21, VMP1*	2
	q25.3	78,792,051-78,943,986	0.152	23	9.00E-03	1	*RPTOR*	100
	q25.3	78,996,511-79,025,995	0.029	26	0.00E+00	2	*BAIAP2, BAIAP2-AS1*	100
	q25.3	79,043,133-79,108,656	0.066	26	0.00E+00	6	*AATK, BAIAP2, MIR1250, MIR3065, MIR338, MIR657*	100
	q25.3	79,204,713-79,252,341	0.048	26	0.00E+00	3	*C17orf89, ENTHD2, SLC38A10*	100
Chr 18	q21.1	46,332,876-46,349,548	0.017	23	0.00E+00	1	*CTIF*	59

APPENDIX 3

Table A1: Significant Common Focal Amplifications in all 74 STS cases (STAC)

	Cytoband	Region	Length (Mb)	Freq. %	P-Value	Gene Count	Genes	% of CNV Overlap
Chr 19	p13.11	18,429,433-18,539,026	0.110	31	0.00E+00	6	*GDF15, LRRC25, LSM4, MIR3189, PGPEP1, SSBP4*	0
	q12	30,459,500-30,521,228	0.062	23	7.00E-03	1	*URI1*	91
	q12	31,213,939-31,839,232	0.625	24	1.00E-03	2	*DKFZp566F0947, TSHZ3*	2
Chr 20	p12.2	10,498,625-10,544,746	0.046	27	0.00E+00	1	*SLX4IP*	0
	p12.2	10,595,495-10,632,707	0.037	27	0.00E+00	2	*JAG1, SLX4IP*	0
	q13.13	48,909,768-49,005,661	0.096	34	8.00E-03	1	*LOC284751*	13
	q13.13	49,144,167-49,195,548	0.051	32	4.10E-02	1	*PTPN1*	0
	q13.33	61,974,698-61,991,078	0.016	35	1.00E-03	1	*CHRNA4*	100
Chr 21	q22.3	46,468,637-46,671,553	0.203	22	1.30E-02	2	*ADARB1, SSR4P1*	1
	q22.3	46,702,083-46,712,289	0.010	22	1.30E-02	2	*LOC642852, POFUT2*	0
	q22.3	46,734,765-46,866,052	0.131	22	1.30E-02	3	*COL18A1, COL18A1-AS1, COL18A1-AS2*	93
	q22.3	47,384,850-47,435,984	0.051	22	1.30E-02	1	*COL6A1*	16
Chr 22	q11.21	20,128,525-20,149,725	0.021	30	0.00E+00	2	*LOC388849, ZDHHC8*	100
	q11.21	20,303,536-20,309,261	0.006	27	1.00E-03	1	*DGCR6L*	100
	q11.23	24,398,622-24,412,555	0.014	24	4.00E-02	2	*CABIN1, GSTTP2*	100
	q12.3	36,665,486-36,730,556	0.065	26	6.00E-03	1	*MYH9*	0
	q13.31	46,449,007-46,491,493	0.042	34	0.00E+00	4	*C22orf26, LOC150381, MIR3619, MIRLET7BHG*	100
	q13.32	48,910,656-49,017,773	0.107	24	4.00E-02	2	*FAM19A5, LOC284933*	100
	q13.32	49,048,111-49,112,319	0.064	24	4.00E-02	1	*FAM19A5*	14
						293		

APPENDIX 3

Table A2: Significant Common Focal Deletions in all 74 STS cases (STAC)

	Cytoband	Region	Length (Mb)	Freq %	p-value	Gene Count	Genes	% of CNV overlap
Chr 1	p36.33	956,608-1,068,631	0.112	27	0.00E+00	3	*AGRN, C1orf159, RNF223*	100
	p36.33	1,075,323-1,107,787	0.032	30	0.00E+00	4	*LOC254099, MIR200A, MIR200B, MIR429*	100
	p36.32	2,371,593-2,419,712	0.048	24	4.00E-03	1	*PLCH2*	100
	p36.32	2,472,466-2,539,213	0.067	24	4.00E-03	5	*FAM213B, LOC100133445, LOC115110, MMEL1, TNFRSF14*	100
	p36.32	3,923,830-4,136,628	0.213	23	3.30E-02	1	*LOC728716*	2
	p36.22	10,016,474-10,200,060	0.184	28	0.00E+00	3	*NMNAT1, RBP7, UBE4B*	0
	p35.3	28,619,691-28,928,183	0.308	27	0.00E+00	11	*MED18, PHACTR4, RAB42, RCC1, SNHG12, SNHG3, SNORA16A, SNORA44, SNORA61, SNORD99, TRNAU1AP*	21
	p35.3	29,009,502-29,055,824	0.046	27	0.00E+00	1	*GMEB1*	1
	p35.2	31,508,693-31,648,449	0.140	23	3.30E-02	1	*PUM1*	1
	p34.3	35,521,725-35,777,359	0.256	24	4.00E-03	3	*SFPQ, ZMYM1, ZMYM4*	12
	p34.1	45,864,086-46,013,665	0.150	24	4.00E-03	4	*CCDC163P, MMACHC, PRDX1, TESK2*	30
	p13.2	113,478,891-113,605,851	0.127	23	3.30E-02	1	*SLC16A1*	100
	p11.2	121,031,118-121,322,377	0.291	23	3.30E-02	2	*EMBP1, SRGAP2D*	100
	q21.2	148,936,653-149,185,842	0.249	23	0.00E+00	1	*LOC645166*	100
	q21.3	153,978,323-154,095,056	0.117	22	0.00E+00	1	*NUP210L*	4
	q25.2	179,343,150-179,387,834	0.045	20	1.00E-03	1	*AXDND1*	4
	q31.3	196,790,444-196,812,218	0.022	26	0.00E+00	1	*CFHR1*	100
	q42.11	224,317,312-224,426,566	0.109	24	0.00E+00	3	*DEGS1, FBXO28, NVL*	96
	q42.3	235,273,856-235,483,275	0.209	23	0.00E+00	5	*ARID4B, MIR4753, RBM34, SNORA14B, TOMM20*	58
	q42.3	235,506,903-235,601,351	0.094	23	0.00E+00	2	*GGPS1, TBCE*	100
	q42.3	235,710,197-235,744,621	0.034	22	0.00E+00	1	*GNG4*	100
	q43	240,317,906-240,395,286	0.077	20	1.00E-03	1	*FMN2*	100
	q43	240,708,989-240,931,477	0.222	23	0.00E+00	1	*GREM2*	0
	q43	242,156,886-242,185,035	0.028	20	1.00E-03	1	*MAP1LC3C*	100
	q44	248,733,443-248,796,977	0.064	24	0.00E+00	3	*OR2T10, OR2T11, OR2T34*	100
	q44	248,983,071-249,250,621	0.268	26	0.00E+00	5	*MIR3124, PGBD2, SH3BP5L, ZNF672, ZNF692*	78
Chr 2	p23.3	24,085,415-24,215,135	0.130	30	0.00E+00	2	*ATAD2B, UBXN2A*	0
	p23.3	26,076,849-26,285,068	0.208	27	1.00E-03	3	*ASXL2, KIF3C, RAB10*	100
	p22.3	32,189,545-32,388,583	0.199	32	0.00E+00	3	*DPY30, MEMO1, SPAST*	2
	p21	42,648,727-42,869,815	0.221	26	1.30E-02	2	*KCNG3, MTA3*	0
	p21	47,566,725-47,610,197	0.043	27	1.00E-03	2	*EPCAM, MIR559*	0
	p15	61,737,824-61,859,483	0.122	34	0.00E+00	1	*XPO1*	72
	q21.3	136,065,879-136,201,097	0.135	28	1.00E-03	1	*ZRANB3*	100
	q33.1	203,075,213-203,095,429	0.020	31	0.00E+00	1	*SUMO1*	1
	q33.1 - q33.2	203,141,016-203,353,381	0.212	32	0.00E+00	5	*BMPR2, NOP58, SNORD11, SNORD11B, SNORD70*	18
	q33.2	203,626,882-203,666,675	0.040	28	1.00E-03	2	*FAM117B, ICA1L*	3
	q33.2	203,975,036-204,095,433	0.120	32	0.00E+00	1	*NBEAL1*	0
	q33.2	204,263,410-204,271,749	0.008	27	1.00E-02	1	*ABI2*	0
	q36.3	228,558,973-228,645,211	0.086	27	1.00E-02	1	*SLC19A3*	0
	q36.3	230,795,681-230,895,774	0.100	35	0.00E+00	1	*FBXO36*	100
	q37.1	232,365,447-232,533,640	0.168	27	1.00E-02	3	*C2orf57, LINC00471, NMUR1*	0
	q37.1	232,744,204-232,758,888	0.015	27	1.00E-02	1	*MIR1471*	100
	q37.1	232,776,466-232,787,523	0.011	27	1.00E-02	1	*NPPC*	100
	q37.1	233,017,240-233,129,262	0.112	28	1.00E-03	1	*DIS3L2*	22
	q37.1	233,419,593-233,427,159	0.008	28	1.00E-03	1	*EIF4E2*	80
	q37.1	233,451,131-233,561,148	0.110	28	1.00E-03	1	*EFHD1*	0
	q37.1	233,968,460-234,082,483	0.114	31	0.00E+00	1	*INPP5D*	0

APPENDIX 3

Table A2: Significant Common Focal Deletions in all 74 STS cases (STAC)

	Cytoband	Region	Length (Mb)	Freq %	p-value	Gene Count	Genes	% of CNV overlap
Chr 2	q37.1	234,984,549-235,056,307	0.072	28	1.00E-03	1	*SPP2*	9
	q37.1	235,166,511-235,422,424	0.256	28	1.00E-03	1	*ARL4C*	0
	q37.2	235,930,332-236,368,103	0.438	30	0.00E+00	1	*SH3BP4*	0
	q37.2	236,464,405-236,970,235	0.506	30	0.00E+00	1	*AGAP1*	28
	q37.3	241,747,277-241,763,176	0.016	34	0.00E+00	1	*KIF1A*	100
	q37.3	242,503,450-242,825,417	0.322	34	0.00E+00	10	*ATG4B, BOK, CXXC11, D2HGDH, DTYMK, GAL3ST2, ING5, NEU4, PDCD1, THAP4*	100
	q37.3	242,930,739-243,199,373	0.269	41	0.00E+00	1	*LOC728323*	64
Chr 3	p22.1	41,835,995-42,009,669	0.174	20	1.00E-03	1	*ULK4*	31
	p21.31	47,839,057-47,880,973	0.042	20	1.00E-03	1	*DHX30*	16
	p21.31	48,294,309-48,307,250	0.013	20	1.00E-03	1	*ZNF589*	100
	p21.31	48,965,902-48,976,775	0.011	20	1.00E-03	1	*ARIH2*	0
	p21.31	49,299,147-49,543,299	0.244	20	1.00E-03	9	*AMT, C3orf62, DAG1, GPX1, MIR4271, NICN1, RHOA, TCTA, USP4*	3
	p21.2	51,469,972-51,621,766	0.152	28	0.00E+00	2	*RAD54L2, VPRBP*	5
	p14.3	57,453,906-57,530,400	0.076	26	0.00E+00	1	*DNAH12*	9
	q13.13	109,113,547-109,143,938	0.030	20	0.00E+00	1	*FLJ25363*	0
	q29	196,000,424-196,206,380	0.206	20	0.00E+00	6	*PCYT1A, RNF168, TCTEX1D2, TM4SF19, TM4SF19-TCTEX1D2, UBXN7*	100
Chr 4	p14	39,662,567-39,715,730	0.053	27	0.00E+00	1	*UBE2K*	27
	p14	39,963,576-40,082,025	0.118	27	0.00E+00	3	*LOC344967, N4BP2, PDS5A*	75
	p14	40,486,730-40,542,205	0.055	23	0.00E+00	2	*MIR4802, RBM47*	13
	q12	57,671,380-57,779,724	0.108	24	1.30E-02	2	*REST, SPINK2*	39
	q31.3	151,923,034-152,012,891	0.090	27	0.00E+00	1	*LRBA*	100
	q35.1	184,425,375-184,842,296	0.417	24	1.30E-02	5	*ING2, LOC389247, RWDD4, STOX2, TRAPPC11*	13
	q35.2	187,113,646-187,446,183	0.333	24	1.30E-02	4	*CYP4V2, F11, KLKB1, LOC285441*	57
	q35.2	188,944,897-189,042,012	0.097	31	0.00E+00	1	*TRIML2*	100
	q35.2	190,125,160-190,427,650	0.302	24	1.30E-02	1	*HSP90AA4P*	100
Chr 5	p15.1	16,888,013-16,994,509	0.106	20	0.00E+00	1	*MYO10*	5
	q12.1	61,583,390-61,735,907	0.153	22	0.00E+00	3	*DIMT1, IPO11, KIF2A*	100
	q13.2	68,428,997-68,564,076	0.135	27	0.00E+00	4	*CCNB1, CDK7, CENPH, MRPS36*	0
	q13.2	68,786,099-68,849,653	0.064	28	0.00E+00	1	*OCLN*	100
	q13.2	70,778,491-70,800,177	0.022	22	0.00E+00	1	*BDP1*	94
	q13.3	76,055,243-76,180,392	0.125	20	0.00E+00	2	*F2RL1, S100Z*	41
	q14.1	78,443,225-78,565,349	0.122	23	0.00E+00	1	*JMY*	25
	q14.1	79,573,522-79,634,971	0.061	24	0.00E+00	2	*LOC644936, SPZ1*	0
	q23.2	125,889,774-126,113,364	0.224	20	0.00E+00	4	*ALDH7A1, C5orf48, LMNB1, PHAX*	2
	q31.1	132,277,458-132,377,408	0.100	23	0.00E+00	2	*AFF4, ZCCHC10*	1
	q31.1	134,104,930-134,137,607	0.033	22	0.00E+00	1	*DDX46*	0
	q35.1	171,529,719-171,703,493	0.174	22	0.00E+00	3	*EFCAB9, STK10, UBTD2*	96
	q35.1	172,408,450-172,524,339	0.116	22	0.00E+00	3	*ATP6V0E1, CREBRF, SNORA74B*	10
	q35.2	176,522,090-176,585,808	0.064	23	0.00E+00	2	*FGFR4, NSD1*	78
Chr 6	p21.32	32,465,385-32,532,311	0.067	20	2.10E-02	2	*HLA-DRB5, HLA-DRB6*	100
	q13	74,050,467-74,067,397	0.017	27	0.00E+00	1	*DPPA5*	0
	q13	74,099,505-74,195,846	0.096	26	0.00E+00	3	*DDX43, MB21D1, MTO1*	2
	q14.1	76,462,056-76,510,198	0.048	22	1.10E-02	1	*MYO6*	0
	q15	90,497,575-90,599,843	0.102	23	0.00E+00	2	*CASP8AP2, MDN1*	46
	q21	107,444,141-107,492,203	0.048	27	0.00E+00	1	*PDSS2*	100
	q21	108,522,069-108,672,999	0.151	22	1.10E-02	2	*LACE1, SNX3*	94
	q21	110,991,764-111,325,700	0.334	22	1.10E-02	4	*AMD1, CDK19, GTF3C6, RPF2*	13
	q25.1	150,086,688-150,096,190	0.010	23	0.00E+00	1	*PCMT1*	0

APPENDIX 3

Table A2: Significant Common Focal Deletions in all 74 STS cases (STAC)

	Cytoband	Region	Length (Mb)	Freq %	p-value	Gene Count	Genes	% of CNV overlap
Chr 7	p22.1	6,058,613-6,125,961	0.067	22	3.00E-03	3	*AIMP2, ANKRD61, EIF2AK1*	100
	p22.1	6,312,742-6,401,731	0.089	20	3.60E-02	1	*FAM220A*	100
	p14.1	40,212,869-40,287,726	0.075	24	0.00E+00	1	*C7orf10*	76
	p13	44,805,831-44,819,347	0.014	22	3.00E-03	1	*ZMIZ2*	0
	p11.2	55,896,581-55,958,866	0.062	23	0.00E+00	1	*SEPT14*	100
	q36.1	152,298,801-152,428,614	0.130	20	0.00E+00	1	*XRCC2*	3
Chr 8	p12	30,521,548-30,639,277	0.118	31	0.00E+00	2	*GSR, UBXN8*	26
	p11.22	39,229,932-39,254,152	0.024	39	0.00E+00	1	*ADAM5*	100
	p11.22	39,354,153-39,368,808	0.015	41	0.00E+00	1	*ADAM3A*	100
Chr 9	p24.1	6,604,848-6,770,652	0.166	24	1.00E-03	2	*GLDC, KDM4C*	100
	p22.1	19,073,617-19,095,812	0.022	24	1.00E-03	1	*HAUS6*	100
	p22.1	19,328,279-19,336,610	0.008	22	4.90E-02	1	*DENND4C*	100
	p13.3	34,020,894-34,172,360	0.151	26	0.00E+00	2	*DCAF12, UBAP2*	28
	p13.3	36,197,774-36,210,370	0.013	23	8.00E-03	1	*CLTA*	100
	p13.3 - p13.2	36,238,083-36,602,433	0.364	23	8.00E-03	3	*GNE, MELK, RNF38*	60
	q21.32	86,538,847-86,579,313	0.040	20	1.10E-02	1	*C9orf64*	100
	q22.1	90,942,429-91,026,524	0.084	27	0.00E+00	1	*SPIN1*	3
	q22.32	99,068,141-99,166,239	0.098	20	1.10E-02	2	*SLC35D2, ZNF367*	10
	q31.2	110,185,871-110,312,825	0.127	22	1.00E-03	1	*KLF4*	3
	q31.3	114,225,250-114,374,017	0.149	22	1.00E-03	4	*KIAA0368, LRRC37A5P, PTGR1, ZNF483*	45
	q31.3 - q32	114,899,274-114,978,317	0.079	23	0.00E+00	2	*MIR3134, SUSD1*	4
	q33.3	127,912,225-128,098,776	0.187	28	0.00E+00	4	*GAPVD1, HSPA5, PPP6C, RABEPK*	5
	q34.11	131,209,624-131,324,980	0.115	20	1.10E-02	3	*GLE1, ODF2, SPTAN1*	2
	q34.11	131,586,233-131,592,085	0.006	31	0.00E+00	1	*C9orf114*	100
	q34.3	137,970,025-138,031,761	0.062	20	1.10E-02	1	*OLFM1*	10
	q34.3	140,569,189-140,833,019	0.264	22	1.00E-03	4	*CACNA1B, EHMT1, FLJ40292, MIR602*	41
Chr 10	p12.31	21,824,360-21,930,936	0.107	39	1.70E-02	1	*MLLT10*	47
	q21.3	70,020,324-70,058,211	0.038	39	1.00E-02	1	*PBLD*	0
	q21.3	70,117,022-70,139,117	0.022	43	0.00E+00	1	*RUFY2*	0
	q21.3	70,448,059-70,560,756	0.113	42	0.00E+00	3	*CCAR1, SNORD98, TET1*	91
	q22.1	70,612,594-70,658,476	0.046	42	0.00E+00	1	*STOX1*	41
	q22.1	74,378,863-74,431,837	0.053	41	1.00E-03	1	*MICU1*	0
	q22.1	74,695,200-74,882,869	0.188	39	1.00E-02	3	*NUDT13, P4HA1, PLA2G12B*	1
	q22.2	75,088,327-75,190,252	0.102	38	4.90E-02	3	*ANXA7, MSS51, TTC18*	3
	q22.2	75,979,373-76,068,813	0.089	38	4.90E-02	1	*ADK*	100
	q22.2	76,539,809-76,624,155	0.084	39	1.00E-02	1	*KAT6B*	4
	q23.33	94,118,854-94,208,531	0.090	38	4.90E-02	1	*MARK2P9*	31
	q23.33	96,313,938-96,405,785	0.092	38	4.90E-02	1	*HELLS*	0
	q23.33	96,483,383-96,501,601	0.018	38	4.90E-02	1	*CYP2C18*	0
	q24.1	98,598,111-98,639,457	0.041	38	4.90E-02	1	*LCOR*	48
Chr 11	p15.5	975,029-1,054,782	0.080	26	3.00E-03	2	*AP2A2, MUC6*	100
	p15.4	3,752,511-3,833,652	0.081	24	2.90E-02	2	*NUP98, PGAP2*	60
	p15.4	9,291,259-9,433,407	0.142	32	0.00E+00	2	*IPO7, TMEM41B*	27
	p15.1	17,216,087-17,328,061	0.112	24	2.90E-02	1	*NUCB2*	5
	p15.1	18,384,389-18,566,743	0.182	26	3.00E-03	6	*GTF2H1, LDHA, LDHAL6A, LDHC, TSG101, UEVLD*	13
	p11.2	47,685,266-47,775,861	0.091	26	3.00E-03	2	*AGBL2, FNBP4*	0
	q22.3	107,770,408-107,898,857	0.128	45	0.00E+00	2	*CUL5, RAB39A*	100
	q23.1	112,001,948-112,013,997	0.012	35	4.70E-02	1	*IL18*	0
	q23.3	116,854,116-117,006,852	0.153	36	9.00E-03	1	*SIK3*	4
	q23.3	118,598,334-118,631,072	0.033	35	4.70E-02	1	*DDX6*	0

APPENDIX 3

Table A2: Significant Common Focal Deletions in all 74 STS cases (STAC)

	Cytoband	Region	Length (Mb)	Freq %	p-value	Gene Count	Genes	% of CNV overlap
Chr 11	q24.1	122,858,214-122,972,257	0.114	36	9.00E-03	3	*CLMP, HSPA8, LOC341056*	8
	q24.1	123,024,491-123,037,487	0.013	36	9.00E-03	1	*CLMP*	0
	q24.2	125,355,927-125,367,110	0.011	35	4.70E-02	1	*FEZ1*	19
	q24.2	125,561,636-126,152,231	0.591	35	4.70E-02	12	*CDON, DDX25, FAM118B, FOXRED1, HYLS1, PATE1, PATE2, PATE3, PATE4, PUS3, RPUSD4, SRPR*	65
	q24.3	129,941,688-130,298,215	0.357	36	9.00E-03	4	*ADAMTS8, APLP2, ST14, ZBTB44*	100
Chr 12	p13.33	1,087,151-1,143,274	0.056	27	0.00E+00	1	*ERC1*	12
	p13.31	7,837,530-7,904,204	0.067	38	0.00E+00	3	*CLEC4C, DPPA3, GDF3*	100
	p13.2	12,378,259-12,486,780	0.109	28	0.00E+00	2	*LRP6, MANSC1*	24
	p13.1	13,005,969-13,035,585	0.030	28	0.00E+00	1	*RPL13AP20*	4
	p12.3	19,494,831-19,609,167	0.114	24	2.90E-02	2	*AEBP2, PLEKHA5*	100
	q13.12	51,127,712-51,247,966	0.120	20	0.00E+00	3	*ATF1, DIP2B, TMPRSS12*	1
	q23.2	101,863,963-101,904,060	0.040	22	0.00E+00	1	*SPIC*	2
	q24.23	118,507,406-118,519,199	0.012	23	0.00E+00	1	*VSIG10*	0
	q24.31	122,469,774-122,665,202	0.195	22	0.00E+00	4	*BCL7A, IL31, LRRC43, MLXIP*	8
	q24.31	123,975,245-124,001,234	0.026	20	0.00E+00	1	*RILPL1*	0
Chr 13	q14.13	45,907,934-45,944,432	0.036	50	0.00E+00	3	*SNORA31, TPT1, TPT1-AS1*	4
	q14.13	46,052,795-46,062,566	0.010	45	2.50E-02	1	*COG3*	1
	q14.2	48,961,498-49,034,756	0.073	49	0.00E+00	2	*LPAR6, RB1*	2
	q14.2	49,229,080-49,281,746	0.053	47	0.00E+00	1	*CYSLTR2*	100
	q14.2	50,015,335-50,058,363	0.043	47	0.00E+00	2	*CAB39L, SETDB2*	1
	q14.2	50,123,069-50,223,067	0.100	47	0.00E+00	2	*ARL11, RCBTB1*	1
	q14.2	50,395,858-50,553,767	0.158	51	0.00E+00	2	*CTAGE10P, SPRYD7*	19
	q14.3	52,907,036-52,975,980	0.069	45	2.50E-02	1	*THSD1*	100
	q22.1	73,587,724-73,617,924	0.030	46	4.00E-03	1	*PIBF1*	53
Chr 14	q11.2	20,224,270-20,405,952	0.182	38	0.00E+00	5	*OR4K1, OR4K2, OR4K5, OR4M1, OR4N2*	100
	q13.1	35,045,934-35,056,862	0.011	32	0.00E+00	1	*SNX6*	100
	q23.2	64,024,743-64,309,341	0.285	31	0.00E+00	2	*SGPP1, WDR89*	16
	q24.2	73,442,657-73,558,351	0.116	28	3.40E-02	2	*RBM25, ZFYVE1*	4
	q24.3	73,996,129-74,029,564	0.033	30	4.00E-03	2	*ACOT1, HEATR4*	100
	q24.3	74,281,839-74,537,345	0.256	32	0.00E+00	7	*ALDH6A1, CCDC176, COQ6, ENTPD5, FAM161B, PTGR2, ZNF410*	0
	q32.33	104,048,530-104,117,267	0.069	28	3.40E-02	2	*APOPT1, KLC1*	0
	q32.33	104,566,967-104,715,087	0.148	31	0.00E+00	4	*ASPG, KIF26A, MIR203, MIR3545*	63
Chr 15	q11.2	21,217,563-22,134,343	0.917	27	0.00E+00	4	*CXADRP2, LOC646214, NF1P2, POTEB*	100
	q11.2	22,311,626-22,395,369	0.084	28	0.00E+00	3	*LOC727924, OR4M2, OR4N4*	100
	q11.2	22,550,035-22,662,192	0.112	32	0.00E+00	1	*REREP3*	100
	q15.1	41,409,196-41,592,148	0.183	24	9.00E-03	3	*CHP1, EXD1, OIP5-AS1*	12
	q15.1	41,665,226-41,739,919	0.075	24	9.00E-03	3	*NDUFAF1, NUSAP1, RTF1*	0
	q15.3	44,479,587-44,509,992	0.030	27	0.00E+00	1	*FRMD5*	0
	q15.3	44,571,354-44,655,162	0.084	27	0.00E+00	1	*CASC4*	0
	q21.2	50,726,619-50,862,207	0.136	26	1.00E-03	3	*TRPM7, USP50, USP8*	68
	q21.2	50,874,701-50,958,619	0.084	26	1.00E-03	1	*TRPM7*	100
	q21.2	51,052,550-51,089,434	0.037	26	1.00E-03	1	*SPPL2A*	100
	q21.3	55,734,880-55,762,580	0.028	31	0.00E+00	2	*DYX1C1, DYX1C1-CCPG1*	0
	q22.31	64,769,454-64,836,946	0.067	28	0.00E+00	1	*ZNF609*	3

APPENDIX 3

Table A2: Significant Common Focal Deletions in all 74 STS cases (STAC)

	Cytoband	Region	Length (Mb)	Freq %	p-value	Gene Count	Genes	% of CNV overlap
Chr 16	q21	58,612,157-58,835,213	0.223	34	5.00E-03	3	*CNOT1, GOT2, SLC38A7*	100
	q22.1	68,761,997-68,928,133	0.166	32	3.50E-02	2	*CDH1, TANGO6*	41
	q22.1	69,215,073-69,271,226	0.056	32	3.50E-02	1	*SNTB2*	46
	q23.1	74,366,198-74,436,209	0.070	34	5.00E-03	1	*LOC283922*	100
	q23.2	81,181,342-81,195,711	0.014	32	3.50E-02	1	*PKD1L2*	100
Chr 17	p13.3	172,499-344,123	0.172	34	0.00E+00	4	*C17orf97, FAM101B, LOC100506388, RPH3AL*	70
	p13.3	511,145-555,195	0.044	32	1.00E-03	1	*VPS53*	20
	p13.3	2,665,936-2,769,575	0.104	32	1.00E-03	1	*RAP1GAP2*	23
	p13.2	5,136,618-5,181,043	0.044	31	3.00E-03	2	*LOC100130950, SCIMP*	3
	p13.1	7,565,085-7,611,407	0.046	31	3.00E-03	3	*EFNB3, TP53, WRAP53*	0
	q11.2	27,735,232-27,764,959	0.030	28	0.00E+00	1	*TAOK1*	0
	q11.2	27,984,473-28,129,484	0.145	24	1.00E-03	1	*SSH2*	0
	q11.2	29,117,045-29,216,285	0.099	31	0.00E+00	2	*ATAD5, CRLF3*	1
	q11.2	30,161,501-30,258,051	0.097	26	0.00E+00	2	*COPRS, UTP6*	38
	q12	35,710,467-35,718,781	0.008	35	0.00E+00	1	*ACACA*	100
	q12	37,461,700-37,519,729	0.058	22	2.20E-02	1	*FBXL20*	26
Chr 18	q11.2	19,209,847-19,239,318	0.029	24	1.00E-03	2	*ABHD3, SNRPD1*	100
Chr 19	q13.32	47,802,665-47,833,309	0.031	24	2.00E-02	1	*C5AR1*	2
	q13.33	48,041,626-48,139,184	0.098	24	2.00E-02	2	*GLTSCR1, ZNF541*	0
	q13.33	48,255,738-48,563,754	0.308	24	2.00E-02	39	*BSPH1, CABP5, CRX, ELSPBP1, GLTSCR2, PLA2G4C, SEPW1, SNAR-A1, SNAR-A10, SNAR-A11, SNAR-A12, SNAR-A13, SNAR-A14, SNAR-A2, SNAR-A3, SNAR-A4, SNAR-A5, SNAR-A6, SNAR-A7, SNAR-A8, SNAR-A9, SNAR-C1, SNAR-C2, SNAR-C3, SNAR-C4, SNAR-C5, SNORD23, SULT2A1, TPRX1*	61
	q13.41	53,501,507-53,551,665	0.050	24	2.00E-02	2	*ERVV-1, ERVV-2*	100
	q13.42	54,235,078-54,300,229	0.065	35	0.00E+00	18	*MIR1283-2, MIR371A, MIR371B, MIR372, MIR373, MIR516A1, MIR516A2, MIR516B1, MIR517C, MIR518A2, MIR518D, MIR519A1, MIR519A2, MIR520H, MIR521-1, MIR522, MIR527, NLRP12*	45
Chr 20	p12.3	5,677,855-5,749,604	0.072	23	0.00E+00	1	*C20orf196*	80
Chr 21	p11.1	11,066,993-11,132,953	0.066	28	0.00E+00	5	*BAGE, BAGE2, BAGE3, BAGE4, BAGE5*	100
	q22.11	34,789,786-34,886,279	0.096	24	1.00E-03	4	*DNAJC28, GART, IFNGR2, TMEM50B*	41
	q22.12	37,616,924-37,740,911	0.124	28	0.00E+00	2	*DOPEY2, MORC3*	100
	q22.3	45,818,895-45,899,492	0.081	23	1.20E-02	3	*LRRC3, LRRC3-AS1, TRPM2*	100
Chr 22	q11.1	17,721,150-17,873,514	0.152	30	9.00E-03	1	*CECR3*	24
	q12.1	29,021,468-29,050,839	0.029	31	1.00E-03	1	*TTC28*	0
	q12.1	29,060,638-29,160,547	0.100	31	1.00E-03	3	*CHEK2, HSCB, TTC28*	54
	q12.2	31,819,400-31,830,361	0.011	30	9.00E-03	1	*DRG1*	100
	q13.32	49,017,773-49,105,738	0.088	32	0.00E+00	1	*FAM19A5*	45
						510		

APPENDIX 3

Table B1: Significant Common Focal Amplifications in all 74 STS cases (GISTIC)

	Peak Region	Extended Region	Q-Bound	G-Score	No. of Genes	Frequency %	Genes in Peak	All genes
Chr 1	172,098,669-172,150,518	172,028,886-172,405,472	5.71E-05	18.4	6	35	*DNM3, MIR312, MIR214, DNM3OS, MIR199A2,*	*DNM3, MIR312, MIR214, DNM3OS, MIR199A2, C1orf105*
	145,112,508-145,166,867	144,009,907-145,340,062	3.85E-03	14.7	14	43	*SEC22B*	*SRGAP2B, LINC00623, LOC728875, PPIAL4B, PPIAL4C, PPIAL4A, LOC728875, PFN1P2, LOC653513, NBPF9, PDE4DIP, SEC22B, NOTCH2NL, NBPF10*
	59,247,642-59,633,875	59,099,590-61,974,758	1.20E-02	13.0	10	32	*JUN, LOC100131060, HSD52*	*MYSM1, JUN, LOC100131060, HSD52, FGGY, MIR4711, HOOK1, CYP2J2, C1orf87, NFIA*
Chr 2	91,799,855-91,906,643	91,693,395-91,906,643	4.03E-02	11.2	1	34	*LOC654342*	*LOC654342*
Chr 3	87,299,611-87,306,626	86,816,036-88,468,010	3.81E-02	11.3	8	21	*CHMP2B*	*VGLL3, MIR4795, CHMP2B, POU1F1, HTR1F, CGGBP1, ZNF654, C3orf38*
Chr 5	14,207,189-14,318,260	11,206,668-14,990,016	3.57E-04	16.9	9	42	*TRIO*	*CTNND2, CT49, DNAH5, TRIO, FAM105A, FAM105B, LOC100130744, ANKH, MIR4637*
Chr 7	38,291,860-38,364,472	38,282,545-38,385,111	1.43E-04	17.6	2	57	*TARP*	*TARP, LOC100506776*
	82,845,459-83,425,216	80,923,212-86,580,804	1.52E-02	12.7	8	28	*SEMA3E*	*HGF, CACNA2D1, PCLO, SEMA3E, SEMA3A, SEMA3D, GRM3, KIAA1324L*
Chr 8	39,267,045-39,326,605	39,254,152-39,368,808	1.89E-02	12.4	2	30	*ADAM3A*	*ADAM5, ADAM3A*
	72,348,163-72,380,339	72,104,731-73,824,911	1.89E-02	12.4	7	20	*EYA1, MSC, LOC100132891, RNU6-83, TRPA1, LOC392232, KCNB2*	*EYA1, MSC, LOC100132891, RNU6-83, TRPA1, LOC392232, KCNB2*
Chr 11	101,888,922-102,084,901	101,597,949-102,391,460	7.75E-06	19.9	8	<20	*C11orf70, YAP1*	*ANGPTL5, KIAA1377, C11orf70, YAP1, BIRC3, BIRC2, TMEM123, MMP7*
Chr 12	69,214,969-69,274,789	69,152,516-69,644,424	1.13E-08	33.2	5	<20	*MDM2, CPM*	*SLC35E3, LOC100130075, MDM2, CPM, CPSF6*
	58,177,011-58,193,558	58,114,609-58,265,541	1.13E-08	26.6	13	<20	*TSFM, AVIL*	*OS9, LOC100130776, AGAP2, TSPAN31, CDK4, MARCH9, CYP27B1, METTL1, METTL21B, TSFM, AVIL, MIR26A2, CTDSP2*
	72,779,573-72,832,009	72,603,962-73,912,299	9.53E-03	13.3	2	<20	*TRHDE*	*TRHDE-AS1, TRHDE*
Chr 15	21,217,563-21,618,233	20,523,183-22,311,626	1.45E-04	17.6	12	<20	-	*HERC2P3, GOLGA6L6, GOLGA8CP, NBEAP1, POTEB, NF1P2, CT60, LOC646214, CXADRP2, POTEB, NF1P2, LOC727924*
Chr 16	704,825-751,634	601,203-804,072	4.44E-02	11.1	20	30	*WDR90, RHOT2, RHBDL1, STUB1, JMJD8, WDR24, FBXL16*	*SOLH, C16orf11, NHLRC4, PIGQ, RAB40C, WFIKKN1, C16orf13, FAM195A, WDR90, RHOT2, RHBDL1, STUB1, JMJD8, WDR24, FBXL16, METRN, FAM173A, CCDC78, HAGHL, NARFL*
Chr 17	17,754,230-17,784,405	17,725,378-17,822,932	4.62E-03	14.4	2	31	*TOM1L2*	*SREBF1, TOM1L2*
Chr 20	48,909,768-49,005,661	48,776,558-49,195,548	1.35E-02	12.9	3	32	*LOC284751*	*CEBPB, LOC284751, PTPN1*
Chr 22	46,449,007-46,491,493	46,439,878-46,507,701	3.50E-02	11.4	5	34	*C22orf26, LOC150381, MIR3619, MIRLET7BHG*	*LOC100271722, C22orf26, LOC150381, MIR3619, MIRLET7BHG*
					137			

APPENDIX 3

Table B2: Significant Common Focal Deletions in all 74 STS cases (GISTIC)

	Peak Region	Extended Region	Q-Bound	G-Score	No of Genes	Frequency %	Genes in Peak	All genes
Chr 1	248,733,443-248,796,977	248,733,443-248,796,977	1.05E-06	20.3	3	24	*OR2T34, OR2T10, OR2T11*	*OR2T34, OR2T10, OR2T11*
	196,790,444-196,812,218	196,790,444-196,812,218	3.38E-02	11.3	1	26	*CFHR1*	*CFHR1*
Chr 2	242,930,739-242,953,346	242,859,272-243,199,373	2.10E-09	26.0	1	41	-	*LOC728323*
Chr 4	69,307,839-69,405,506	69,307,839-69,546,914	1.37E-03	14.8	3	22	*TMPRSS11E, UGT2B17,*	*TMPRSS11E, UGT2B17, UGT2B15*
Chr 6	32,465,385-32,532,311	32,465,385-32,532,311	7.59E-07	20.5	2	20	*HLA-DRB5, HLA-DRB6*	*HLA-DRB5, HLA-DRB6*
Chr 8	39,354,153-39,368,808	39,349,089-39,383,632	2.08E-09	58.5	1	41	*ADAM3A*	*ADAM3A*
Chr 9	21,936,953-22,013,183	21,824,934-23,653,364	3.01E-02	11.4	6	<20	*C9orf53, CDKN2A, CDKN2B, CDKN2B-AS1*	*MTAP, C9orf53, CDKN2A CDKN2B, CDKN2B-AS1, DMRTA1, FLJ35282*
Chr 10	70,612,594-70,641,422	70,117,022-70,813,318	1.85E-04	16.9	9	43	*STOX1*	*RUFY2, DNA2, SLC25A16, TET1, SNORD98, CCAR1, STOX1, DDX50, DDX21*
	21,824,360-21,930,936	21,824,360-21,930,936	5.43E-04	15.8	1	39	*MLLT10*	*MLLT10*
Chr 11	55,393,589-55,416,668	55,381,793-55,444,601	8.32E-07	20.5	3	26	*OR4P4*	*OR4P4, OR4S2, OR4C6*
	134,790,729-135,006,516	128,837,921-135,006,516	1.31E-03	14.9	29	36	-	*ARHGAP32, BARX2, TMEM45B, NFRKB, PRDM10, LINC00167, APLP2, ST14, ZBTB44, ADAMTS8, ADAMTS15, C11orf44, SNX19, NTM, OPCML, SPATA19, MIR4697, LOC283174, IGSF9B, LOC100128239, JAM3, NCAPD3, VPS26B, THYN1, ACAD8, GLB1L3, GLB1L2, B3GAT1, LOC283177*
	9,291,259-9,433,407	9,230,469-9,519,928	1.25E-02	12.3	6	32	*TMEM41B, IPO7*	*DENND5A, TMEM41B, IPO7, SNORA23, LOC644656, ZNF143*
Chr 13	48,992,576-49,034,756	48,413,942-49,467,005	2.08E-09	35.5	11	49	*RB1, LPAR6*	*SUCLA2, NUDT15, MED4-AS1, MED4, ITM2B, LINC00441, RB1, LPAR6, RCBTB2, LINC00462, CYSLTR2*
Chr 14	20,224,270-20,369,804	19,376,762-20,420,707	1.74E-02	11.9	15	38	*OR4M1 OR4N2 OR4K2*	*OR11H12, LOC642426, POTEG, LINC00516, LOC101101776, LINC00516, LOC101101776, POTEM, OR11H2, OR4Q3, OR4M1, OR4N2, OR4K2, OR4K5, OR4K1*
Chr 15	22,505,176-22,534,833	22,497,998-22,662,192	1.48E-02	12.1	1	32	-	*REREP3*
Chr 16	74,366,198-74,436,209	73,260,280-74,790,855	1.72E-03	14.6	8	31	*LOC283922*	*LOC100506172, PSMD7, LOC283922, CLEC18B, GLG1, RFWD3, MLKL, FA2H*

APPENDIX 3

Table B2: Significant Common Focal Deletions in all 74 STS cases (GISTIC)

	Peak Region	Extended Region	Q-Bound	G-Score	No of Genes	Frequency %	Genes in Peak	All genes
Chr 17	7,611,407-7,618,079	7,111,674-7,650,119	1.34E-05	18.7	50	31	*EFNB3*	*DLG4, ACADVL, MIR324, DVL2, PHF23, GABARAP, CTDNEP1, ELP5, CLDN7, SLC2A4, YBX2, EIF5A, GPS2, NEURL4, ACAP1, KCTD11, TMEM95, TNK1, PLSCR3, C17orf61-PLSCR3, TMEM256, NLGN2, SPEM1, C17orf74, TMEM102, FGF11, CHRNB1, ZBTB4, SLC35G6, POLR2A, TNFSF12, TNFSF12-TNFSF13, TNFSF13, SENP3, SENP3-EIF4A1, SNORA48, EIF4A1, SNORD10, SNORA67, CD68, MPDU1, SOX15, FXR2, SHBG, SAT2, ATP1B2, TP53, WRAP53, EFNB3, DNAH2*
	35,710,467-35,718,781	35,672,592-35,819,066	7.64E-03	12.9	3	35	*ACACA*	*ACACA, C17orf78, TADA2A*
Chr 18	77,944,237-78,077,248	77,622,155-78,077,248	1.70E-02	12.0	8	22	*PARD6G*	*KCNG2, PQLC1, HSBP1L1, TXNL4A, RBFA, ADNP2, PARD6G-AS1, PARD6G*
Chr 19	54,235,078-54,300,229	52,885,325-55,554,280	1.23E-02	12.3	146	45	*MIR518D MIR516B1 MIR518A2 MIR517C MIR520H MIR521-1 MIR522 MIR519A1 MIR527 MIR516A1 MIR1283-2 MIR516A2 MIR519A2 MIR371A MIR371B MIR372 MIR373 NLRP12*	*ZNF880, ZNF528, ZNF534, ZNF578, ZNF808, ZNF701, ZNF137P, ZNF83, ZNF611, ZNF600, ZNF28, ZNF468, ZNF320, ZNF321P, ZNF816-ZNF321P, ZNF816, ZNF702P, ERVV-1, ERVV-2, ZNF160, ZNF415, ZNF347, ZNF665, ZNF818P, ZNF677, VN1R2, VN1R4, FAM90A27P, BIRC8, ZNF845, ZNF525, ZNF765, TPM3P9, ZNF761, ZNF813, ZNF331, LOC284379, DPRX, MIR512-1, MIR512-2, MIR512-1, MIR512-2, MIR1323, MIR498, MIR520E, MIR515-1, MIR515-2, MIR519E, MIR520F, MIR515-1, MIR515-2, MIR519C, MIR1283-1, MIR520A, MIR526B, MIR519B, MIR525, MIR523, MIR518F, MIR520B, MIR518B, MIR526A1, MIR520C, MIR518C, MIR524, MIR517A, MIR519D, MIR521-2, MIR520D, MIR517B, MIR520G, MIR516B2, MIR526A2, MIR518E, MIR518A1, MIR518D, MIR516B1, MIR518A2, MIR517C, MIR520H, MIR521-1, MIR522, MIR519A1, MIR527, MIR516A1, MIR1283-2, MIR516A2, MIR519A2, MIR371A, MIR371B, MIR372, MIR373, NLRP12, MYADM, PRKCG, CACNG7, CACNG8, MIR935, CACNG6, VSTM1, TARM1, OSCAR, NDUFA3, TFPT, PRPF31, CNOT3, LENG1, TMC4, MBOAT7, TSEN34, RPS9, LILRB3, LILRA6, LILRB5, LILRB2, MIR4752, LILRA3, LILRA5, LILRA4, LAIR1, TTYH1, LENG8, LENG9, CDC42EP5, LAIR2, KIR3DX1, LILRA2, LILRA1, LILRB1, LILRB4, LILRP2, KIR3DL3, KIR2DL3, KIR2DL1, LOC100287534, KIR2DL4, KIR3DL1, KIR2DS4, KIR3DL2, FCAR, NCR1, NLRP7, RNU6-35, RNU6-64, NLRP2, GP6*
Chr 21	11,066,993-11,132,953	10,948,745-11,132,953	9.58E-03	12.6	5	28	*BAGE, BAGE2, BAGE3, BAGE4, BAGE5*	*BAGE, BAGE2, BAGE3, BAGE4, BAGE5*
Chr 22	24,342,813-24,398,622	24,342,813-24,398,622	7.25E-07	20.6	5	<20	*GSTTP1, LOC3913, 22, GSTT1, GSTTP2*	*GSTTP1, LOC3913, 22, GSTT1, GSTTP2*
					317			

APPENDIX 4

Table A1: Significant Common Focal Amplifications in Leiomyosarcoma (STAC)

	Cytoband	Region	Length (Mb)	Freq %	P-Value	Gene Count	Genes	% of CNV Overlap
Chr 1	p12 - p11.2	120,563,106-120,756,223	0.19	48	0.0E+00	1	*NOTCH2*	100
	q21.1	144,009,907-144,995,110	0.99	71	0.0E+00	8	*NBPF9, PDE4DIP, PFN1P2, PPIAL4A, PPIAL4B, PPIAL4C, SRGAP2B*	100
	q21.1	145,112,508-145,278,200	0.17	76	0.0E+00	2	*NOTCH2NL, SEC22B*	100
	q21.2	149,311,087-149,596,705	0.29	62	1.6E-02	6	*FAM91A2, FCGR1C, HIST2H2BF, PPIAL4A, PPIAL4B, PPIAL4C*	100
	q21.2	149,818,426-149,873,111	0.05	62	1.6E-02	11	*BOLA1, HIST2H2AA3, HIST2H2AA4, HIST2H2AB, HIST2H2AC, HIST2H2BC, HIST2H2BE, HIST2H3A, HIST2H3C, HIST2H4A, HIST2H4B*	100
	q21.3	154,899,255-154,958,577	0.06	62	1.6E-02	6	*CKS1B, FLAD1, PBXIP1, PMVK, PYGO2, SHC1*	0
	q22	155,177,125-155,253,539	0.08	62	1.6E-02	8	*CLK2, FAM189B, GBA, GBAP1, HCN3, MTX1, SCAMP3, THBS3*	100
	q22	155,894,065-156,045,700	0.15	62	1.6E-02	8	*ARHGEF2, LAMTOR2, MEX3A, RAB25, RXFP4, SCARNA4, SSR2, UBQLN4*	16
	q24.2	168,235,779-168,258,696	0.02	62	1.6E-02	1	*TBX19*	100
Chr 2	q22.3	145,162,871-145,304,149	0.14	29	1.6E-02	1	*ZEB2*	3
	p22.2	37,823,758-38,245,167	0.42	33	9.0E-03	2	*CDC42EP3, FAM82A1*	75
	q11.2	97,527,264-97,623,299	0.01	38	0.0E+00	2	*FAM178B, SEMA4C*	54
	q13.13	048,766,079-49,005,661	0.24	52	0.0E+00	3	*CEBPB, TMEM189, TMEM189-UBE2V1*	65
Chr 3	q26.31	171,958,976-172,051,964	0.09	52	3.0E-03	1	*FNDC3B*	0
Chr 5	p13.2	37,012,156-37,067,386	0.06	52	1.5E-02	1	*NIPBL*	0
	q11.2	54,728,700-54,928,310	0.20	43	4.3E-02	3	*PPAP2A, RNF138P1, SLC38A9*	70
	q12.3	64,264,133-64,412,850	0.15	43	4.3E-02	1	*CWC27*	1
	q13.1	67,067,899-67,548,574	0.48	43	4.3E-02	1	*PIK3R1*	18
	q13.2	71,430,376-71,606,831	0.18	48	1.0E-03	2	*MAP1B, MRPS27*	6
	q14.3	90,402,321-90,686,262	0.28	48	1.0E-03	2	*ARRDC3, GPR98*	8
	q32	148,717,819-148,929,994	0.21	43	4.3E-02	5	*AFAP1L1, CSNK1A1, GRPEL2, IL17B, PCYOX1L*	55
Chr 6	p22.2	26,124,423-26,259,251	0.13	38	7.0E-03	17	*HIST1H1D, HIST1H1E, HIST1H2AC, HIST1H2AD, HIST1H2AE, HIST1H2BD, HIST1H2BE, HIST1H2BF, HIST1H2BG, HIST1H2BH, HIST1H3D, HIST1H3E, HIST1H3F, HIST1H4D, HIST1H4E, HIST1H4F, HIST1H4G*	47
Chr 9	p22.1	19,095,812-19,317,927	0.22	29	3.3E-02	3	*DENND4C, HAUS6, PLIN2*	100
	q22.31	94,034,499-94,186,403	0.15	38	1.2E-02	2	*AUH, NFIL3*	36

APPENDIX 4

Table A1: Significant Common Focal Amplifications in Leiomyosarcoma (STAC)

	Cytoband	Region	Length (Mb)	Freq %	P-Value	Gene Count	Genes	% of CNV Overlap
Chr 11	p13	135,110,040-35,316,017	0.20	29	6.0E-03	2	*CD44, SLC1A2*	63
	q12.3	162,314,338-62,641,063	0.33	24	3.0E-03	35	*B3GAT3, BSCL2, EEF1G, EML3, GANAB, GNG3, HNRNPUL2, HNRNPUL2-BSCL2, INTS5, LRRN4CL, METTL12, MTA2, NXF1, POLR2G, ROM1, SLC3A2, SNHG1, SNORA57, SNORD22, SNORD25, SNORD26, SNORD27, SNORD28, SNORD29, SNORD30, SNORD31, STX5, TAF6L, TMEM179B, TMEM223, TTC9C, TUT1, UBXN1, WDR74, ZBTB3*	85
Chr 12	q24.11	2109,123,797-109,147,883	0.02	33	4.0E-03	1	*CORO1C*	0
Chr 13	q34	3110,918,305-111,046,505	0.13	43	0.0E+00	2	*COL4A1, COL4A2*	8
Chr 14	q22.1	451,964,027-52,127,715	0.16	57	2.0E-03	1	*FRMD6*	5
	q24.1	469,144,273-69,519,867	0.38	57	2.0E-03	3	*ACTN1, DCAF5, ZFP36L1*	2
Chr 16	q22.1	669,228,790-69,753,047	0.52	29	4.3E-02	10	*COG8, CYB5B, NFAT5, NIP7, NQO1, PDF, SNTB2, TERF2, TMED6, VPS4A*	18
	q22.3	672,964,471-73,143,118	0.18	33	1.0E-03	2	*HTA, ZFHX3*	28
	q24.1	684,963,962-85,068,069	0.10	33	1.0E-03	1	*ZDHHC7*	3
Chr 17	q22	755,948,928-56,056,018	0.11	38	2.3E-02	2	*CUEDC1, VEZF1*	40
	q23.1	757,816,707-57,935,524	0.12	38	2.3E-02	1	*VMP1*	2
Chr 19	p13.11	918,429,433-18,583,639	0.15	48	3.0E-03	7	*ELL, GDF15, ISYNA1, LRRC25, LSM4,, PGPEP1, SSBP4*	0
Chr 22	q12.3	236,665,486-36,786,805	0.12	48	1.3E-02	1	*MYH9*	45
						165		

APPENDIX 4

Table A2: Significant Common Focal Deletions in Leiomyosarcoma (STAC)

	Cytoband	Region	Length (Mb)	Freq %	P-Value	Gene Count	Genes	% of CNV Overlap
Chr 1	p36.33	1,075,323-1,151,667	0.08	62	0.00E+00	3	*TNFRSF18, TNFRSF4, TTLL10*	100
	p36.33	1,870,020-2,195,283	0.33	38	9.00E-03	3	*GABRD, PRKCZ, SKI*	88
	p36.32	2,472,466-2,539,213	0.07	52	0.00E+00	3	*FAM213B, MMEL1, TNFRSF14*	100
	p36.32	3,184,391-3,448,830	0.26	48	0.00E+00	3	*ARHGEF16, MEGF6, PRDM16*	100
	p36.32	4,793,782-5,354,520	0.56	38	9.00E-03	1	*AJAP1*	26
	q44	248,983,071-249,250,621	0.27	33	0.00E+00	4	*PGBD2, SH3BP5L, ZNF672, ZNF692*	78
Chr 2	q36.3	229,013,827-229,907,621	0.89	48	2.40E-02	2	*PID1, SPHKAP*	10
	q36.3	230,795,681-230,895,774	0.10	48	2.40E-02	1	*FBXO36*	100
	q37.1	233,213,865-233,427,159	0.21	52	1.00E-03	10	*ALPI, ALPP, ALPPL2, CHRND, CHRNG, ECEL1, ECEL1P2, EIF4E2, PRSS56, TIGD1*	99
	q37.1	233,903,041-234,336,019	0.43	48	2.40E-02	6	*ATG16L1, DGKD, INPP5D, SAG, SCARNA5, SCARNA6*	13
	q37.1	234,507,893-235,056,307	0.55	48	2.40E-02	14	*DNAJB3, HJURP, MSL3P1, SPP2, TRPM8, UGT1A1, UGT1A10, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9*	16
	q37.2	235,880,185-236,368,103	0.49	52	1.00E-03	1	*SH3BP4*	0
	q37.3	241,538,072-241,782,200	0.24	57	0.00E+00	5	*AQP12A, AQP12B, CAPN10, GPR35, KIF1A*	100
	q37.3	242,503,450-242,825,417	0.32	57	0.00E+00	10	*ATG4B, BOK, CXXC11, D2HGDH, DTYMK, GAL3ST2, ING5, NEU4, PDCD1, THAP4*	100
Chr 4	p16.1	8,771,782-8,886,137	0.11	33	2.00E-03	1	*HMX1*	100
Chr 6	p25.3	1,696,679-1,919,998	0.22	38	1.90E-02	1	*GMDS*	0
Chr 8	q24.3	143,656,045-144,082,427	0.43	33	1.90E-02	12	*ARC, CYP11B1, CYP11B2, GML, JRK, LY6D, LY6K, LYNX1, LYPD2, PSCA, SLURP1, THEM6*	71
Chr 9	q34.11 - q34.12	133,145,616-133,553,257	0.41	43	3.80E-02	3	*ASS1, FUBP3, PRDM12*	41
	q34.3	137,870,941-138,031,761	0.16	48	2.00E-03	1	*OLFM1*	6
	q34.3	140,630,365-140,833,019	0.20	57	0.00E+00	2	*CACNA1B, EHMT1*	33
Chr 11	p15.5	1975,029-1,093,963	0.12	43	3.00E-03	3	*AP2A2, MUC2, MUC6*	100
Chr 12	p13.31	27,796,205-7,921,641	0.13	38	2.60E-02	5	*APOBEC1, CLEC4C, DPPA3, GDF3, NANOGNB*	100
Chr 14	q32.33	4104,561,057-104,763,502	0.20	33	2.00E-03	2	*ASPG, KIF26A*	70
Chr 19	p13.3	90-307,937	0.31	29	2.00E-03	6	*FAM138A, FAM138F, MIER2, OR4F17, PPAP2C, WASH5P*	80
Chr 20	q13.33	060,072,327-60,546,948	0.47	29	0.00E+00	1	*CDH4*	48
	q13.33	061,064,386-61,378,893	0.31	24	2.30E-02	2	*NTSR1, SLCO4A1*	100
	q13.33	061,638,941-61,818,375	0.18	24	2.30E-02	2	*HAR1A, HAR1B,*	100
	q13.33	062,844,161-63,025,520	0.18	33	0.00E+00	2	*MYT1, PCMTD2*	67
Chr 21	q22.3	145,818,895-45,951,075	0.13	33	0.00E+00	3	*LRRC3, TRPM2, TSPEAR*	100
Chr 22	q13.32	249,017,773-49,105,738	0.09	52	0.00E+00	1	*FAM19A5*	45
						113		

APPENDIX 4

Table B1: Significant Common Focal Amplifications in Leiomyosarcoma (GISTIC)

	Peak Region	Extended Region	Length (Mb)	Freq (%)	G-Score	Q-Bound	No. of Genes	Genes in Peak	All Genes
Chr 1	145,112,508 - 145,278,200	144,009,907 - 145,492,482	1.48	81	6.848	0.035	15	*SEC22B NOTCH2NL*	*SRGAP2B, PPIAL4B, PPIAL4C, PPIAL4A, PFN1P2, NBPF9, PDE4DIP, SEC22B, NOTCH2NL, NBPF10, HFE2, TXNIP, POLR3GL, ANKRD34A, LIX1L*
	230,350,913 - 230,412,939	228,255,947 - 231,657,562	3.40	24	6.493	0.049	36	*GALNT2*	*ARF1, MRPL55, GUK1, GJC2, IBA57, OBSCN, TRIM11, TRIM17, HIST3H3, HIST3H2A, HIST3H2BB, RNF187, DUSP5P1, RHOU, RAB4A, SPHAR, CCSAP, ACTA1, NUP133 ABCB10, TAF5L, URB2 GALNT2, PGBD5, COG2 AGT, CAPN9, TTC13, ARV1, FAM89A, TRIM67,GNPAT, EXOC8, SPRTN, EGLN1, SNRPD2P2*
Chr 3	162,504,616 - 162,630,126	150,855,624 - 165,822,232	14.97	67	7.412	0.022	55	-	*GPR171, P2RY14, MED12L, GPR87, P2RY13, P2RY12, IGSF10, AADACL2, AADAC, SUCNR1, MBNL1-AS1, TMEM14E, MBNL1, P2RY1, RAP2B, ARHGEF26-AS1, ARHGEF26, DHX36, GPR149, MME, PLCH1, SLC33A1, GMPS, KCNAB1-AS2, KCNAB1, KCNAB1-AS1, SSR3, TIPARP-AS1, TIPARP, PA2G4P4, LEKR, CCNL1, VEPH1, PTX3, SHOX2, RSRC1, MLF1, GFM1, LXN, RARRES1, MFSD1, IQCJ, IQCJ-SCHIP1, SCHIP1, IL12A, IFT80, SMC4, TRIM59, SCARNA7, KPNA4, ARL14, PPM1L, B3GALNT1, NMD3, SPTSSB, OTOL1, CT64, SI, SLITRK3, BCHE*
Chr 14	22,588,308 - 22,958,571	22,394,168 - 23,535,702	1.14	76	6.643	0.040	16	-	*DAD1, ABHD4, OXA1L, SLC7A7, MRPL52, MMP14, LRP10, REM2, RBM23, PRMT5, HAUS4, AJUBA, PSMB5, PSMB11, CDH24, ACIN1*
Chr 17	12,159,777 - 12,638,186	11,235,749 - 14,746,194	3.51	43	8.084	0.017	13	*SHISA6, DNAH9, ZNF18, MAP2K4, MYOCD, ARHGAP44, ELAC2, HS3ST3A1, CDRT15P1, COX10-AS1, COX10, CDRT15, MGC12916, HS3ST3B1*	*SHISA6, DNAH9, ZNF18, MAP2K4, MYOCD, ARHGAP44, ELAC2, HS3ST3A1, CDRT15P1, COX10-AS1, COX10, CDRT15, MGC12916, HS3ST3B1*
							135		

APPENDIX 4

Table B2: Significant Common Focal Deletions in Leiomyosarcoma (GISTIC)

	Peak Region	Extended Region	Length (Mb)	Freq (%)	Q-Bound	G-Score	No. of Genes	Genes in Peak	All Genes
Chr 1	3,277,968 - 3,402,895	2,285,830 - 4,637,196	2.35	48	0.002	9.69	23	*PRDM16, ARHGEF16*	*MORN1, RER1, PEX10, PLCH2, PANK4, HES5, TNFRSF14, FAM213B, MMEL1, TTC34, ACTRT2, FLJ42875, PRDM16, ARHGEF16, MEGF6, TPRG1L, WRAP73, TP73, TP73-AS1, CCDC27, SMIM1, LRRC47, CEP104, DFFB*
Chr 2	242,989,305 - 243,199,373	233,732,597 - 243,199,373	9.47	67	0.000	13.59	88	-	*NGEF, NEU2, INPP5D, ATG16L1, SCARNA5, SCARNA6, SAG, DGKD, USP40, UGT1A8, UGT1A10, UGT1A9, UGT1A7, UGT1A6, UGT1A5, DNAJB3, UGT1A4, UGT1A3, UGT1A1, HJURP, MSL3P1, TRPM8, SPP2, ARL4C, SH3BP4, AGAP1, GBX2, ASB18, IQCA1, CXCR7, COPS8, COL6A3, MLPH, PRLH, RAB17, LRRFIP1, RBM44, RAMP1, UBE2F, UBE2F-SCLY, SCLY, ESPNL, KLHL30, ILKAP, HES6, PER2, TRAF3IP1, ASB1, TWIST2, FLJ43879, MGC16025, HDAC4, NDUFA10, OR6B2, PRR21, OR6B3, MYEOV2, OTOS, GPC1, PP14571, ANKMY1, DUSP28, RNPEPL1, CAPN10, GPR35, AQP12B, AQP12A, KIF1A, AGXT, SNED1, MTERFD2, PASK, PPP1R7, ANO7, HDLBP, SEPT2, FARP2, STK25, BOK-AS1, BOK, THAP4, ATG4B, DTYMK, ING5, D2HGDH, GAL3ST2, NEU4, PDCD1, CXXC11*
Chr 10	89,569,335 - 90,067,948	81,212,702 - 99,855,478	18.64	52	0.030	7.97	133	*ATAD1, CFL1P1, KLLN, PTEN, RNLS*	*EIF5AL1, SFTPA2, SFTPA1, MBL1P, SFTPD, TMEM254-AS1, TMEM254, PLAC9, ANXA11, MAT1A, DYDC1, DYDC2, FAM213A, TSPAN14, SH2D4B, NRG3, GHITM, CDHR1, LRIT2, LRIT1, RGR, CCSER2, GRID1-AS1, GRID1, WAPAL, OPN4, LDB3, BMPR1A, MMRN2, SNCG, AGAP11, FAM25A, GLUD1, FAM35A, FAM22A, FAM22D, MINPP1, PAPSS2, ATAD1, CFL1P1, KLLN, PTEN, RNLS, LIPJ, LIPF, LIPK, LIPN, LIPM, ANKRD22, STAMBPL1, ACTA2, FAS-AS1, FAS, CH25H, LIPA, IFIT2, IFIT3, IFIT1B, IFIT1, IFIT5, SLC16A12, PANK1, FLJ37201, KIF20B, HTR7, RPP30, ANKRD1, NUDT9P1, PCGF5, HECTD2, PPP1R3C, TNKS2, FGFBP3, BTAF1, CPEB3, MARCH5, MARK2P9, IDE, KIF11, HHEX, EXOC6, CYP26C1, CYP26A1, MYOF, CEP55, FFAR4, RBP4, PDE6C, FRA10AC1, LGI1, SLC35G1, PIPSL, PLCE1, NOC3L, TBC1D12, HELLS, CYP2C18, CYP2C19, CYP2C9, CYP2C8, PDLIM1, SORBS1, ALDH18A1, TCTN3, ENTPD1, LOC728558, C10orf131, CC2D2B, CCNJ, MIR3157, ZNF518A, BLNK, DNTT, OPALIN, TLL2, TM9SF3, PIK3AP1, LCOR, C10orf12, SLIT1, LOC100505540, ARHGAP19-SLIT1, ARHGAP19, FRAT1, FRAT2, RRP12, PGAM1, EXOSC1, ZDHHC16, MMS19, UBTD1, ANKRD2, C10orf62, MIR5692C2, HOGA1, MORN4, PI4K2A, AVPI1, MARVELD1, ZFYVE27, SFRP5, GOLGA7B, CRTAC1*
Chr 13	49,838,258 - 50,073,617	46,723,507 - 53,957,341	7.23	67	0.000	18.78	53	*CDADC1, CAB39L, SETDB2, PHF11*	*LCP1, LRCH1, ESD, HTR2A, SUCLA2, NUDT15, MED4-AS1, MED4, ITM2B, RB1, LPAR6, RCBTB2, CYSLTR2, FNDC3A, MLNR, CDADC1, CAB39L, SETDB2, PHF11, RCBTB1, ARL11, EBPL, KPNA3, CTAGE10P, SPRYD7, TRIM13, KCNRG, DLEU2, DLEU1, ST13P4, DLEU7, DLEU7-AS1, RNASEH2B-AS1, RNASEH2B, GUCY1B2, FAM124A, SERPINE3, INTS6, WDFY2, DHRS12, CCDC70, ATP7B, ALG11, UTP14C, NEK5, NEK3, MRPS31P5, THSD1, VPS36, CKAP2, TPTE2P3, HNRNPA1L2, SUGT1, LECT1, PCDH8, OLFM4*
Chr 17	7,565,085 - 7,618,079	7,111,674 - 7,650,119	0.05	52	0.000	11.18	47	*TP53, WRAP53, EFNB3*	*DLG4, ACADVL, DVL2, PHF23, GABARAP, CTDNEP1, ELP5, CLDN7, SLC2A4, YBX2, EIF5A, GPS2, NEURL4, ACAP1, KCTD11, TMEM95, TNK1, PLSCR3, PLSCR3, TMEM256, NLGN2, SPEM1, TMEM102, FGF11, CHRNB1, ZBTB4, SLC35G6, POLR2A, TNFSF12, TNFSF12-TNFSF13, TNFSF13, SENP3, SENP3-EIF4A1, SNORA48, EIF4A1, SNORD10, SNORA67, CD68, MPDU1, SOX15, FXR2, SHBG, SAT2, ATP1B2, TP53, WRAP53, EFNB3, DNAH2*
							344		

APPENDIX 4

Table C1: Over-represented (Enriched) Pathways among Commonly-amplified Genes in Leiomyosarcoma

Category	Term	Count	Fisher p-value	EASE p-value	Genes	Fold Enrichment
BIOCARTA h_akap13Pathway	Rho-Selective Guanine Exchange Factor AKAP13 Mediates Stress Fiber Formation	4	2.90E-03	2.65E-02	*PRKAR2A, AKAP13, PRKACA, LPAR1*	5.59
BIOCARTA h_pelp1Pathway	Pelp1 Modulation of Estrogen Receptor Activity	4	2.90E-03	2.65E-02	*MAPK1, CREBBP, ESR1, SRC*	5.59
BIOCARTA h_nthiPathway	NFkB activation by Nontypeable Hemophilus influenzae	10	5.20E-05	3.12E-04	*TNF, MYD88, DUSP1, MAP2K3, CREBBP, TGFBR2, SMAD3, NFKBIA, NR3C1, IKBKB*	4.07
BIOCARTA h_cdMacPathway	Cadmium induces DNA synthesis and proliferation in macrophages	5	6.80E-03	3.53E-02	*MAPK1, FOS, TNF, JUN, NFKBIA*	3.76
BIOCARTA h_etsPathway	METS affect on Macrophage Differentiation	6	3.50E-03	1.75E-02	*FOS, ETS2, JUN, DDX20, NCOR2, ETV3*	3.67
BIOCARTA h_il6Pathway	IL 6 signaling pathway	7	1.80E-03	8.79E-03	*FOS, CEBPB, GRB2, JUN, SHC1, IL6R, SRF*	3.60
BIOCARTA h_ecmPathway	Erk and PI-3 Kinase Are Necessary for Collagen Binding in Corneal Epithelia	8	9.10E-04	4.45E-03	*MAPK1, ARHGAP5, ROCK1, ROCK1P1, DIAPH1, FYN, SHC1, SRC, MYLK*	3.55
BIOCARTA h_ptenPathway	PTEN dependent cell cycle arrest and apoptosis	7	4.60E-03	1.87E-02	*MAPK1, GRB2, BCAR1, PIK3CA, SHC1, FOXO3B, PIK3R1*	3.11
BIOCARTA h_insulinPathway	Insulin Signaling Pathway	6	9.10E-03	3.63E-02	*FOS, GRB2, JUN, SHC1, SRF, INSR*	3.09
BIOCARTA h_pyk2Pathway	Links between Pyk2 and Map Kinases	8	3.10E-03	1.21E-02	*MAPK1, GNAQ, GRB2, JUN, MAP2K3, BCAR1, SHC1, SRC*	3.01
BIOCARTA h_rhoPathway	Rho cell motility signaling pathway	7	7.90E-03	2.85E-02	*ARHGAP5, ROCK1, ROCK1P1, DIAPH1, PIP5K1B, SRC, MYLK, ARHGEF11*	2.85
KEGG hsa05020	Prion diseases	10	2.30E-03	8.02E-03	*PRKACG, NCAM2, MAPK1, C7, FYN, C6, PRKACA, HSPA1A, LAMC1, HSPA5, HSPA1B*	2.74
BIOCARTA h_carm-erPathway	CARM1 and Regulation of the Estrogen Receptor	7	1.00E-02	3.44E-02	*MEF2C, GTF2E1, RLN3, CREBBP, ESR1, NCOR2, NRIP1*	2.74
KEGG hsa04320	Dorso-ventral axis formation	7	1.20E-02	3.94E-02	*NOTCH3, MAPK1, NOTCH2, HSD3B7, GRB2, ETS2, SOS2*	2.68
BIOCARTA h_egfPathway	EGF Signaling Pathway	7	1.30E-02	4.11E-02	*FOS, GRB2, JUN, PIK3CA, SHC1, SRF, PIK3R1*	2.63
BIOCARTA h_integrinPathway	Integrin Signaling Pathway	9	5.40E-03	1.71E-02	*MAPK1, ROCK1, GRB2, ROCK1P1, FYN, JUN, BCAR1, RAP1A, SHC1, SRC*	2.59
KEGG hsa04130	SNARE interactions in vesicular transport	10	4.50E-03	1.40E-02	*STX5, STX4, VAMP8, STX18, VAMP5, USE1, SEC22B, VAMP2, STX1B, STX10*	2.52
KEGG hsa05322	Systemic lupus erythematosus	25	1.80E-05	5.31E-05	*HIST1H2AC, C7, HIST2H2AA3, HIST2H2AA4, TNF, HIST1H4K, HIST1H2AD, C6, HIST1H2AE, HIST2H4A, HIST2H4B, HIST2H2AB, HIST1H2BM, HIST1H2BN, FCGR1C, HIST1H2BL, HIST2H2AC, HIST1H4E, HIST1H4F, HIST1H4D, HIST3H3, HIST1H4J, HIST1H4G, HIST2H3A, HIST1H2BD, HIST1H2BE, HIST1H2BF, HIST1H2BG, HIST1H2BH, GRIN2A, ACTN1, HIST2H3C, HIST2H2BE, HIST2H2BF, HIST1H2AI, HIST1H2AK, HIST1H3D, HIST1H3E, HIST1H2AJ, HIST1H3F, HIST1H3H*	2.42

APPENDIX 4

Table C1: Over-represented (Enriched) Pathways among Commonly-amplified Genes in Leiomyosarcoma

Category	Term	Count	Fisher p-value	EASE p-value	Genes	Fold Enrichment
KEGG hsa05222	Small cell lung cancer	21	9.90E-05	2.87E-04	*E2F2, CKS1B, FHIT, COL4A2, COL4A1, PTGS2, RXRB, NFKBIA, BCL2L1, CCNE1, LAMA4, LAMA3, BCL2, PIK3CA, LAMC2, RARB, LAMC1, IKBKB, MYC, PIK3R1, TRAF4*	2.39
BIOCARTA h_pparaPathway	Mechanism of Gene Regulation by Peroxisome Proliferators via PPARa(alpha)	10	7.60E-03	2.14E-02	*TNF, DUSP1, JUN, CREBBP, NFKBIA, HSPA1A, HSPA1B, NCOR1, NCOR2, NR2F1, NRIP1*	2.33
KEGG hsa04512	ECM-receptor interaction	19	8.20E-04	2.10E-03	*COL4A2, COL4A1, COL3A1, ITGA11, ITGB5, COL5A2, ITGA9, LAMA4, LAMA3, CD44, ITGA5, TNR, COL6A2, COL6A1, LAMC2, LAMC1, THBS2, THBS3, THBS4*	2.17
KEGG hsa05220	Chronic myeloid leukemia	16	3.90E-03	9.34E-03	*E2F2, GRB2, TGFBR2, NFKBIA, SMAD3, BCL2L1, MAPK1, CBLB, SOS2, PIK3CA, SHC1, IKBKB, SHC3, MYC, PIK3R1, SHC4*	2.04
KEGG hsa05223	Non-small cell lung cancer	11	2.20E-02	4.95E-02	*FHIT, E2F2, MAPK1, GRB2, RXRB, RASSF1, SOS2, PIK3CA, RARB, FOXO3B, PIK3R1*	1.95
KEGG hsa04510	Focal adhesion	40	3.40E-05	7.42E-05	*CAV2, DIAPH1, GRB2, BCAR1, COL3A1, ITGA11, ITGB5, SRC, ARHGAP5, TNR, BCL2, SOS2, COL6A2, PIK3CA, COL6A1, SHC1, SHC3, THBS2, THBS3, PIK3R1, SHC4, THBS4, COL4A2, COL4A1, VAV3, ROCK1, MYLPF, ACTN1, COL5A2, MAPK1, ITGA9, LAMA4, LAMA3, FYN, ITGA5, ROCK1P1, JUN, RAP1A, LAMC2, LAMC1, MYLK*	1.91
BIOCARTA h_mapkPathway	MAPKinase Signaling Pathway	15	1.20E-02	2.63E-02	*GRB2, MAP2K3, NFKBIA, MAPKAPK2, DAXX, RPS6KA5, FOS, MAPK1, MAP4K5, MAP3K4, MAPK6, JUN, SHC1, MAPK7, IKBKB*	1.83
KEGG hsa04660	T cell receptor signaling pathway	19	1.50E-02	2.91E-02	*PTPRC, ITK, TNF, VAV3, GRB2, NFKBIA, MAPK1, FOS, CBLB, FYN, JUN, NCK1, SOS2, NFAT5, PIK3CA, NFATC4, NFATC2, IKBKB, PIK3R1*	1.68
KEGG hsa04912	GnRH signaling pathway	17	2.40E-02	4.51E-02	*ADCY4, PLA2G10, GRB2, MAP2K3, ITPR1, SRC, PRKACG, MAPK1, PLA2G4A, MAP3K4, GNAQ, ADCY9, JUN, SOS2, PLA2G6, PRKACA, MAPK7*	1.66
KEGG hsa04530	Tight junction	23	1.10E-02	2.00E-02	*PARD6B, CLDN18, MYH15, MAGI1, MYH1, MYH3, MYH2, CSNK2B, MYH4, MYLPF, PRKCH, ACTN1, MYH9, MYH8, SRC, CGN, ASH1L, MYH11, MYH13, RAB13, PPP2R2C, TJP2, MYH10*	1.64
KEGG hsa04910	Insulin signaling pathway	23	1.20E-02	2.16E-02	*SREBF1, SOCS3, GRB2, PHKG2, PRKAB2, RHOQ, SOCS4, PCK2, PRKACG, MAPK1, PRKAR2A, CBLB, PPP1R3B, SOS2, PIK3CA, PRKACA, SHC1, PRKAA1, SHC3, IKBKB, INSR, PIK3R1, SHC4*	1.63

APPENDIX 4

Table C2: Over-represented (Enriched) Pathways among Commonly-deleted Genes in Leiomyosarcoma

Category	Term	Gene Count	Fisher p-value	EASE p-value	Genes	Fold Enrichment
BIOCARTA h_cremPathway	Regulation of Spermatogenesis by CREM	5	1.20E-04	1.97E-03	*XPO1, ADCY1, CREM, FSHB, FSHR*	7.66
BIOCARTA h_pmlPathway	Regulation of transcriptional activity by PML	5	9.10E-03	4.28E-02	*SUMO1, TP53, RB1, FAS, SIRT1*	3.57
KEGG hsa00040	Pentose and glucuronate interconversions	6	4.50E-03	2.14E-02	*LOC729020, UGT1A1, UGT1A7, UGT1A10, UGT1A6, UGT2B17, UGT1A9, RPE, UGT1A8, UGT1A3, UGT1A5, UGT1A4, UGT2B4, UGT2A1, UGT2B15*	3.56
KEGG hsa00140	Steroid hormone biosynthesis	13	2.00E-04	7.77E-04	*HSD17B2, CYP11B1, CYP11B2, UGT1A1, AKR1C3, UGT1A7, AKR1C2, UGT1A10, UGT1A6, CYP17A1, UGT2B17, UGT1A9, AKR1C4, UGT1A8, UGT1A3, UGT1A5, UGT1A4, UGT2A1, UGT2B4, SRD5A2, UGT2B15*	3.02
KEGG hsa00980	Metabolism of xenobiotics by cytochrome P450	16	8.30E-05	2.93E-04	*CYP2C19, CYP2C18, CYP2C9, ALDH3B2, CYP2E1, UGT1A1, ALDH3B1, AKR1C3, UGT1A7, AKR1C2, UGT1A10, UGT1A6, UGT2B17, UGT1A9, AKR1C4, UGT1A8, UGT1A3, UGT1A5, UGT1A4, UGT2A1, UGT2B4, UGT2B15, MGST1, GSTP1*	2.85
KEGG hsa00150	Androgen and estrogen metabolism	9	5.90E-03	1.86E-02	*HSD17B2, CYP11B1, UGT1A1, UGT1A7, UGT1A10, UGT1A6, UGT2B17, UGT1A9, AKR1C4, UGT1A8, UGT1A3, UGT1A5, UGT1A4, UGT2B4, UGT2A1, SRD5A2, UGT2B15*	2.60
KEGG hsa00982	Drug metabolism	13	4.10E-03	1.09E-02	*CYP2C19, CYP2C18, CYP2C9, ALDH3B2, CYP2E1, UGT1A1, ALDH3B1, UGT1A7, UGT1A10, UGT1A6, UGT2B17, UGT1A9, UGT1A8, UGT1A3, UGT1A5, UGT1A4, UGT2A1, UGT2B4, UGT2B15, MGST1, GSTP1*	2.24
KEGG hsa00830	Retinol metabolism	11	1.00E-02	2.58E-02	*BCMO1, CYP2C19, CYP2C18, CYP2C9, UGT1A1, UGT1A7, UGT1A10, UGT1A6, DHRS3, UGT2B17, UGT1A9, UGT1A8, UGT1A3, UGT1A5, UGT1A4, CYP26B1, UGT2B4, UGT2A1, UGT2B15*	2.18
KEGG hsa04740	Olfactory transduction	66	2.20E-07	4.63E-07	*OR5C1, OR1A1, OR1A2, OR4D11, OR4D10, OR4C6, OR4S2, OR2T10, OR2T11, OR4C11, OR8A1, OR8G1, OR5AN1, OR10S1, OR1K1, OR5M1, OR4F17, OR4M2, OR4D9, OR4M1, OR1D2, OR1D5, OR4D6, OR1S2, OR1D4, OR1S1, OR4D5, OR6M1, OR8B8, OR4N4, OR4N2, OR8B4, OR6Q1, OR8U8, OR1L6, OR1L4, OR5A1, OR5A2, OR9Q1, OR10G4, OR1G1, OR9Q2, OR10G7, OR10G8, OR9I1, OR10G9, PDE1C, CALML3, OR4Q3, CALML5, OR2T34, OR4K5, OR5P3, OR4K2, OR5P2, OR5M10, OR5M11, OR8G2, OR8G5, OR6T1, OR4K1, OR8B12, OR10Q1, OR8D1, OR8D2, OR8D4, OR4P4, OR6X1*	1.86

APPENDIX 4

Table D1: Differential Amplifications in Leiomyosarcoma compared with other STS subtypes

	Region	Cytoband Location	Length (Mb)	Freq. in LMS (%)	Freq. in Other STS (%)	Difference	q-bound	p-value	Count	Genes
Chr 1	203,369,354-203,514,881	q32.1	0.15	52.4	9.4	43.0	0.21	0.0010	2	*PRELP OPTC*
	183,311,689-183,353,653	q25.3	0.04	57.1	15.6	41.5	0.25	0.0025	1	*NMNAT2*
Chr 14	69,333,317-69,519,867	q24.1	0.19	57.1	9.4	47.8	0.18	0.0003	2	*ACTN1 DCAF5*
Chr 16	15,891,532-15,957,512	p13.11	0.07	52.4	6.3	46.1	0.15	0.0002	1	*MYH11*
									6	

Table D2: Differential Deletions in Leiomyosarcoma compared with other STS subtypes

	Region	Cytoband Location	Length (Mb)	Freq. in LMS (%)	Freq. in Other STS (%)	Difference	p-value	q-bound	Count	Genes
Chr 1	1,151,667-1,161,955	p36.33	0.01	57.1	3.1	54.0	0.0000	0.08	1	*SDF4*
	3,184,391-3,448,830	p36.32	0.26	47.6	3.1	44.5	0.0002	0.11	3	*PRDM16*
Chr 10	103,015,529-103,301,856	q24.32	0.29	61.9	21.9	40.0	0.0045	0.19	1	***BTRC***
Chr 11	78,557,943-78,714,348	q14.1	0.16	47.6	3.1	44.5	0.0002	0.11	1	*TENM4*
	92,589,653-92,644,197	q14.3	0.05	57.1	15.6	41.5	0.0025	0.17	1	*FAT3*
									7	

APPENDIX 5

Table A: Significant Focal Common Amplifications in Gastrointestinal Stromal Tumours (STAC)

	Cytoband	Region	Length (Mb)	Frequency (%)	p-value	Genes	Gene Symbols	% of CNV Overlap
Chr 1	q24.2	168,378,978-168,685,455	0.31	27	4.00E-03	3	*DPT, XCL1, XCL2*	3
	q32.2	207,984,254-208,137,993	0.15	45	0.00E+00	1	*CD34*	3
Chr 2	q12.1 - q12.2	105,949,598-106,147,840	0.20	36	0.00E+00	1	*FXYD6, FXYD6-FXYD2, TMPRSS13*	0
Chr 5	p15.33	1,035,059-1,353,444	0.32	45	0.00E+00	6	*ANO1, FADD*	6
	q31.3	141,696,250-142,040,249	0.34	45	0.00E+00	2	*APOLD1, CDKN1B*	0
Chr 6	p25.2	2,836,816-3,001,306	0.16	36	2.00E-02	6	*NCOR2, SCARB1*	20
	p22.2	26,124,423-26,250,955	0.13	64	0.00E+00	16	*ALOXE3, AURKB, HES7, PER1, TMEM107, VAMP2*	66
	p21.31	34,958,314-35,188,392	0.23	36	2.00E-02	3	*ELL, GDF15, ISYNA1, JUND, LRRC25, LSM4, PGPEP1, SSBP4*	12
Chr 7	q36.1	150,913,717-150,954,753	0.04	45	0.00E+00	3	*FHL2*	0
Chr 9	q22.32	97,847,752-98,088,612	0.24	36	7.00E-03	1	*ARFGAP1, BIRC7, CHRNA4, COL20A1, EEF1A2, FLJ16779, KCNQ2, NKAIN4, YTHDF1*	100
Chr 11	q23.3	117,706,762-117,781,447	0.07	36	0.00E+00	3	*CLPTM1L, NKD2, SLC12A7, SLC6A18, SLC6A19, TERT*	68
	q13.3	69,906,662-70,052,101	0.15	36	0.00E+00	2	*FGF1, SPRY4*	4
Chr 12	p13.1	12,863,830-12,896,612	0.03	27	0.00E+00	2	*DKFZP686I15217, MGC39372, NQO2, SERPINB1, SERPINB6, SERPINB9*	7
	q24.31	124,988,479-125,313,180	0.32	27	1.00E-03	2	*HIST1H1D, HIST1H1E, HIST1H2AC, HIST1H2AD, HIST1H2AE, HIST1H2BD, HIST1H2BE, HIST1H2BF, HIST1H2BG, HIST1H3D, HIST1H3E, HIST1H3F, HIST1H4D, HIST1H4E, HIST1H4F, HIST1H4G*	50
Chr 17	p13.1	8,017,959-8,127,962	0.11	27	2.00E-03	6	*ANKS1A, SCUBE3, TCP11*	12
Chr 19	p13.11	18,380,369-18,583,639	0.20	45	0.00E+00	9	*ABCF2, CHPF2, SMARCD3*	17
Chr 20	q13.33	61,818,375-62,138,166	0.32	27	1.00E-02	9	*FANCC,*	22
						75		

APPENDIX 5

Table B: Significant Focal Common Deletions in Gastrointestinal Stromal Tumours (GISTIC)

	Peak Region	Extended Region	Length (Mb)	Q-Bound	G-Score	No. of Genes	Genes in Peak	Genes
Chr 1	72,765,091-72,812,744	0-121,322,377	121.30	0.00	4.91	1266	-	*TNFRSF14, PRDM16, RPL22, CAMTA1, SDHB, PAX7, MDS2, ARID1A, LCK, SFPQ, THRAP3, MYCL1, MPL, MUTYH, TAL1, STIL, CDKN2C, EPS15, JUN, JAK1, FUBP1, BCL10, BCAR3, RBM15, TRIM33, NRAS, FAM46C, NOTCH2, OR4F5, SAMD11, NOC2L, KLHL17, PLEKHN1, C1orf170, HES4, ISG15, AGRN, TTLL10, TNFRSF18, TNFRSF4, SDF4, B3GALT6, FAM132A, UBE2J2, SCNN1D, ACAP3, PUSL1, CPSF3L, TAS1R3, DVL1, MXRA8, AURKAIP1, CCNL2, MRPL20, VWA1, ATAD3C, ATAD3B, ATAD3A, SSU72, MIB2, MMP23B, CDK11B, CDK11A, SLC35E2, NADK, GNB1, CALML6, TMEM52, KIAA1751, GABRD, PRKCZ, C1orf86, SKI, MORN1, RER1, PEX10, PLCH2, PANK4, HES5, MMEL1, ACTRT2, ARHGEF16, MEGF6, TPRG1L, WRAP73, TP73, CCDC27, LRRC47, CEP104, DFFB, C1orf174, AJAP1, NPHP4, KCNAB2, CHD5, RNF207, ICMT, HES3, GPR153, ACOT7, HES2, ESPN, TNFRSF25, PLEKHG5, NOL9, TAS1R1, ZBTB48, KLHL21, PHF13, THAP3, DNAJC11, VAMP3, PER3, UTS2, TNFRSF9, PARK7, ERRFI1, SLC45A1, RERE, ENO1, CA6, SLC2A7, SLC2A5, GPR157, H6PD, SPSB1, SLC25A33, TMEM201, C1orf200, PIK3CD, CLSTN1, CTNNBIP1, LZIC, NMNAT1, RBP7, UBE4B, KIF1B, PGD, APITD1, CORT, DFFA, PEX14, CASZ1, C1orf127, TARDBP, MASP2, SRM, EXOSC10, MTOR, ANGPTL7, UBIAD1, PTCHD2, FBXO2, FBXO44, FBXO6, MAD2L2, AGTRAP, MTHFR, CLCN6, NPPA, NPPB, KIAA2013, PLOD1, MFN2, MIIP, TNFRSF8, TNFRSF1B, VPS13D, DHRS3, AADACL4, AADACL3, C1orf158, PRAMEF12, PRAMEF1, PRAMEF11, HNRNPCL1, PRAMEF2, PRAMEF4, PRAMEF10, PRAMEF7, PRAMEF22, PRAMEF5, PRAMEF3, PRAMEF22, PRAMEF5, PRAMEF13, PRAMEF14, PRAMEF18, PRAMEF20, PRAMEF7, PRAMEF14, PRAMEF18, PRAMEF20, PDPN, PRDM2, KAZN, TMEM51, FHAD1, EFHD2, CTRC, CELA2A, CELA2B, CASP9, DNAJC16, AGMAT, DDI2, RSC1A1, PLEKHM2, SLC25A34, TMEM82, FBLIM1, SPEN, ZBTB17, C1orf64, HSPB7, CLCNKA, CLCNKB, FAM131C, EPHA2, ARHGEF19, RSG1, FBXO42, SPATA21, NECAP2, CROCC, MFAP2, ATP13A2, PADI2, PADI1, PADI3, PADI4, PADI6, RCC2, ARHGEF10L, ACTL8, IGSF21, KLHDC7A, TAS1R2, ALDH4A1, UBR4, MRTO4, AKR7L, AKR7A3, AKR7A2, PQLC2, CAPZB, NBL1, HTR6, TMCO4, RNF186, OTUD3, PLA2G2E, PLA2G2A, PLA2G5, PLA2G2D, PLA2G2F, PLA2G2C, UBXN10, VWA5B1, CAMK2N1, MUL1, FAM43B, CDA, PINK1, DDOST, KIF17, SH2D5, HP1BP3, EIF4G3, ECE1, NBPF3, ALPL, RAP1GAP, USP48, LDLRAD2, HSPG2, CELA3B, CELA3A, CDC42, WNT4, ZBTB40, EPHA8, C1QA, C1QC, C1QB, EPHB2, KDM1A, LUZP1, HTR1D, HNRNPR, ZNF436, C1orf213, ASAP3, E2F2, ID3, RPL11, TCEB3, PITHD1, LYPLA2, GALE, HMGCL, FUCA1, CNR2, PNRC2, MYOM3, IL22RA1, GRHL3, NIPAL3, RCAN3, SRRM1, CLIC4, RUNX3, SYF2, C1orf63, RHD, TMEM50A, RHCE, TMEM57, LDLRAP1, MAN1C1, SEPN1, PAQR7, STMN1, PAFAH2, EXTL1, SLC30A2, TRIM63, PDIK1L, ZNF593, CNKSR1, CATSPER4, CEP85, SH3BGRL3, UBXN11, CD52, AIM1L, ZNF683, LIN28A, DHDDS, HMGN2, RPS6KA1, PIGV, ZDHHC18, SFN, GPN2, GPATCH3, NR0B2, NUDC, C1orf172, TRNP1, FAM46B, SLC9A1, WDTC1, TMEM222, SYTL1, MAP3K6, FCN3, CD164L2, GPR3, WASF2, AHDC1, FGR, IFI6, FAM76A, STX12, PPP1R8, RPA2, SMPDL3B, XKR8, EYA3, PTAFR, DNAJC8, ATPIF1, SESN2, MED18, PHACTR4, RCC1, TRNAU1AP, RAB42, TAF12, GMEB1, YTHDF2, OPRD1, EPB41, TMEM200B, SRSF4, MECR, PTPRU, MATN1, LAPTM5, SDC3, PUM1, NKAIN1, SNRNP40, ZCCHC17, FABP3, SERINC2, TINAGL1, HCRTR1, PEF1, COL16A1, BAI2, SPOCD1, PTP4A2, KHDRBS1, TMEM39B, KPNA6, TXLNA, CCDC28B, IQCC, DCDC2B, TMEM234, EIF3I, FAM167B, HDAC1, MARCKSL1, TSSK3, BSDC1, ZBTB8A, ZBTB8OS, RBBP4, SYNC, KIAA1522, YARS, S100PBP, FNDC5, HPCA, TMEM54, RNF19B, AK2, ADC, TRIM62, ZNF362, PHC2, ZSCAN20, CSMD2, HMGB4, C1orf94, GJB5, GJB4, GJB3, GJA4, DLGAP3, ZMYM6, ZMYM1, ZMYM4, KIAA0319L, NCDN, TFAP2E, PSMB2, C1orf216, CLSPN, EIF2C4, EIF2C1, EIF2C3, TEKT2, ADPRHL2, COL8A2, TRAPPC3, MAP7D1, SH3D21, STK40, LSM10, OSCP1, MRPS15, CSF3R, GRIK3, ZC3H12A, MEAF6, SNIP1, DNALI1, GNL2, RSPO1, C1orf109, CDCA8, EPHA10, MANEAL, YRDC, C1orf122, MTF1, INPP5B, SF3A3, FHL3, UTP11L, POU3F1, RRAGC, MYCBP, GJA9, RHBDL2, AKIRIN1, NDUFS5, MACF1, KIAA0754, BMP8A, PABPC4, HEYL, NT5C1A, HPCAL4, PPIE, OXCT2, BMP8B, TRIT1, MFSD2A, CAP1, PPT1, RLF, TMCO2, ZMPSTE24, COL9A2, SMAP2, ZNF684, RIMS3, NFYC, KCNQ4, CITED4, SLFNL1, SCMH1, EDN2, HIVEP3, GUCA2B, GUCA2A, FOXJ3, RIMKLA, ZMYND12, PPCS, CCDC30, PPIH, YBX1, CLDN19, LEPRE1, C1orf50, CCDC23, ERMAP, ZNF691, SLC2A1, FAM183A, EBNA1BP2, WDR65, TMEM125, C1orf210, TIE1, CDC20, ELOVL1, MED8, SZT2, HYI, PTPRF, KDM4A, ST3GAL3, ARTN, IPO13, DPH2, ATP6V0B, B4GALT2, CCDC24, SLC6A9, KLF17, DMAP1, ERI3, RNF220, TMEM53, KIF2C, RPS8, BEST4, PLK3, TCTEX1D4, PTCH2, EIF2B3, HECTD3, UROD, ZSWIM5, HPDL, TOE1, TESK2, MMACHC, PRDX1, AKR1A1, NASP, CCDC17, GPBP1L1, TMEM69, IPP, MAST2, PIK3R3, TSPAN1, POMGNT1, RAD54L, LRRC41, UQCRH, NSUN4, FAAH, DMBX1, KNCN, MKNK1, ATPAF1, CYP4B1, CYP4A11, CYP4X1, CYP4Z1, CYP4A22, PDZK1IP1, CMPK1, FOXE3, FOXD2, SLC5A9, SPATA6, BEND5, AGBL4, ELAVL4, DMRTA2, FAF1, RNF11, TTC39A, OSBPL9, NRD1, RAB3B, KTI12, TXNDC12, BTF3L4, ZFYVE9, CC2D1B, ORC1, PRPF38A, ZCCHC11, GPX7, FAM159A, SELRC1, ZYG11B, ZYG11A, ECHDC2, SCP2, PODN, SLC1A7, CPT2, C1orf123, MAGOH, LRP8, DMRTB1, GLIS1, TMEM48, YIPF1, DIO1, HSPB11, LRRC42, LDLRAD1, TMEM59, TCEANC2, CDCP2, MRPL37, SSBP3, ACOT11, FAM151A, TTC4, PARS2, TTC22, C1orf177, DHCR24, TMEM61, BSND, PCSK9, USP24, PPAP2B, PRKAA2, C1orf168, C8A, C8B, DAB1, OMA1,*

APPENDIX 5

Table B: Significant Focal Common Deletions in Gastrointestinal Stromal Tumours (GISTIC)

	Peak Region	Extended Region	Length (Mb)	Q-Bound	G-Score	No. of Genes	Genes in Peak	Genes
Chr 1	72,765,091-72,812,744	0-121,322,377	121.30	0.00	4.91	1266	-	*DNAJC6, LEPROT, LEPR, PDE4B, SGIP1, TCTEX1D1, INSL5, WDR78, MIER1, SLC35D1, C1orf141, IL23R, IL12RB2, SERBP1, GADD45A, GNG12, DIRAS3, WLS, RPE65, DEPDC1, LRRC7, LRRC40, SRSF11, ANKRD13C, HHLA3, CTH, PTGER3, ZRANB2, NEGR1, LRRIQ3, FPGT, TNNI3K, C1orf173, CRYZ, TYW3, LHX8, SLC44A5, ACADM, RABGGTB, MSH4, ASB17, ST6GALNAC3, ST6GALNAC5, PIGK, AK5, ZZZ3, USP33, FAM73A, NEXN, DNAJB4, GIPC2, PTGFR, IFI44L, IFI44, ELTD1, LPHN2, TTLL7, PRKACB, SAMD13, DNASE2B, RPF1, GNG5, CTBS, SSX2IP, LPAR3, MCOLN2, MCOLN3, WDR63, SYDE2, C1orf52, DDAH1, CYR61, ZNHIT6, COL24A1, ODF2L, CLCA2, CLCA1, CLCA4, CLCA3P, SH3GLB1, SEP15, HS2ST1, LMO4, PKN2, GTF2B, CCBL2, GBP3, GBP1, GBP2, GBP7, GBP4, GBP5, GBP6, LRRC8B, LRRC8C, LRRC8D, ZNF326, BARHL2, ZNF644, HFM1, CDC7, TGFBR3, BRDT, EPHX4, BTBD8, KIAA1107, C1orf146, GLMN, RPAP2, GFI1, EVI5, RPL5, FAM69A, MTF2, TMED5, CCDC18, DR1, FNBP1L, DNTTIP2, GCLM, ABCA4, ARHGAP29, ABCD3, F3, SLC44A3, CNN3, ALG14, TMEM56, RWDD3, PTBP2, DPYD, SNX7, LPPR4, PALMD, FRRS1, AGL, SLC35A3, HIAT1, SASS6, LRRC39, DBT, CDC14A, GPR88, VCAM1, EXTL2, SLC30A7, DPH5, S1PR1, DNAJA1P5, OLFM3, COL11A1, RNPC3, AMY2B, AMY2A, AMY1C, AMY1C, AMY1C, PRMT6, NTNG1, VAV3, SLC25A24, NBPF4, FAM102B, HENMT1, PRPF38B, FNDC7, STXBP3, AKNAD1, GPSM2, CLCC1, WDR47, TAF13, TMEM167B, C1orf194, KIAA1324, SARS, CELSR2, PSRC1, MYBPHL, SORT1, PSMA5, SYPL2, ATXN7L2, CYB561D1, AMIGO1, GPR61, GNAI3, GNAT2, AMPD2, GSTM4, GSTM2, GSTM1, GSTM5, GSTM3, EPS8L3, CSF1, AHCYL1, ALX3, UBL4B, SLC6A17, KCNC4, SLC16A4, PROK1, KCNA10, KCNA2, KCNA3, CD53, LRIF1, DRAM2, CEPT1, DENND2D, CHI3L2, CHIA, OVGP1, WDR77, ATP5F1, C1orf162, ADORA3, RAP1A, DDX20, KCND3, CTTNBP2NL, WNT2B, ST7L, CAPZA1, MOV10, RHOC, PPM1J, FAM19A3, SLC16A1, LRIG2, MAGI3, PHTF1, RSBN1, PTPN22, BCL2L15, AP4B1, DCLRE1B, HIPK1, OLFML3, SYT6, BCAS2, DENND2C, AMPD1, CSDE1, SIKE1, SYCP1, TSHB, TSPAN2, NGF, VANGL1, CASQ2, NHLH2, SLC22A15, MAB21L3, ATP1A1, CD58, IGSF3, CD2, PTGFRN, CD101, TTF2, TRIM45, VTCN1, MAN1A2, GDAP2, WDR3, SPAG17, TBX15, WARS2, HAO2, HSD3B2, HSD3B1, ZNF697, PHGDH, HMGCS2, REG4, NBPF7, ADAM30, FAM72B, FCGR1B, DDX11L1, WASH7P, FAM138A, FAM138F, LOC729737, LOC100132062, LOC100132287, LOC100133331, OR4F16, OR4F29, OR4F3, OR4F16, OR4F29, OR4F3, LOC100133331, LOC100288069, LINC00115, LOC643837, FAM41C, LOC100130417, RNF223, C1orf159, LOC254099, MIR200B, MIR200A, MIR429, GLTPD1, LOC148413, ANKRD65, TMEM88B, TMEM240, C1orf233, MMP23A, SLC35E2B, MMP23A, LOC100129534, LOC115110, LOC100133445, FAM213B, TTC34, FLJ42875, MIR4251, MIR551A, TP73-AS1, SMIM1, LOC100133612, LOC728716, LOC284661, MIR4417, MIR4689, MIR4252, ENO1-AS1, MIR34A, APITD1-CORT, MTOR-AS1, DRAXIN, NPPA-AS1, MIR4632, SNORA59A, SNORA59B, LOC649330, PRAMEF8, PRAMEF6, LOC440563, PRAMEF6, PRAMEF8, PRAMEF15, PRAMEF9, PRAMEF19, PRAMEF16, PRAMEF21, PRAMEF8, PRAMEF15, PRAMEF9, PRAMEF19, PRAMEF17, PRAMEF21, LRRC38, C1orf126, UQCRHL, FLJ37453, SZRD1, CROCCP3, MIR3675, NBPF1, CROCCP2, MST1P2, MIR3675, ESPNP, MST1L, MIR3675, MIR4695, IFFO2, EMC1, LOC100506730, MINOS1, C1orf151-NBL1, LOC339505, LOC100506801, LINC00339, MIR4684, MIR4253, LOC729059, MIR3115, TCEA3, LOC100506963, MIR378F, SRSF10, IFNLR1, LOC284632, STPG1, RCAN3AS, NCMAP, LOC646471, MTFR1L, AUNIP, MIR3917, FAM110D, MIR1976, LOC644961, SCARNA1, THEMIS2, SNHG3, SNORD99, SNORA61, SNHG12, SNORA44, SNORA16A, RNU11, MATN1-AS1, MIR4420, SNORD103A, SNORD103B, SNORD103A, SNORD103B, SNORD85, LOC149086, LOC284551, MIR4254, MTMR9LP, FAM229A, ZBTB8B, MIR3605, LOC402779, SMIM12, LOC653160, ZMYM6NB, EVA1B, MIR4255, LOC728431, MIR5581, MIR3659, LOC339442, GJA9-MYCBP, PPIEL, SNORA55, ZFP69B, ZFP69, EXOS, LOC100130557, MIR30E, MIR30C1, CTPS1, SLFNL1-AS1, LOC100129924, SLC2A1-AS1, KDM4A-AS1, MIR5584, C1orf228, SNORD55, SNORD46, SNORD38A, SNORD38B, BTBD19, LOC400752, CCDC163P, RPS15AP10, LURAP1, LOC729041, MKNK1-AS1, MOB3C, TEX38, EFCAB14-AS1, EFCAB14, CYP4Z2P, PDZK1IP1-AS1, FOXD2-AS1, TRABD2B, SKINTL, C1orf185, MIR761, LOC100507564, SLC25A3P1, MIR4781, CYB5RL, MROH7, HEATR8-TTC4, LOC100507634, MIR4422, LOC100131060, HSD52, MIR4711, MGC34796, LINC00466, MIR4794, MIR3671, MIR101-1, MIR3117, GNG12-AS1, MIR1262, PIN1P1, ZRANB2-AS1, MIR186, ZRANB2-AS2, NEGR1-IT1, FPGT-TNNI3K, SNORD45C, SNORD45A, SNORD45B, MGC27382, MIR548AP, UOX, C1orf180, MIR4423, LOC646626, LOC339524, LOC100505768, RBMXL1, GBP1P1, FLJ27354, GEMIN8P4, HSP90B3P, SNORD21, SNORA66, LOC100131564, LOC100129046, MIR760, LOC729970, TMEM56-RWDD3, FLJ31662, DPYD-AS1, MIR137HG, MIR2682, MIR137, LOC729987, LPPR5, LOC100129620, MIR548AA1, MIR548D1, TRMT13, RTCA, MIR553, LOC100128787, RNU6-6, ACTG1P4, AMY1A, AMY1B, AMY1B, AMY1A, AMY1A, AMY1B, LOC100129138, VAV3-AS1, NBPF6, SRG7, SCARNA2, MIR197, STRIP1, LOC440600, LAMTOR5, CYMP, CHIAP2, PIFO, PGCP1, LOC100129269, FAM212B, LOC100506343, LOC643355, MIR4256, AKR7A2P1, LOC643441, AP4B1-AS1, ATP1A1OS, MIR320B1, MIR942, HSD3BP4, LINC00622, HIST2H2BA, SRGAP2D, EMBP1, MST1P9, TACSTD2, MYSM1, FGGY, HOOK1, CYP2J2, C1orf87, NFIA, TM2D1, INADL, L1TD1, KANK4, USP1, DOCK7, ANGPTL3, ATG4C, FOXD3, ALG6, ITGB3BP, EFCAB7, DLEU2L, PGM1, ROR1, UBE2U, CACHD1, RAVER2, AK4,*

APPENDIX 5

Table B: Significant Focal Common Deletions in Gastrointestinal Stromal Tumours (GISTIC)

	Peak Region	Extended Region	Length (Mb)	Q-Bound	G-Score	No. of Genes	Genes in Peak	Genes
Chr 14	106,883,633-106,919,036	43,345,398-107,349,540	64.00	0.00	7.04	562	-	*FSCB, C14orf28, KLHL28, FAM179B, PRPF39, SNORD127, FKBP3, FANCM, MIS18BP1, RPL10L, MDGA2, MIR548Y, LINC00648, RPS29, LRR1, RPL36AL, MGAT2, DNAAF2, POLE2, KLHDC1, KLHDC2, NEMF, ARF6, C14orf182, C14orf183, METTL21D, SOS2, L2HGDH, ATP5S, CDKL1, MAP4K5, ATL1, SAV1, NIN, ABHD12B, PYGL, TRIM9, TMX1, LINC00640, FRMD6-AS2, FRMD6, FRMD6-AS1, GNG2, C14orf166, NID2, PTGDR, PTGER2, TXNDC16, GPR137C, ERO1L, PSMC6, STYX, GNPNAT1, FERMT2, DDHD1, MIR5580, BMP4, CDKN3, CNIH, GMFB, CGRRF1, SAMD4A, GCH1, MIR4308, WDHD1, SOCS4, MAPK1IP1L, LGALS3, DLGAP5, FBXO34, ATG14, TBPL2, KTN1-AS1, KTN1, RPL13AP3, LINC00520, PELI2, C14orf101, OTX2, OTX2-AS1, EXOC5, AP5M1, NAA30, C14orf105, SLC35F4, C14orf37, ACTR10, PSMA3, FLJ31306, ARID4A, TOMM20L, TIMM9, KIAA0586, DACT1, DAAM1, GPR135, L3HYPDH, JKAMP, CCDC175, RTN1, PCNXL4, DHRS7, PPM1A, C14orf39, SIX6, SIX1, SIX4, MNAT1, TRMT5, SLC38A6, TMEM30B, PRKCH, FLJ22447, HIF1A, HIF1A-AS2, SNAPC1, SYT16, LINC00643, KCNH5, RHOJ, GPHB5, PPP2R5E, WDR89, SGPP1, SYNE2, ESR2, TEX21P, MTHFD1, AKAP5, ZBTB25, ZBTB1, HSPA2, PPP1R36, PLEKHG3, SPTB, CHURC1, GPX2, RAB15, CHURC1-FNTB, FNTB, MIR4706, MAX, LOC100506321, MIR4708, FUT8-AS1, FUT8, LINC00238, GPHN, FAM71D, MPP5, ATP6V1D, EIF2S1, PLEK2, TMEM229B, PLEKHH1, PIGH, ARG2, VTI1B, RDH11, RDH12, ZFYVE26, RAD51B, ZFP36L1, ACTN1, ACTN1-AS1, DCAF5, EXD2, GALNTL1, ERH, SLC39A9, PLEKHD1, CCDC177, KIAA0247, LOC100289511, SRSF5, SLC10A1, SMOC1, SLC8A3, ADAM21P1, COX16, SYNJ2BP-COX16, SYNJ2BP, ADAM21, ADAM20P1, ADAM20, MED6, TTC9, MAP3K9, PCNX, SNORD56B, LOC145474, SIPA1L1, RGS6, DPF3, DCAF4, ZFYVE1, RBM25, PSEN1, PAPLN, NUMB, C14orf169, HEATR4, ACOT1, ACOT2, ACOT4, ACOT6, DNAL1, PNMA1, ELMSAN1, MIR4505, PTGR2, ZNF410, FAM161B, COQ6, ENTPD5, CCDC176, ALDH6A1, LIN52, VSX2, ABCD4, VRTN, SYNDIG1L, MIR4709, NPC2, ISCA2, LTBP2, KIAA0317, FCF1, YLPM1, PROX2, DLST, RPS6KL1, PGF, EIF2B2, MLH3, ACYP1, ZC2HC1C, NEK9, TMED10, FOS, LOC731223, JDP2, BATF, FLVCR2, C14orf1, TTLL5, TGFB3, IFT43, GPATCH2L, RNU6-6, RNU6-19, ESRRB, VASH1, ANGEL1, C14orf166B, IRF2BPL, KIAA1737, ZDHHC22, TMEM63C, MIR1260A, NGB, POMT2, GSTZ1, TMED8, SAMD15, NOXRED1, VIPAS39, AHSA1, ISM2, SPTLC2, ALKBH1, SLIRP, SNW1, C14orf178, ADCK1, NRXN3, DIO2, DIO2-AS1, CEP128, TSHR, GTF2A1, SNORA79, STON2, SEL1L, FLRT2, LOC283585, GALC, GPR65, LOC283587, KCNK10, SPATA7, PTPN21, ZC3H14, EML5, TTC8, FOXN3, FOXN3-AS1, FOXN3-AS2, EFCAB11, TDP1, KCNK13, PSMC1, NRDE2, CALM1, LINC00642, TTC7B, RPS6KA5, SNORA11B, C14orf159, GPR68, CCDC88C, SMEK1, CATSPERB, TC2N, FBLN5, TRIP11, ATXN3, NDUFB1, CPSF2, SLC24A4, RIN3, LGMN, GOLGA5, CHGA, ITPK1, ITPK1-AS1, MOAP1, TMEM251, C14orf142, UBR7, BTBD7, COX8C, UNC79, PRIMA1, FAM181A-AS1, FAM181A, ASB2, LINC00521, OTUB2, DDX24, IFI27L1, IFI27, IFI27L2, PPP4R4, SERPINA10, SERPINA6, SERPINA1, SERPINA11, SERPINA9, SERPINA12, SERPINA4, SERPINA5, SERPINA3, SERPINA13P, GSC, DICER1, MIR3173, DICER1-AS1, CLMN, LINC00341, SYNE3, SCARNA13, SNHG10, GLRX5, TCL6, TCL1B, TCL1A, LINC00617, C14orf132, BDKRB2, BDKRB1, ATG2B, GSKIP, AK7, PAPOLA, VRK1, LOC100129345, C14orf64, C14orf177, BCL11B, SETD3, CCNK, CCDC85C, HHIPL1, MIR5698, CYP46A1, EML1, EVL, DEGS2, YY1, SLC25A29, MIR345, SLC25A47, WARS, WDR25, BEGAIN, LINC00523, DLK1, MIR2392, MEG3, MIR770, MIR493, MIR337, MIR665, MIR431, MIR433, RTL1, MIR127, MIR432, MIR136, MEG8, SNORD113-1, SNORD113-2, SNORD113-4, SNORD113-5, SNORD113-6, SNORD113-7, SNORD113-9, SNORD114-1, SNORD114-2, SNORD114-3, SNORD114-4, SNORD114-5, SNORD114-6, SNORD114-7, SNORD114-8, SNORD114-9, SNORD114-10, SNORD114-11, SNORD114-12, SNORD114-13, SNORD114-14, SNORD114-15, SNORD114-16, SNORD114-17, SNORD114-18, SNORD114-19, SNORD114-20, SNORD114-21, SNORD114-22, SNORD114-23, SNORD114-24, SNORD114-25, SNORD114-26, SNORD114-27, SNORD114-28, SNORD114-29, SNORD114-30, SNORD114-31, MIR379, MIR411, MIR299, MIR380, MIR1197, MIR323A, MIR758, MIR329-1, MIR329-2, MIR494, MIR1193, MIR543, MIR495, MIR376C, MIR376A2, MIR654, MIR376B, MIR376A1, MIR300, MIR1185-1, MIR1185-2, MIR381, MIR487B, MIR539, MIR889, MIR655, MIR487A, MIR382, MIR134, MIR668, MIR485, MIR323B, MIR154, MIR496, MIR377, MIR541, MIR409, MIR412, MIR369, MIR410, MIR656, MEG9, DIO3OS, MIR1247, DIO3, LINC00239, PPP2R5C, DYNC1H1, HSP90AA1, WDR20, MOK, ZNF839, CINP, TECPR2, ANKRD9, MIR4309, RCOR1, TRAF3, AMN, CDC42BPB, EXOC3L4, TNFAIP2, LINC00605, SNORA28, EIF5, MARK3, CKB, TRMT61A, BAG5, APOPT1, KLC1, XRCC3, ZFYVE21, PPP1R13B, LINC00637, C14orf2, RD3L, TDRD9, ASPG, MIR203, MIR3545, KIF26A, C14orf180, TMEM179, MIR4710, INF2, ADSSL1, SIVA1, AKT1, ZBTB42, LINC00638, CEP170B, PLD4, AHNAK2, C14orf79, CDCA4, GPR132, JAG2, NUDT14, BTBD6, BRF1, PACS2, TEX22, MTA1, CRIP2, CRIP1, C14orf80, TMEM121, ELK2AP, KIAA0125, ADAM6, LINC00226, LINC00221*

APPENDIX 5

Table B: Significant Focal Common Deletions in Gastrointestinal Stromal Tumours (GISTIC)

	Peak Region	Extended Region	Length (Mb)	Q-Bound	G-Score	No. of Genes	Genes in Peak	Genes
Chr 15	102,156,415-102,237,854	72,998,432-102,531,392	29.50	0.05	3.15	297	*TM2D3 TARSL2*	*BBS4, ADPGK, ADPGK-AS1, NEO1, HCN4, C15orf60, NPTN, CD276, C15orf59, TBC1D21, LOXL1-AS1, LOXL1, STOML1, PML, GOLGA6A, LOC283731, ISLR2, ISLR, STRA6, CCDC33, CYP11A1, LOC729739, SEMA7A, UBL7, LOC440288, ARID3B, CLK3, EDC3, CYP1A1, CYP1A2, MIR4513, CSK, LMAN1L, CPLX3, ULK3, SCAMP2, MPI, FAM219B, COX5A, RPP25, SCAMP5, PPCDC, C15orf39, GOLGA6C, GOLGA6D, COMMD4, NEIL1, MIR631, MAN2C1, SIN3A, PTPN9, SNUPN, IMP3, SNX33, CSPG4, ODF3L1, DNM1P35, MIR4313, UBE2Q2, FBXO22, FBXO22-AS1, NRG4, C15orf27, TYRO3P, ETFA, ISL2, SCAPER, RCN2, PSTPIP1, TSPAN3, LINC00597, PEAK1, HMG20A, LINGO1, LOC253044, LOC645752, LOC91450, TBC1D2B, MIR5003, SH2D7, CIB2, IDH3A, ACSBG1, DNAJA4, WDR61, CRABP1, IREB2, AGPHD1, PSMA4, CHRNA5, CHRNA3, CHRNB4, LOC646938, ADAMTS7, MORF4L1, CTSH, RASGRF1, MIR184, LOC729911, ANKRD34C, TMED3, KIAA1024, MTHFS, ST20-MTHFS, ST20, C15orf37, BCL2A1, ZFAND6, FAH, LOC283688, ARNT2, FAM108C1, MIR549, KIAA1199, MESDC2, MIR4514, MESDC1, C15orf26, IL16, STARD5, TMC3, MEX3B, EFTUD1, FAM154B, LOC390660, GOLGA6L10, UBE2Q2P2, UBE2Q2P3, GOLGA6L9, LOC440297, LOC727849, AGSK1, RPS17L, RPS17, LOC390660, AGSK1, LOC727849, LOC440297, GOLGA6L10, UBE2Q2P2, UBE2Q2P3, GOLGA6L9, LOC440297, LOC727849, AGSK1, RPS17L, RPS17, CPEB1, LOC283692, AP3B2, LOC338963, LOC283693, SCARNA15, FSD2, WHAMM, HOMER2, FAM103A1, C15orf40, BTBD1, MIR4515, TM6SF1, HDGFRP3, BNC1, SH3GL3, ADAMTSL3, EFTUD1P1, LOC100505679, LOC642423, LOC440300, LOC388152, GOLGA6L4, DNM1P41, GOLGA6L5, UBE2Q2P1, LOC100506874, ZSCAN2, SCAND2, WDR73, NMB, SEC11A, ZNF592, ALPK3, SLC28A1, PDE8A, LOC642423, AKAP13, MIR1276, KLHL25, LOC727915, AGBL1, LINC00052, NTRK3, NTRK3-AS1, MRPL46, MRPS11, DET1, MIR1179, MIR7-2, MIR3529, AEN, ISG20, ACAN, HAPLN3, MFGE8, ABHD2, RLBP1, FANCI, POLG, MIR9-3, LOC254559, RHCG, LOC283761, TICRR, KIF7, PLIN1, PEX11A, WDR93, MESP1, MESP2, ANPEP, MIR5094, AP3S2, C15orf38-AP3S2, C15orf38, ZNF710, IDH2, SEMA4B, CIB1, GDPGP1, TTLL13, NGRN, GABARAPL3, ZNF774, IQGAP1, CRTC3, BLM, FURIN, FES, MAN2A2, HDDC3, UNC45A, RCCD1, LOC100507118, PRC1, VPS33B, SV2B, SLCO3A1, ST8SIA2, C15orf32, LOC100144604, FAM174B, ASB9P1, LOC100507217, MIR3175, CHD2, RGMA, MCTP2, LOC400456, LOC145820, NR2F2, MIR1469, SPATA8, LOC91948, ARRDC4, FAM169B, MIR4714, IGF1R, PGPEP1L, SYNM, TTC23, HSP90B2P, LRRC28, MEF2A, LYSMD4, DNM1P46, ADAMTS17, FLJ42289, CERS3, PRKXP1, LINS, ASB7, ALDH1A3, LRRK1, CHSY1, VIMP, SNRPA1, LOC100507472, PCSK6, TM2D3, TARSL2, OR4F6, OR4F15, OR4F13P, OR4F4, FAM138E, WASH3P, DDX1*
Chr 22	24,342,813-24,398,622	16,133,474-51,304,566	35.20	0.00	5.56	552	*GSTTP1 GSTT1 GSTTP2*	*LINC00516, LOC101101776, POTEH, OR11H1, CCT8L2, TPTEP1, ANKRD62P1-PARP4P3, XKR3, HSFY1P1, GAB4, CECR7, IL17RA, CECR6, CECR5, CECR5-AS1, CECR1, CECR3, CECR2, SLC25A18, ATP6V1E1, BCL2L13, BID, MIR3198-1, MICAL3, MIR648, FLJ41941, PEX26, TUBA8, USP18, GGT3P, DGCR6, PRODH, DGCR5, DGCR9, DGCR10, DGCR11, DGCR2, TSSK2, DGCR14, GSC2, SLC25A1, CLTCL1, HIRA, MRPL40, C22orf39, UFD1L, CDC45, CLDN5, LOC150185, SEPT5, SEPT5-GP1BB, GP1BB, TBX1, GNB1L, C22orf29, TXNRD2, COMT, MIR4761, ARVCF, MIR185, TANGO2, MIR3618, MIR1306, DGCR8, TRMT2A, RANBP1, ZDHHC8, LOC388849, LOC284865, LOC150197, MIR1286, RTN4R, DGCR6L, LOC729444, TMEM191B, PI4KAP1, RIMBP3, ZNF74, SCARF2, KLHL22, MED15, POM121L4P, TMEM191A, SERPIND1, PI4KA, SNAP29, CRKL, AIFM3, LZTR1, THAP7, THAP7-AS1, TUBA3FP, P2RX6, SLC7A4, P2RX6P, LOC400891, BCRP2, POM121L8P, RIMBP3C, RIMBP3B, HIC2, TMEM191C, PI4KAP2, RIMBP3B, RIMBP3C, UBE2L3, YDJC, CCDC116, SDF2L1, MIR301B, MIR130B, PPIL2, YPEL1, MAPK1, PPM1F, TOP3B, VPREB1, LOC96610, ZNF280B, ZNF280A, PRAME, LOC648691, POM121L1P, GGTLC2, MIR650, IGLL5, GNAZ, RTDR1, RAB36, BCR, FBXW4P1, CES5AP1, ZDHHC8P1, IGLL1, C22orf43, GUSBP11, RGL4, ZNF70, VPREB3, C22orf15, CHCHD10, MMP11, SMARCB1, DERL3, SLC2A11, MIF, LOC284889, GSTT2B, GSTT2, DDTL, DDT, GSTT2, GSTTP1, LOC391322, GSTT1, GSTTP2, CABIN1, SUSD2, GGT5, POM121L9P, SPECC1L, ADORA2A, ADORA2A-AS1, UPB1, GUCD1, SNRPD3, FAM211B, GGT1, BCRP3, POM121L10P, PIWIL3, TOP1P2, SGSM1, TMEM211, LOC100128531, KIAA1671, CRYBB3, CRYBB2, IGLL3P, LRP5L, CRYBB2P1, ADRBK2, MYO18B, SEZ6L, ASPHD2, HPS4, SRRD, TFIP11, MIR548J, TPST2, CRYBB1, CRYBA4, MIAT, MN1, PITPNB, MIR3199-1, MIR3199-2, TTC28-AS1, TTC28, CHEK2, HSCB, CCDC117, XBP1, ZNRF3, ZNRF3-AS1, C22orf31, KREMEN1, EMID1, RHBDD3, EWSR1, GAS2L1, RASL10A, SNORD125, MIR3653, AP1B1, RFPL1-AS1, RFPL1, NEFH, THOC5, NIPSNAP1, NF2, CABP7, ZMAT5, UQCR10, ASCC2, MTMR3, HORMAD2, LIF, OSM, GATSL3, TBC1D10A, SF3A1, CCDC157, KIAA1656, RNF215, SEC14L2, MTFP1, SEC14L3, SDC4P, SEC14L4, SEC14L6, GAL3ST1, PES1, TCN2, SLC35E4, DUSP18, MIR3200, OSBP2, MORC2-AS1, MORC2, TUG1, SMTN, SELM, INPP5J, PLA2G3, MIR3928, RNF185, LIMK2, PIK3IP1, PATZ1, DRG1, EIF4ENIF1, SFI1, PISD, PRR14L, DEPDC5, C22orf24, YWHAH, SLC5A1, AP1B1P1, C22orf42, RFPL2, SLC5A4, RFPL3, RFPL3-AS1, LOC339666, C22orf28, BPIFC, FBXO7, SYN3, TIMP3, MIR4764, LARGE, LOC100506195, ISX, HMGXB4, TOM1, MIR3909, HMOX1, MCM5, RASD2, MB, APOL6, APOL5, RBFOX2, APOL3, APOL4, APOL2, APOL1, MYH9, TXN2, FOXRED2, EIF3D, CACNG2, IFT27, PVALB, NCF4, CSF2RB, LOC100506241, TEX33, TST, MPST, KCTD17,*

APPENDIX 5

Table B: Significant Focal Common Deletions in Gastrointestinal Stromal Tumours (GISTIC)

	Peak Region	Extended Region	Length (Mb)	Q-Bound	G-Score	No. of Genes	Genes in Peak	Genes
Chr 22	24,342,813-24,398,622	16,133,474-51,304,566	35.20	0.00	5.56	552	*GSTTP1 GSTT1 GSTTP2*	*TMPRSS6, IL2RB, C1QTNF6, SSTR3, RAC2, CYTH4, ELFN2, MFNG, CARD10, CDC42EP1, LGALS2, GGA1, SH3BP1, PDXP, LGALS1, NOL12, TRIOBP, H1F0, GCAT, GALR3, ANKRD54, MIR658, MIR659, EIF3L, MICALL1, C22orf23, POLR2F, SOX10, MIR4534, PICK1, SLC16A8, BAIAP2L2, PLA2G6, MAFF, TMEM184B, CSNK1E, LOC400927, KCNJ4, KDELR3, DDX17, DMC1, FAM227A, CBY1, TOMM22, JOSD1, GTPBP1, SUN2, DNAL4, NPTXR, CBX6, APOBEC3A, APOBEC3A_B, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G, APOBEC3H, CBX7, PDGFB, SNORD83B, SNORD83A, RPL3, RNU86, SNORD43, SYNGR1, TAB1, LOC100506472, MGAT3, SMCR7L, ATF4, RPS19BP1, CACNA1I, ENTHD1, GRAP2, FAM83F, LOC100130899, TNRC6B, ADSL, SGSM3, MKL1, MCHR1, SLC25A17, MIR4766, ST13, DNAJB7, XPNPEP3, RBX1, MIR1281, EP300, L3MBTL2, CHADL, RANGAP1, ZC3H7B, TEF, TOB2, PHF5A, ACO2, POLR3H, CSDC2, PMM1, DESI1, XRCC6, NHP2L1, C22orf46, MEI1, CCDC134, SREBF2, MIR33A, SHISA8, TNFRSF13C, CENPM, LINC00634, SEPT3, WBP2NL, NAGA, FAM109B, C22orf32, NDUFA6, LOC100132273, CYP2D6, CYP2D7P1, TCF20, LOC388906, NFAM1, SERHL, RRP7A, SERHL2, RRP7B, POLDIP3, RNU12, CYB5R3, ATP5L2, A4GALT, ARFGAP3, PACSIN2, TTLL1, BIK, MCAT, TSPO, TTLL12, SCUBE1, MPPED1, EFCAB6-AS1, EFCAB6, SULT4A1, PNPLA5, PNPLA3, SAMM50, PARVB, PARVG, KIAA1644, LDOC1L, LINC00207, LINC00229, PRR5, PRR5-ARHGAP8, ARHGAP8, PHF21B, LOC100506714, NUP50, MIR1249, KIAA0930, UPK3A, FAM118A, SMC1B, RIBC2, FBLN1, ATXN10, MIR4762, WNT7B, LOC730668, LOC100271722, C22orf26, LOC150381, MIR3619, MIRLET7BHG, MIRLET7A3, MIR4763, MIRLET7B, PPARA, CDPF1, PKDREJ, TTC38, CN5H6.4, GTSE1, TRMU, CELSR1, GRAMD4, CERK, TBC1D22A, FLJ46257, MIR3201, LOC284933, FAM19A5, MIR4535, LOC100128946, C22orf34, LOC90834, BRD1, ZBED4, ALG12, CRELD2, PIM3, IL17REL, MLC1, MOV10L1, PANX2, TRABD, SELO, TUBGCP6, HDAC10, MAPK12, MAPK11, PLXNB2, DENND6B, PPP6R2, SBF1, ADM2, MIOX, LMF2, NCAPH2, SCO2, TYMP, ODF3B, KLHDC7B, SYCE3, CPT1B, CHKB-CPT1B, CHKB, LOC100144603, MAPK8IP2, ARSA, SHANK3, ACR, RABL2B, RPL23AP82*
						2677		

APPENDIX 5

Table C1: Over-represented (Enriched) Pathways among Commonly-amplified genes in Gastrointestinal Stromal Tumours

Pathway ID	Term	Fisher p-value	EASE p-value	Gene Count	Genes	Fold Enrichment
KEGG hsa05322	Systemic lupus erythematosus	5.80E-04	2.98E-03	7	HIST1H2AC, HIST1H2BD, HIST1H2BE, HIST1H2BF, HIST1H2AD, HIST1H2BG, HIST1H2AE, HIST1H4E, HIST1H4F, HIST1H3D, HIST1H4D, HIST1H3E, HIST1H3F, HIST1H4G	4.79

Table C2: Over-represented (Enriched) Pathways among Commonly-deleted genes in Gastrointestinal Stromal Tumours

Pathway ID	Term	Fisher p-value	EASE p-value	Gene Count	Genes	Fold Enrichment
BIOCARTA h_cpsfPathway	Polyadenylation of mRNA	2.3E-02	2.33E-02	3	*OR10A5, OR10A6, OR10A3, OR10A4, OR52H1, OR52A1, OR51D1, OR10A2, OR52A5, OR4K17, OR4L1, OR4S1, OR51Q1, OR6S1, OR52D1, OR4C3, OR4C13, OR5AU1, OR52L1, OR51T1, OR4C12, OR51L1, OR52B2, OR52B4, OR52B6, OR4A47, OR51G1, CNGA4, OR51G2, OR4E2, OR52K2, OR52K1, OR4K15, OR4K14, OR4N5, OR4K13, OR2D2, OR2D3, OR51B2, OR52J3, OR51F1, ADRBK2, OR51F2, OR11G2, OR10G2, OR10G3, OR52N2, OR56B4, OR52N5, OR52N4, OR52N1, OR6A2, OR56B1, OR51B5, OR51B6, OR51B4, OR51V1, OR52R1, OR4K5, OR5P3, OR2AG1, OR5P2, OR2AG2, OR52I1, OR52I2, OR52E2, OR4B1, OR11H4, OR52E4, OR51E1, OR51E2, OR4K1, OR56A3, OR56A4, OR51M1, OR4X1, OR56A1, OR52M1, OR51I2, OR51S1, OR4X2, OR52E6, OR11H6, OR52E8, OR52W1, OR51A2, OR4C45, OR51A4, OR51I1, OR51A7, OR56A5, CALM1*	11.34
BIOCARTA h_circadianPathway	Circadian Rhythms	2.4E-04	4.69E-03	4	*CRY2, CSNK1E, ARNTL, LOC400927*	10.08
KEGG hsa04740	Olfactory transduction	5.3E-28	2.16E-27	92	*PSMB5, PSMA1, PSMC6, PSMD13, PSME1, PSME2, PSMC3, PSMA3, PSMC1, PSMB11*	3.35
KEGG hsa03050	Proteasome	1.6E-03	5.73E-03	10	*SPTLC2, SGPP1, GALC, ARSA, SMPD1, CERK, GAL3ST1, DEGS2*	2.93
KEGG hsa00600	Sphingolipid metabolism	6.0E-03	2.01E-02	8	*PABPN1, PAPOLA, CSTF3*	2.83

APPENDIX 6

Table A1: Significant Common Focal Amplifications in Undifferentiated Pleomorphic Sarcoma (STAC)

	Cytoband	Region	Length (Mb)	Freq %	P-Value	Genes	Gene Symbols	% of CNV Overlap
Chr 1	q21.1	145,070,378-145,278,200	0.21	68.75	0.00E+00	3	*NOTCH2NL, PDE4DIP, SEC22B*	100
	p32.1	59,228,562-59,722,944	0.49	75	0.00E+00	2	*HSD52, JUN*	34
Chr 2	q12.3	108,867,623-108,889,968	0.02	37.5	2.40E-02	1	*SULT1C3*	100
Chr 3	p12.1	86,880,017-87,076,959	0.20	68.75	1.20E-02	1	*VGLL3*	2
Chr 4	q12	54,400,737-54,592,199	0.19	43.75	1.30E-02	1	*LNX1*	21
Chr 7	q33	135,033,284-135,102,970	0.07	62.5	2.20E-02	1	*CNOT4*	0
Chr 8	q12.1	59,724,608-59,965,524	0.24	50	4.20E-02	1	*TOX*	0
Chr 11	p13	35,080,110-35,282,073	0.20	43.75	0.00E+00	2	*CD44, SLC1A2*	79
Chr 14	q24.1	68,824,970-69,144,273	0.32	56.25	2.70E-02	1	*RAD51B*	60
Chr 21	q22.12	36,617,496-36,764,360	0.15	43.75	1.00E-03	1	*RUNX1*	100
						14		

Table A2: Significant Common Focal Deletions in Undifferentiated Pleomorphic Sarcoma (STAC)

	Cytoband	Region	Length (Mb)	Freq (%)	P-Value	Genes	Gene Symbols	% of CNV Overlap
Chr 4	q35.1	184,425,375-185,143,402	0.72	56.25	3.20E-02	5	*ENPP6, ING2, RWDD4, STOX2, TRAPPC11*	16
	q35.1 - q35.2	186,797,948-187,792,442	0.99	56.25	3.20E-02	9	*CYP4V2, F11, FAM149A, FAT1, FLJ38576, KLKB1,MTNR1A, SORBS2, TLR3*	40
	q35.2	188,398,868-189,739,168	1.34	56.25	3.20E-02	3	*TRIML1, TRIML2, ZFP42*	100
	q35.2	190,624,209-191,154,276	0.53	62.5	1.00E-03	37	*DUX2, DUX4, DUX4L2, DUX4L3, DUX4L4, DUX4L5, DUX4L6, FRG1, FRG2*	74
Chr 9	p21.3	21,936,953-22,013,183	0.08	43.75	1.00E-02	2	*CDKN2A, CDKN2B*	100
Chr 11	q24.1	121,300,330-121,397,853	0.10	62.5	1.80E-02	1	*SORL1*	37
	q24.1	122,766,398-123,471,184	0.70	62.5	1.80E-02	4	*BSX, CLMP, GRAMD1B, HSPA8,*	32
	q24.2	125,355,927-126,358,069	1.00	62.5	1.80E-02	22	*ACRV1, CDON, CHEK1, DCPS, DDX25, EI24, FAM118B, FEZ1, FLJ39051, FOXRED1, HYLS1, KIRREL3, PATE1, PATE2, PATE3, PATE4, PUS3, RPUSD4, SRPR, ST3GAL4, STT3A, TIRAP*	60
	q24.3	129,023,372-130,446,982	1.42	62.5	1.80E-02	10	*ADAMTS15, ADAMTS8, APLP2, ARHGAP32, BARX2, NFRKB, PRDM10, ST14, TMEM45B, ZBTB44*	38
Chr 14	q32.33	104,566,967-104,715,087	0.15	31.25	4.20E-02	2	*ASPG, KIF26A*	63
Chr 17	q11.2	26,726,585-27,042,613	0.32	50	6.00E-03	13	*ALDOC, FOXN1, PIGS, PROCA1, RAB34, SARM1, SDF2, SGK494, SLC13A2, SLC46A1, SPAG5, SUPT6H, UNC119*	1
	q11.2	27,203,586-27,211,735	0.01	56.25	0.00E+00	1	*FLOT2*	0
	q11.2	27,623,194-27,880,882	0.26	62.5	0.00E+00	1	*TAOK1*	9
	q11.2	28,424,381-28,525,378	0.10	56.25	0.00E+00	3	*EFCAB5, NSRP1, SLC6A4*	7
	q11.2	29,041,847-29,371,754	0.33	62.5	0.00E+00	7	*ADAP2, ATAD5, CRLF3, DPRXP4, RNF135, SUZ12P1, TEFM*	16
Chr 19	p13.3	0-350,965	0.35	43.75	1.60E-02	7	*FAM138A, FAM138F, FLJ45445, MIER2, OR4F17, PPAP2C, WASH5P*	83
Chr 20	p13	0-210,288	0.21	37.5	4.30E-02	5	*DEFB125, DEFB126, DEFB127, DEFB128, DEFB129*	69
	p13	398,763-434,066	0.04	37.5	4.30E-02	2	*RBCK1, TBC1D20*	20
Chr 21	q22.3	47,760,929-47,790,354	0.03	56.25	0.00E+00	1	*PCNT*	22
						135		

APPENDIX 6

Table B1: Significant Common Focal Amplifications in Undifferentiated Pleomorphic Sarcoma (GISTIC)

	Peak Region	Extended Region	Length (Mb)	Frequency (%)	G-Score	Q-Bound	No. of Genes	Genes in Peak	Genes
Chr 3	86,816,036-88,468,010	87,299,611-87,306,626	1.7	63	8.19	0.03	6	*CHMP2B*	*VGLL3*, *CHMP2B*, *POU1F1*, *HTR1F*, *CGGBP1*, *ZNF654*
Chr 5	10,606,546-16,879,792	15,942,852-16,064,034	6.3	56	7.96	0.03	14	-	*ANKRD33B*, *DAP*, *CTNND2*, *CT49*, *DNAH5*, *TRIO*, *FAM105A*, *FAM105B*, *ANKH*, *FBXL7*, *MARCH11*, *ZNF622*, *FAM134B*, *MYO10*
Chr 7	77,437,091-94,074,985	82,845,459-83,425,216	16.6	38	8.06	0.03	57	*SEMA3E*	*PHTF2*, *RPL13AP17*, *MAGI2*, *GNAI1*, *GNAT3*, *CD36*, *SEMA3C*, *HGF*, *CACNA2D1*, *PCLO*, *SEMA3E*, *SEMA3A*, *SEMA3D*, *GRM3*, *DMTF1*, *TMEM243*, *TP53TG1*, *CROT*, *ABCB4*, *ABCB1*, *RUNDC3B*, *SLC25A40*, *DBF4*, *ADAM22*, *SRI*, *STEAP4*, *ZNF804B*, *DPY19L2P4*, *STEAP1*, *STEAP2*, *GTPBP10*, *CLDN12*, *CDK14*, *FZD1*, *MTERF*, *AKAP9*, *CYP51A1*, *LRRD1*, *KRIT1*, *ANKIB1*, *GATAD1*, *PEX1*, *RBM48*, *MGC16142*, *FAM133B*, *FAM133DP*, *CDK6*, *SAMD9*, *SAMD9L*, *HEPACAM2*, *CCDC132*, *CALCR*, *TFPI2*, *GNGT1*, *GNG11*, *BET1*, *COL1A2*
Chr 11	101,263,543-103,276,617	101,597,949-101,646,626	2.0	50	14.43	0.00	18	-	*TRPC6*, *ANGPTL5*, *YAP1*, *BIRC3*, *BIRC2*, *TMEM123*, *MMP7*, *MMP20*, *MMP27*, *MMP8*, *MMP10*, *MMP1*, *WTAPP1*, *MMP3*, *MMP12*, *MMP13*, *DCUN1D5*, *DYNC2H1*
							95		

APPENDIX 6

Table B2: Significant Common Focal Deletions in Undifferentiated Pleomorphic Sarcoma (GISTIC)

	Peak Region	Extended Region	Length (Mb)	Frequency (%)	G-Score	Q-Bound	No. of Genes	Genes in Peak	Genes
Chr 1	248,733,443-249,250,621	248,733,443-248,811,992	0.5	69	9.49	0.00006	11	*OR2T34, OR2T10, OR2T11, OR2T35*	*OR2T34*, *OR2T10*, *OR2T11*, *OR2T35*, *OR2T27*, *OR14I1*, *LYPD8*, *SH3BP5L*, *ZNF672*, *ZNF692*, *PGBD2*
Chr 11	128,837,921-135,006,516	134,590,936-134,623,445	6.2	56	6.13	0.03319	23	-	*ARHGAP32*, *BARX2*, *TMEM45B*, *NFRKB*, *PRDM10*, *APLP2*, *ST14*, *ZBTB44*, *ADAMTS8*, *ADAMTS15*, *SNX19*, *NTM*, *OPCML*, *SPATA19*, *IGSF9B*, *JAM3*, *NCAPD3*, *VPS26B*, *THYN1*, *ACAD8*, *GLB1L3*, *GLB1L2*, *B3GAT1*
Chr 13	47,927,728-52,108,681	48,961,498-49,034,756	4.2	81	10.80	0.00001	31	*RB1, LPAR6*	*SUCLA2*, *NUDT15*, *MED4*, *ITM2B*, *RB1*, *LPAR6*, *RCBTB2*, *CYSLTR2*, *FNDC3A*, *MLNR*, *CDADC1*, *CAB39L*, *SETDB2*, *PHF11*, *RCBTB1*, *ARL11*, *EBPL*, *KPNA3*, *CTAGE10P*, *SPRYD7*, *TRIM13*, *KCNRG*, *DLEU2*, *DLEU1*, *ST13P4*, *DLEU7*, *RNASEH2B*, *GUCY1B2*, *FAM124A*, *SERPINE3*, *INTS6*
Chr 17	27,203,586-29,571,405	27,845,429-27,862,668	2.4	63	6.91	0.01042	34	*TAOK1*	*FLOT2*, *DHRS13*, *PHF12*, *SEZ6*, *PIPOX*, *TIAF1*, *MYO18A*, *CRYBA1*, *NUFIP2*, *TAOK1*, *ABHD15*, *TP53I13*, *GIT1*, *ANKRD13B*, *CORO6*, *SSH2*, *EFCAB5*, *NSRP1*, *SLC6A4*, *BLMH*, *TMIGD1*, *CPD*, *GOSR1*, *TBC1D29*, *LRRC37BP1*, *SH3GL1P2*, *SUZ12P1*, *CRLF3*, *ATAD5*, *TEFM*, *ADAP2*, *DPRXP4*, *RNF135*, *NF1*
							99		

APPENDIX 6

Table C1: Over-represented (Enriched) Pathways among Commonly-amplified genes in Undifferentiated Pleomorphic Sarcoma

Category	Term	Count	EASE p-value	Genes	Fold Enrichment
BIOCARTA h_raccPathway	Ion Channels and Their Functional Role in Vascular Endothelium	5	0.0498	*GNAS, GNGT1, NOS3, KCNQ4, SGCB*	3.27
KEGG hsa04742	Taste transduction	26	0.0000	*GNAS, ADCY4, ADCY8, GNG13, PLCB2, TAS1R3, TAS2R1, TAS2R10, TAS2R13, TAS2R14, TAS2R19, TAS2R20, TAS2R3, TAS2R31, TAS2R38, TAS2R4, TAS2R41, TAS2R42, TAS2R43, TAS2R46, TAS2R5, TAS2R50, TAS2R60, TAS2R7, TAS2R8, TAS2R9*	3.06
BIOCARTA h_hesPathway	Segmentation Clock	7	0.0199	*LFNG, NOTCH1, NOTCH2, DVL1, HES7, HEY2, RBPJ*	2.94
BIOCARTA h_il1rPathway	Signal transduction through IL1R	12	0.0052	*ECSIT, IL1RAP, IL1R1, IL1A, IL6, JUN, MAP2K3, MAP2K6, MAP3K14, MAP3K7, NFKBIA, TGFB1*	2.44
BIOCARTA h_nthiPathway	NFkB activation by Nontypeable Hemophilus influenzae	9	0.0370	*SMAD3, MAP2K3, MAP2K6, MAP3K14, MAP3K7, NFKBIA, TGFBR1, TGFBR2, RELA*	2.21
KEGG hsa05020	Prion diseases	12	0.0195	*FYN, NOTCH1, C1QA, C1QB, C1QC, C7, C8A, C8B, C8G, HSPA5, IL1A, IL6*	2.10
KEGG hsa04512	ECM-receptor interaction	28	0.0002	*CD44, AGRN, COL1A2, COL3A1, COL4A1, COL4A2, COL5A1, COL5A2, COL6A1, COL6A2, COL6A3, COL11A1, GP1BA, ITGA1, ITGA11, ITGA2, ITGA9, ITGB5, LAMA1, LAMA3, LAMA4, LAMA5, LAMB1, LAMB4, RELN, SV2A, TNC, TNR*	2.04
BIOCARTA h_keratinocytePathway	Keratinocyte Differentiation	12	0.0434	*CEBPA, TRAF2, EGFR, HOXA7, JUN, MAP2K3, MAP2K4, MAP2K6, MAP2K7, MAP3K14, NFKBIA, PRKCA*	1.86
KEGG hsa05212	Pancreatic cancer	20	0.0191	*BAD, BCL2L1, SMAD3, ACVR1B, CDK6, CDKN2A, EGFR, RALBP1, PIK3CG, RALGDS, RAC2, RAC3, TGFB1, TGFB3, TGFBR1, TGFBR2, BRAF, RALB, RELA, VEGFB*	1.70
KEGG hsa05222	Small cell lung cancer	23	0.0134	*BCL2L1, MAX, TRAF2, COL4A1, COL4A2, CDK6, CDKN1B, CDKN2B, FHIT, ITGA2, LAMA1, LAMA3, LAMA4, LAMA5, LAMB1, LAMB4, NFKBIA, PTENP1, PIK3CG, RARB, RXRA, RXRG, RELA*	1.68
KEGG hsa04360	Axon guidance	34	0.0042	*EPHA1, EPHA2, EPHA3, EPHA6, EPHA8, EPHB2, EPHB6, FYN, SRGAP3, EFNA2, EFNA5, FES, GNAI2, MET, NFATC2, NFATC4, PAK1, PAK4, PAK6, PLXNA2, PLXNB1, PPP3R2, RAC2, RAC3, ROBO1, ROBO2, SEMA3A, SEMA3E, SEMA3F, SEMA4D, SEMA5A, SEMA6A, SLIT2, SLIT3*	1.61
KEGG hsa04520	Adherens junction	20	0.0372	*BAIAP2, FYN, SMAD3, WASF2, WASL, ACVR1B, CTNNB1, EGFR, INSR, MET, MAP3K7, PTPN1, PTPRF, RAC2, RAC3, MLLT4, SNAI2, SSX2IP, TGFBR1, TGFBR2*	1.59

APPENDIX 6

Table C1: Over-represented (Enriched) Pathways among Commonly-amplified genes in Undifferentiated Pleomorphic Sarcoma

Category	Term	Count	EASE p-value	Genes	Fold Enrichment
KEGG hsa04510	Focal adhesion	52	0.0005	*BAD, FYN, RAPGEF1, ARHGAP5, SHC3, CTNNB1, CAV3, COL1A2, COL3A1, COL4A1, COL4A2, COL5A1, COL5A2, COL6A1, COL6A2, COL6A3, COL11A1, EGFR, HGF, ITGA1, ITGA11, ITGA2, ITGA9, ITGB5, JUN, LAMA1, LAMA3, LAMA4, LAMA5, LAMB1, LAMB4, MET, MYL12A, MYL12B, PAK1, PAK4, PAK6, PARVA, PTENP1, PIK3CG, PDGFB, PRKCA, RAC2, RAC3, RELN, TLN2, TNC, TNR, CRK, BRAF, VEGFB, VAV2*	1.58
BIOCARTA h_mapkPathway	MAPKinase Signaling Pathway	21	0.0379	*CEBPA, MKNK1, MAX, TRAF2, JUN, MAPK7, MAP2K3, MAP2K4, MAP2K6, MAP2K7, MAP3K11, MAP3K14, MAP3K2, MAP3K6, MAP3K7, MAP4K2, NFKBIA, PAK1, TGFB1, TGFBR1, BRAF*	1.55
KEGG hsa04010	MAPK signaling pathway	62	0.0027	*ECSIT, MKNK1, MAX, RASGRP1, TRAF2, ACVR1B, ARRB1, CACNA1D, CACNA1E, CACNA1I, CACNA2D3, CACNG1, CACNG4, CACNG5, DUSP7, MECOM, EGFR, FGF10, FGF20, GADD45A, GNA12, GNG12, FGF7, CACNA1C, IL1R1, IL1A, JUND, JUN, MAPK7, MAP2K3, MAP2K4, MAP2K6, MAP2K7, MAP3K11, MAP3K14, MAP3K2, MAP3K6, MAP3K7, TAB2, MAP4K2, NTRK2, NTF3, NFATC2, NFATC4, NR4A1, PAK1, PLA2G2C, PLA2G4E, PDGFB, PRKCA, PPP3R2, RAC2, RAC3, STK3, TGFB1, TGFB3, TGFBR1, TGFBR2, CRK, MOS, BRAF, RELA*	1.42
KEGG hsa05200	Pathways in cancer	72	0.0053	*BAD, BCL2L1, CEBPA, GLI3, MAX, SMAD3, TRAF2, TFG, ACVR1B, APC2, ARNT2, AXIN2, CTNNB1, COL4A1, COL4A2, CDK6, CDKN1B, CDKN2A, CDKN2B, DAPK1, DVL1, MECOM, EGLN2, EGLN3, EGFR, FGF10, FGF20, HGF, RALBP1, FGF7, ITGA2, IL6, JUN, LAMA1, LAMA3, LAMA4, LAMA5, LAMB1, LAMB4, MMP1, MET, MITF, MLH1, NFKBIA, PTCH1, PTENP1, PIK3CG, PDGFB, PRKCA, RALGDS, RAC2, RAC3, RET, RARB, RXRA, RXRG, RUNX1T1, KIT, SHH, TGFB1, TGFB3, TGFBR1, TGFBR2, TPR, CRK, BRAF, RALB, RELA, VEGFB, WNT2, WNT16, WNT5A*	1.34
KEGG hsa04080	Neuroactive ligand-receptor interaction	53	0.0478	*HTR1F, MAS1, ADORA2A, ADRA2C, BDKRB1, BDKRB2, CALCR, CHRM2, C3AR1, CRHR1, DRD5, GRIN1, GRIN2B, GRIN3A, GRIK1, GRIK2, GRIK3, GRM1, GRM3, GRM5, GRM7, GRM8, HRH1, HRH2, HCRTR1, ADORA2B, LEPR, LTB4R, LTB4R2, LPAR2, LPAR3, MC2R, MC5R, NPBWR2, NTSR1, OPRL1, OPRD1, OXTR, PTH1R, PRLR, PTGDR, PTGER2, PTGER3, PRSS1, PRSS2, P2RY13, P2RY14, SSTR1, SSTR3, S1PR3, THRB, TRHR, TSPO*	1.27

APPENDIX 6

Table C2: Over-represented (Enriched) Pathways among Commonly-deleted genes in Undifferentiated Pleomorphic Sarcoma

Pathway ID	Term	Gene Count	Fisher p-value	EASE p-value	Genes	Fold Enrichment
BIOCARTA h_rbPathway	RB Tumor Suppressor/Checkpoint Signaling in response to DNA damage	7	3.10E-04	2.43E-03	*CHEK1, RASGRF1, CDK1, MYT1, NPAT, RB1, TP53*	4.34
KEGG hsa04020	Calcium signaling pathway	38	2.10E+03	3.82E-03	*HTR7, ATP2A1, ATP2B4, ATP2A3, PTK2B, ADCY9, ADRA1A, ADRB3, AVPR1B, CACNA1S, CACNA1H, CALML3, CALML5, CHRM5, F2R, CYSLTR2, DRD1, EDNRB, GRIN2A, GRIN2D, GRM1, ITPKA, MYLK3, PDE1A, PLCG2, PHKB, PHKG1, PRKCG, PPP3CB, PPP3CC, P2RX1, P2RX5, RYR2, SLC8A2, TACR3, ERBB4, VDAC1, VDAC3*	1.58
BIOCARTA h_amiPathway	Acute Myocardial Infarction	7	6.00E-04	4.01E-03	*F2, F2R, F7, F10, AHSP, FGA, PLG*	4.01
BIOCARTA h_ctbp1Pathway	SUMOylation as a mechanism to modulate CtBP-dependent gene responses	5	7.00E-04	7.79E-03	*CTBP1, SUMO1, SAE1, CDH1, ZEB1*	5.32
KEGG hsa00640	Propanoate metabolism	11	2.40E-03	7.84E-03	*HIBCH, ABAT, ACAT1, ACAT2, ACACA, ALDH2, ALDH7A1, LDHA, LDHAL6A, LDHC, SUCLA2*	2.52
BIOCARTA h_cdc25Pathway	cdc25 and chk1 Regulatory Pathway in response to DNA damage	5	1.70E-03	1.40E-02	*CHEK1, RASGRF1, CDK1, MYT1, NPAT*	4.65
BIOCARTA h_intrinsicPathway	Intrinsic Prothrombin Activation Pathway	7	4.20E-03	1.82E-02	*F2, F2R, F10, F11, CYP4V2, FGA, KLKB1*	3.07
BIOCARTA h_caspasePathway	Caspase Cascade in Apoptosis	8	9.90E-03	3.19E-02	*APAF1, CASP10, CASP3, CASP4, CASP7, CASP8, LMNB1, PRF1*	2.48
KEGG hsa04610	Complement and coagulation cascades	16	2.10E-02	4.15E-02	*F2, F2R, F7, F10, F11, CR1, CR2, C4BPA, C4BPB, C5AR1, FGA, FGB, FGG, KLKB1, PLG, SERPINF2*	1.70
BIOCARTA h_fasPathway	FAS signaling pathway (CD95)	9	1.70E-02	4.49E-02	*CFLAR, FAS, CASP10, CASP3, CASP7, CASP8, LMNB1, PTPN13, RB1*	2.16
BIOCARTA h_extrinsicPathway	Extrinsic Prothrombin Activation Pathway	5	9.70E-03	4.76E-02	*F2, F2R, F7, F10, FGA*	3.38

APPENDIX 6

Table D1: Differential Amplifications in Undifferentiated Pleomorphic Sarcoma compared with other STS subtypes

	Region	Cytoband	Length (Mb)	Freq. in , NOS (%)	Freq. in Others (%)	Difference	p-value	q-bound	Count	Genes
Chr 1	17,236,783-17,280,446	p36.13	0.04	56.25	16.22	40.03	0.0064	0.41	1	*CROCC*
	33,570,565-33,600,936	p35.1	0.03	56.25	10.81	45.44	0.0010	0.37	1	*ADC*
	34,489,528-34,777,012	p35.1 - p34.3	0.29	56.25	16.22	40.03	0.0064	0.41	1	*CSMD2*
	54,724,167-54,779,132	p32.3	0.05	50.00	8.11	41.89	0.0014	0.37	1	*SSBP3*
	59,228,562-59,247,642	p32.1	0.02	75.00	29.73	45.27	0.0056	0.41	1	***JUN***
	63,085,146-63,097,309	p31.3	0.01	50.00	8.11	41.89	0.0014	0.37	1	*DOCK7*
Chr 3,	63,310,435-63,415,348	p14.2	0.1	50.00	8.11	41.89	0.0014	0.37	1	*SYNPR*
	85,152,448-86,003,801	p12.1	0.69	50.00	5.41	44.59	0.0005	0.35	1	*CADM2*
	86,880,017-87,076,959	p12.1	0.2	68.75	13.51	55.24	0.0001	0.33	1	***VGLL3***
	87,727,265-88,074,845	p11.2 - p11.1	0.35	56.25	8.11	48.14	0.0004	0.33	1	*HTR1F*
	89,441,269-89,908,861	p11.1	0.47	50.00	5.41	44.59	0.0005	0.35	1	***EPHA3***
Chr 7	15,302,504-15,335,686	p21.2	0.03	62.50	18.92	43.58	0.0034	0.41	1	*AGMO*
	15,964,268-16,510,513	p21.2 - p21.1	0.55	56.25	16.22	40.03	0.0064	0.41	2	*ISPD, SOSTDC1*
	17,985,333-18,169,473	p21.1	0.18	62.50	18.92	43.58	0.0034	0.41	1	*PRPS1L1*
	18,349,312-19,091,057	p21.1	0.12	68.75	24.32	44.43	0.0045	0.41	1	*HDAC9*
	19,808,883-20,020,725	p21.1	0.21	62.50	18.92	43.58	0.0034	0.41	1	*TMEM196*
	20,390,109-20,400,245	p21.1	0.01	62.50	21.62	40.88	0.0098	0.44	1	*ITGB8*
	22,696,520-22,969,146	p15.3	0.27	62.50	18.92	43.58	0.0034	0.41	3	*GPNMB, MALSU1, IGF2BP3*
	23,301,525-23,362,857	p15.3	0.06	56.25	16.22	40.03	0.0064	0.41	3	*IL6, TOMM7, SNORD93*
	24,593,062-24,738,029	p15.3	0.14	62.50	16.22	46.28	0.0023	0.40	2	*MPP6, DFNA5*
	25,241,387-25,274,221	p15.3	0.03	68.75	21.62	47.13	0.0017	0.37	1	*NPVF*
	26,387,705-26,400,185	p15.2	0.01	56.25	16.22	40.03	0.0064	0.41	1	*SNX10*
	27,792,941-27,803,481	p15.2	0.01	56.25	13.51	42.74	0.0023	0.40	1	*TAX1BP1*
	30,417,431-30,561,717	p14.3	0.14	56.25	16.22	40.03	0.0064	0.41	2	*NOD1, GGCT*
	116,680,093-116,921,396	q31.2	0.24	50.00	8.11	41.89	0.0014	0.37	2	*ST7, WNT2*
	135,033,284-135,102,970	q33	0.07	62.50	21.62	40.88	0.0098	0.44	1	*CNOT4*
Chr 9	77,295,023-77,307,482	q21.13	0.01	56.25	10.81	45.44	0.0010	0.37	1	*RORB*
	100,186,536-100,230,335	q22.33	0.04	50.00	8.11	41.89	0.0014	0.37	1	*TDRD7*
Chr 11	101,384,809-101,437,354	q22.1	0.05	43.75	0.00	43.75	0.0001	0.33	1	*TRPC6*
	102,169,359-102,234,053	q22.2	0.06	43.75	2.70	41.05	0.0005	0.35	2	***BIRC3, BIRC2***
	102,393,563-102,740,197	q22.2	0.35	43.75	0.00	43.75	0.0001	0.33	9	*MMP7, MMP20, MMP27, MMP8, MMP10, MMP1, WTAPP1, MMP3, MMP12*
	102,822,550-103,076,857	q22.2 - q22.3	0.25	43.75	0.00	43.75	0.0001	0.33	3	*MMP13, DCUN1D5, DYNC2H1*
Chr 20	31,009,725-31,128,856	q11.21	0.12	56.25	16.22	40.03	0.0064	0.41	1	***ASXL1***
	31,866,250-31,890,835	q11.21	0.02	56.25	16.22	40.03	0.0064	0.41	1	*BFIFB1*
									53	

APPENDIX 6

Table D2: Differential Deletions in Undifferentiated Pleomorphic Sarcomas compared with other STS subtypes

	Region	Cytoband	Length (Mb)	Freq. in NOS (%)	Freq. in Others (%)	Difference	p-value	q-bound	Count	Genes
Chr 1	213,443,234-213,697,129	q32.3	0.25	43.75	2.70	41.05	0.0005	0.38	1	*RPSKC1*
	213,980,880-214,338,265	q32.3	0.36	43.75	2.70	41.05	0.0005	0.38	1	*PROX1*
	215,302,569-215,751,141	q41	0.45	43.75	2.70	41.05	0.0005	0.38	2	*KCNK2, KCTD3*
	216,043,174-216,684,943	q41	0.64	43.75	2.70	41.05	0.0005	0.38	2	*USH2A, ESRRG*
	216,731,644-217,224,745	q41	0.49	43.75	2.70	41.05	0.0005	0.38	1	*ESRRG*
	234,357,539-234,437,898	q42.2	0.08	50.00	5.41	44.59	0.0005	0.38	1	*SLC35F3*
	242,489,058-242,528,211	q43	0.39	43.75	2.70	41.05	0.0005	0.38	1	*PLD5*
	245,514,462-245,689,504	q44	0.18	50.00	5.41	44.59	0.0005	0.38	1	*KIF26B*
Chr 13	70,404,055-70,857,794	q21.33	0.21	75.00	32.43	42.57	0.0066	0.47	2	*KLHL1, ATXN8OS*
	19,296,544-19,550,188	q11 - q12.11	0.25	75.00	29.73	45.27	0.0056	0.47	1	*ANKRD20A9P*
	24,831,951-24,864,286	q12.12	0.03	68.75	27.03	41.72	0.0064	0.47	1	*SPATA13*
	26,218,732-26,263,107	q12.13	0.04	81.25	32.43	48.82	0.0021	0.43	1	*ATP8A2*
	37,588,232-37,598,673	q13.3	0.01	68.75	27.03	41.72	0.0064	0.47	1	*SUPT20H*
	58,178,513-58,211,706	q21.1	0.03	81.25	37.84	43.41	0.0063	0.47	1	*PCDH17*
	67,442,419-67,451,208	q21.32	0.01	75.00	32.43	42.57	0.0066	0.47	1	*PCDH9*
	74,152,253-74,320,849	q22.1	0.17	81.25	32.43	48.82	0.0021	0.43	1	*KLF12*
	92,058,034-92,596,337	q31.3	0.05	68.75	24.32	44.43	0.0045	0.47	1	*GPC5*
	96,808,522-97,090,785	q32.1	0.28	68.75	24.32	44.43	0.0045	0.47	1	*HS6ST3*
Chr 15	83,773,547-83,780,826	q25.2	0.01	43.75	2.70	41.05	0.0005	0.38	1	*TM6SF1*
Chr 16	6,127,389-6,503,206	p13.3	0.38	50.00	8.11	41.89	0.0014	0.42	1	*RBFOX1*
Chr 17	26,726,585-27,042,613	q11.2	0.31	50.00	5.41	44.59	0.0005	0.38	14	*SARM1, SLC46A1, SLC13A2, FOXN1, UNC119, PIGS, ALDOC, SPAG5, SGK494, KIAA0100, SDF2, SUPT6H, PROCA1, RAB34*
	27,203,586-27,211,735	q11.2	0.01	56.25	8.11	48.14	0.0004	0.38	1	*FLOT2*
	27,862,668-27,880,882	q11.2	0.02	62.50	13.51	48.99	0.0006	0.42	1	*TAOK1*
	28,458,050-28,525,378	q11.2	0.07	56.25	13.51	42.74	0.0023	0.43	2	*NSRP1, SLC6A4*
	29,464,657-29,476,900	q11.2	0.01	56.25	16.22	40.03	0.0064	0.47	1	***NF1***
	8,574,964-8,676,013	p13.1	0.1	56.25	16.22	40.03	0.0064	0.47	2	*CCDC42, SPDYE4*
Chr 21	47,760,929-47,790,354	q22.3	0.03	56.25	10.81	45.44	0.0010	0.42	1	*PCNT*
									45	