Thesis for the Master's degree in Molecular Biosciences
Main field of study in biochemistry

Profiling of DNA copy number in sarcomas by array comparative genomic hybridisation and identification of candidate cancer genes

Ana H. Barragán Lid

60 study points

Department of Tumor Biology
Institute for Cancer Research
The Norwegian Radium Hospital

Department of Molecular Biosciences
Faculty of Mathematics and Natural Sciences
University of Oslo 2006

Acknowledgments

The work presented in this thesis has been carried out at the Department of Tumour Biology, Institute for Cancer Research, The Norwegian Radium Hospital during 2005-2006.

I will initially like to thank my supervisor Professor Dr. philos Ola Myklebost for providing me with excellent advice and for helping me during the process of writing this thesis. Being a part of his group has given me the opportunity to learn many aspects of cancer research, and for that I will always be grateful.

I am also grateful to Dr. philos Leonardo Meza-Zepeda and Stine H. Kresse for their guiding through the exciting field of array CGH. Having taught me most of the practical methods and background theory, they had always time to answer my questions. I am especially grateful to Stine; your patience with my endless row of questions has been enormously impressing, I am sure you will be a wise professor in the future.

I will also like to thank Heidi M. Namløs for the expression data on LMS and the microarray core facility at The Norwegian Radium Hospital for all assistance during the microarray experiments.

To all the staff at the Department of Tumour Biology, thank you for making a stimulating and enjoyable environment to work in. Especially Ola Myklebost's group, Magne and Erik; you have been very supportive with your positive attitude at all times.

Finally, I will thank my parents in Colombia; without their endless hard work and preoccupation for giving me an education I would definitely not be here. To my Norwegian host family; thank you for opening your home for me. And last but not least, my husband Knut Andreas for his patience, support and encouragement.

Contents

Acknowledgments ... ii
Abstract ... 1
1 Introduction ... 2
1.1 Cancer in general ... 2
1.2 Genetics of cancer ... 2
1.2.1 Chromosomes ... 2
1.2.3 Gene amplification ... 6
1.3 Sarcomas ... 9
1.3.1 Genetics of Sarcoma ... 11
1.4 Gastrointestinal Stromal Tumours ... 12
1.5 Leiomyosarcoma ... 13
1.6 Aims of this study ... 14

2. Materials and methods ... 15
2.1 Tumour material ... 15
2.1.1 Isolation of genomic DNA ... 16
2.2 Array Comparative Genomic Hybridisation ... 17
2.2.1 Theory ... 17
2.2.2 Labelling and setup for array CGH ... 19
2.3 Artificial Chromosomes ... 22
2.3.1 Isolation of BAC clones ... 22
2.4 Fluorescence in Situ Hybridisation ... 23
2.4.1 FISH Procedure ... 25

3. Results ... 29
3.1 Array Comparative Genomic Hybridisation ... 29
3.1.1 Hierarchical clustering of tumours ... 29
3.1.2 Genetic Alterations in LMS ... 31
3.1.3 Genetic Alterations in GIST ... 37
3.2 Fluorescence in Situ Hybridisation ... 40

4. Discussion ... 43
4.1 Microarray analyses of DNA copy number variation ... 43
4.1.1 Array CGH as a classification tool ... 47
4.1.2 Gene alterations in LMS ... 48
4.1.3 Genetic Alterations in GIST ... 52
4.1.4 Patterns Distinguishing GIST from LMS ... 52
4.2 Validation of Array CGH analyses by FISH ... 53

5. Conclusions and future prospects ... 56

References ... 59
Appendix A: Solutions ... 67
Appendix B: Overview of the clone set in 17p13.3-p11.2 ... 68

Abbreviations

ABL	v-abl Abelson murine Leukemia viral oncogene homolog 1
ACE	Analysis of Copy Number
ALDH3A2	Aldehyde Dehydrogenase 3 family member A2
AURKB	Aurora Kinase B
BAC	Bacterial Artificial Chromosome
BCR	Breakpoint Cluster Region
CAD	Carbamoyl-phosphate synthetase 2, Aspartate transcarbamylase and Dihydroorotase
CD34	CD34 molecule
CCND1	Cyclin D1
CCND3	Cyclin D3
CDK4	Cyclin-Dependent Kinase 4
CDKN2A	Cyclin-Dependent Kinase inhibitor 2A
Cen	Centromere
CGH	Comparative genomic hybridization
c-KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
CTP	Cytidine trisphosphate
CTNNA3	Catenin (cadherin-associated protein), alpha 3
Cy3/5	Cyanide dye 3 and 5
DAPI	4'-6-Diamidino-2-Phenylindole
DBM	Dopamine Beta-hydroxylase
dH_2O	Distillated water
DHFR	Dihydrofolate Reductase
DF	Denaturating Solution
DLEU7	Deleted in Lymphocytic Leukemia, 7
DM	Double Minutes
DNA	Deoxyribonucleic acid
Dpn	Diphosphopyridine nucleotide
EDTA	Ethylendiamine tetraacetic acid
ERBB1	v-erb-b2 erythroblastic leukemia viral oncogene homolog 1
ERBB2	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2
FAM10A4	Family with sequence similarity 10, member A4
FGF4	Fibroblast Growth Factor 4
FISH	Fluorescent in situ Hybridisation
FLF	Flowering Locus C
GI	GastroIntestinal
GIST	Gastrointestinal Stromal Tumour
GLI	Glioma-associated oncogene homolog
HIV	Human Immunodeficiency Virus
HMGIC	High mobility group AT-hook
HSR	Homogeneously Staining chromosomal Regions
ICC	Interstitial Cells of Cajal
INT2	Integrator complex subunit 2
LMS	Leiomyosarcoma
LOH	Loss of heterozygosity
KCNRG	Potassium channel regulator
KRAS	v-Ki-ras2 Kirsten Rat Sarcoma viral oncogene homolog
MAP2K4	Mitogen-Activated Protein Kinase Kinase 4
MAPK7	Mitogen-Activated Protein Kinase 7
MDM2	Mouse Double Minute 2
MFAP4	Microfibrillar-associated Protein 4
MFH	Malignant Fibrous Histiocytoma
MPNST	Malignant Peripheral Nerve Sheath Tumour
MYB	v-myb MyeloBlastosis viral oncogene homolog
MYC	v-myc MyeloCytomatosis viral oncogene homolog
NaAc	Sodium Acetate
NF1	NeuroFibromin 1
N-MYC	v-myc MyeloCytomatosis viral related oncogene, neuroblastoma derived
NRAS	Neuroblastoma RAS viral (v-ras) oncogene homolog
p/q	p/q chromosome arm
PAC	P1 phage-derived Artificial Chromosome
PBS	Phosphate Buffered Saline
PCR	Polymerase chain reaction
PDGFRA	Platelet Derived Growth Factor Receptor A
PDGFRB	Platelet Derived Growth Factor Receptor B
PRUNE	Prune Homolog
RAS	Rat Sarcoma oncogene
RB1	Retinoblastoma 1
RDA	Representional Difference Analysis
RLGS	Restriction Landmark Genome Scanning
RT	Room Temparature
SAM	Significance Analysis of Microarrays
SAS	Sarcoma Amplified Sequence
SDS	Sodium Dodecyl Sulfate
SHMT1	Serine Hydroxymethyltransferase 1
SMA	Spinal Muscular Atrophy
SPECC1	Spectrin Domain with Coiled-Coils 1
SREBF1	Sterol Regulatory Element Binding transcription Factor 1
SSC	Sodium Citrate
TE	Tris EDTA
Ter	Terminus (Telomere)
TP53	Tumor Protein p53
TSG	Tumour Suppressor Gene
TY	Tryptone-Yeast

Abstract

Cancer is a disease resulting from an accumulation of acquired genetic mutations. The consequence is an uncontrolled growth of cells and disruption of normal control mechanisms. Complex chromosomal aberrations such as amplification and deletion of DNA copy number can lead to the activation and deregulation of oncogenes and tumour suppressor genes respectively, leading to uncontrolled cell growth and giving rise to tumours.

In sarcomas, rare malignant tumours of mesenchymal origin, aberrations such as amplifications and losses of DNA are frequently seen. In this project, a panel of 13 leiomyosarcomas (LMS) and seven gastrointestinal stromal tumours (GIST) were analysed by array comparative genomic hybridisation (array CGH). This technique makes it possible to map DNA copy number changes and identify chromosomal regions containing "target genes" responsible for tumour development and/or progression.

The most frequent aberrations observed in GISTs were losses of the whole or parts of chromosome 22, seen in all tumours with a minimal recurrent region in 22q12.2-q13.31, as well as chromosome 14, 1p36.32-p13.1, 13q12.11-q33.2, 15q13.2-qtel and 9q13-q34.2.

In leiomyosarcomas, the most recurrent aberrations were loss of 10q21.13 and 13q14.2-q14.3. The region in chromosome 17p13.1-p11.2 presented high amplification and its analysis revealed nine candidate genes. Four genomic clones within this region were tested in three LMS samples by fluorescence *in situ* hybridisation (FISH). LMS1, -10 and -25 showed different levels of DNA copy number although LMS10 was expected to have normal copy number in this region.

Only two genes previously cited in literature were contained in the clones tested by FISH although other clones within the amplicon could contain the actual "target" genes; those were *MAP2K4* often mutated in many tumour types and *SPECC1* involved in juvenile myelomonocytic leukaemia. These genes may be useful in studies of the biology of LMS and should be investigated further.

1 Introduction

1.1 Cancer in general

Cancer is one of the most frequent causes of death in humans. Although there are many kinds of cancer, they all start because of uncontrolled growth of cells. This growing mass of abnormal cells gives rise to a tumour - or neoplasm.

Normal cells follow a cell cycle where they grow, divide and die in a controlled manner. Cancer cells continue to grow and divide, but instead of dying, they will form new abnormal cells. As long as the tumour cells remain together but separate from surrounding normal tissues, the tumour is said to be benign. If the tumour cells gain the ability to invade surrounding tissue, then it is a cancer and becomes malignant (Alberts et al. 2002 for review).

The abnormal cells will pass its abnormalities to its progeny and accumulation of further genetic and epigenetic changes within the cell will lead to further changes in gene activity. Several factors as genetic variation in response to external/endogenous carcinogens, DNA damage and disturbance of gene regulation can influence the evolution of cancer.

1.2 Genetics of cancer

1.2.1 Chromosomes

The genetic information of a cell is contained in their genes residing in chromosomes. Chromosomes are composed of chromatin (DNA and DNA binding protein complex) tightly packed in the cell nucleus. Usually, the chromosomes are not visible in the interphase nucleus (non-dividing phase) because the chromatin is so loosely packed that single chromatin threads are not detectable. In metaphase, the chromosomes can be easily distinguished. Metaphase chromosomes have two sister chromatids held together by the centromere. The centromere divides each of the chromatids in two arms; the short arm of each chromosome is designated “p” (for petit) and the long arm is “q”. The ends of chromosomes are called telomeres (See Speicher 2005 for review).

Normal human cells have 22 matching pairs of non-sex chromosome and one pair of sex chromosomes. The 44 non-sex chromosomes are known as autosomes. The homologous pairs are arranged by geneticists after their decreasing size to produce a karyotype (in humans, 1 is

the longest autosome and 22 the shortest). In other words, a karyotype is a visual description of a set of chromosomes in one cell (Hartwell L. 2000 for review).

The position of a gene can be identified on maps that describe the cytogenetic location according distinctive pattern of light and dark bands that appear when the chromosome is Giemsa stained. The position is usually designated by two digits (representing a region and a band), which are sometimes followed by a decimal point and one or more additional digits (representing sub-bands within a light or dark area). The number indicating the gene position increases with distance from the centromere. For example: 14q21 represents position 21 on the long arm of chromosome 14. 14q21 is closer to the centromere than 14q22. The abbreviations "cen" or "ter" are also used to describe a gene's cytogenetic location. "cen" indicates that the gene is very close to the *centromere* while "ter" stands for *terminus*, which indicates that the gene is very close to the end of the p or q arm.

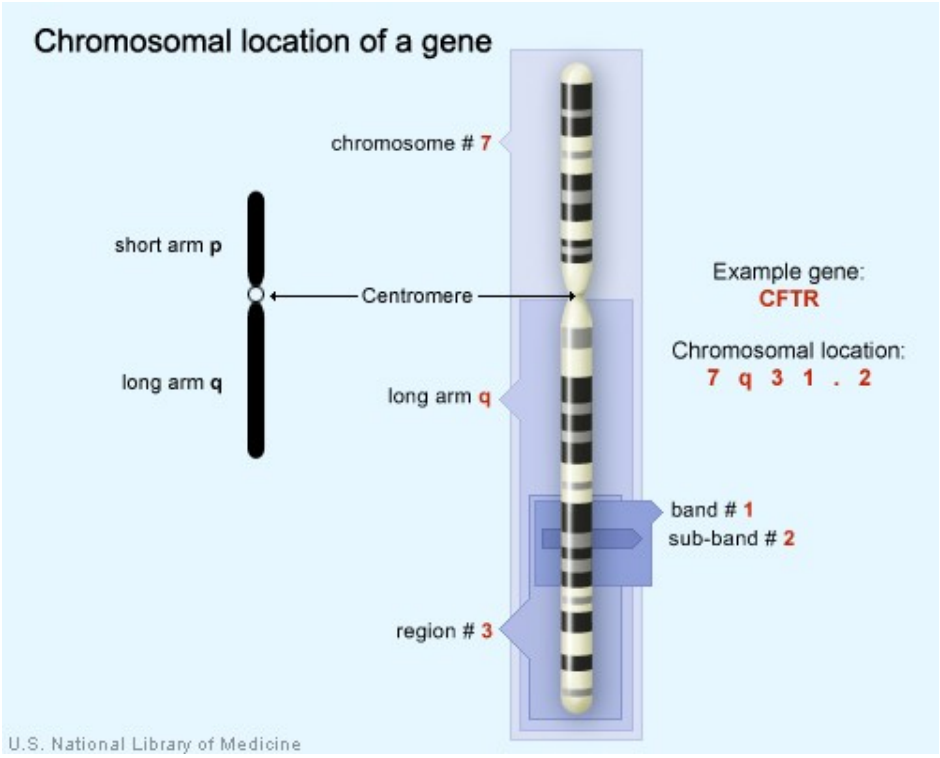

Figure 1.1 How to indicate the localization of a gene (from Genetics Home Reference, U.S. National Library of Medicine)

1.2.2 Cancer

A normal cell has to change many aspects of its phenotype to become a cancer cell, and these changes probably require the alterations of many genes. Cancer is thus, a malignancy involving the accumulation of acquired genetic and epigenetic aberrations.

These alterations can be divided into six categories; i) point mutations or deletion/insertion of single or few nucleotides; ii) Alterations in chromosome number involving losses or gains of whole chromosomes (aneuploidy); iii) Chromosome translocations detected as fusion of different chromosomes or genes, where the fused gene will induce tumorigenic properties; iv) epigenetic modifications, as loss or gain of genomic DNA methylation, local CpG island hypermethylation-associated gene silencing and histone modification patterns (Esteller 2006); v) gene deletions causing the loss of one or both copies of genes involved in carcinogenesis; and vi) gene amplifications, where several copies of genes (or regions –amplicon) can be seen. These amplicons are different from duplications of larger regions that result from aneuploidy and translocations (Lengauer et al. 1998).

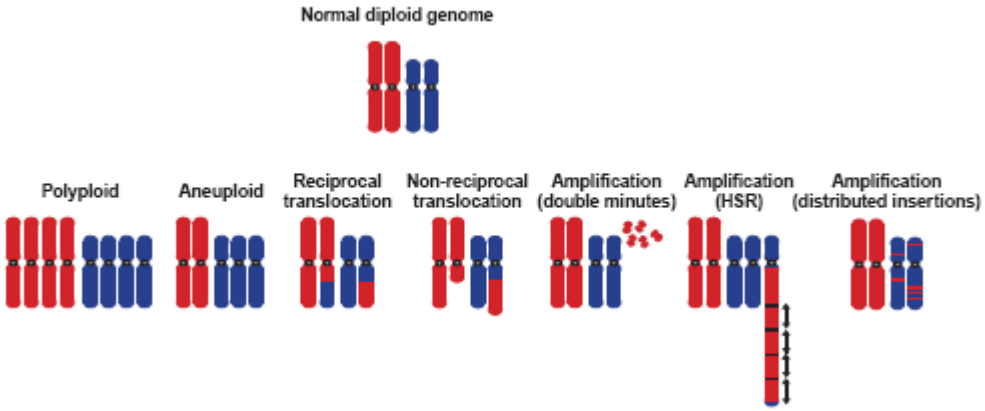

Figure 1.2 Chromosomal aberrations leading to aneuploidy (Figure from Albertson et al. 2003)

The past years, with the advent of new molecular biology technology, and DNA sequencing in particular, both smaller and more abundant alterations have been observed. Such differences include single nucleotide polymorphisms (SNPs), various repetitive elements that involve short DNA sequences (as micro- and microsatellites), and small insertions, deletions, inversions and duplications, also known as copy number polymorphisms (CNPs) (Feuk et al. 2006). CNPs involve gain or losses of several kilobases to hundreds of kilobases of genomic DNA among phenotypically normal individuals.

Identification of disease-specific chromosome aberrations is an important step towards defining the genes involved, particularly since the net effect is the gain or loss of specific gene function. There are two general types of mutations found in tumours: those that improperly activate and those that inactivate protein function. The mutant alleles leading to cancer are referred as cancer genes: Oncogenes and tumour suppressor genes (TSG).

Oncogenes act dominantly (Fig 1.3A). They become activated either by structural alteration or amplification. Chromosomal aberrations affecting oncogenes are mutation, translocation, inversion and amplification. Normal genes that become oncogenes by mutation are known as proto-oncogenes. Examples of common oncogenes are *MYC, MYB, RAF* and *RAS,* often related to leukaemia, neuroblastoma and other neoplasms (Alitalo et al. 1984; Lengauer et al. 1998; Dang 1999; Pinson et al. 2001).

TSGs contribute to cancer in a recessive manner (Fig 1.3B); meaning that loss or inactivation of both alleles is required to target TSGs (Knudson 1971). Chromosome loss is one mechanism for inactivation of TSGs, as well as partial deletions, mutations and epigenetic silencing. They result in removal of their checkpoint and/or inhibitory effect on cell growth and many vital cell pathways (See Roberts 2001 for review). Inheritance of a single mutant allele of many TSGs increases the risk for developing certain types of cancer, e.g. *RB1* in Retinoblastoma, *TP53* in Li-Fraumeni syndrome and *APC* in colon cancer (Li et al. 1969; Knudson 1971; Lindahl 1996).

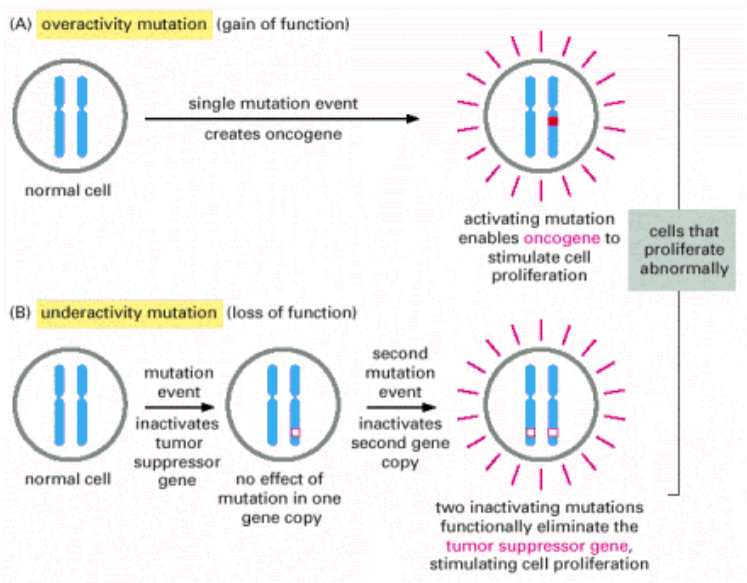

Figure 1.3 Cancer genes act dominantly or recessive; A. One single mutation activates an oncogene stimulating proliferation. B. Two mutations are needed to inactivate TSGs stimulating proliferation (Figure from Alberts B 2002)

Searching for new oncogenes and TSGs is of great importance as many of these genes control essential regulatory steps in the cell. Understanding more about these genes will lead to improvements in cancer diagnosis and treatment, making it selective as we could target directly the genes (and pathways) regulating critical cell mechanisms.

1.2.3 Gene amplification

Genomic amplification is a selective increase in copy number of subchromosomal DNA sequences. Amplification occurs more easily in cancer cells than in normal cells, that is a very important oncogenic process, as genes may become highly up-regulated and that is clearly associated with tumour progression (Lengauer et al. 1998).

Amplicons (amplified regions) can be large and complex, and within them there can be present many genes. Identifying and defining amplified DNA in cancer cells has been a strategy for the isolation of many proto-oncogenes involved in growth control and possibly in tumorigenesis.

Oncogene activation may involve exchange of material between two chromosomes in a balanced or unbalanced order. The significance of amplification emerged from the analysis of tumour cells carrying chromosomal abnormalities, double minutes (DMs) and homogeneously staining chromosomal regions (HSR) which has been known to signal the presence of amplified DNA (See Schwab 1999 for review). DMs appear as small, spherical, chromosome-like structures and may contain circular DNA in chromatin form. They are products of extrachromosomal amplification. HSRs may also contain genetic material of different chromosomal origins (See Myllykangas and Knuutila 2006 for review)

Amplified oncogenes can be assembled in two different ways; i) first the amplified DNA can be intra-chromosomal, residing at the chromosomal site of the single copy gene involved. This amplification could further proceed by unequal but homologous sister chromatide exchanges and secondary arrangements; ii) Second, the DNA can be amplified extrachromosomally or in a chromosomal region distant from the resident site of the single copy gene (e.g. double minutes and episomes). This type of amplification gives rather short co-amplified DNA (Schwab 1999).

According to the breakage-fusion-bridge model (McClintock 1942), the initiating event in HSR formation is the breakage of double chromatid in e.g. fragile sites. After replication, the two sister chromosomes fuse, as a consequence of the action of DNA repair proteins. In mitosis the fused chromatids form a bridge, where two copies of an oncogene would be arranged head to head. If this structure breaks asymmetrically, the daughter cells will receive either a duplicated oncogene or a deleted one (See Fig. 1.4)

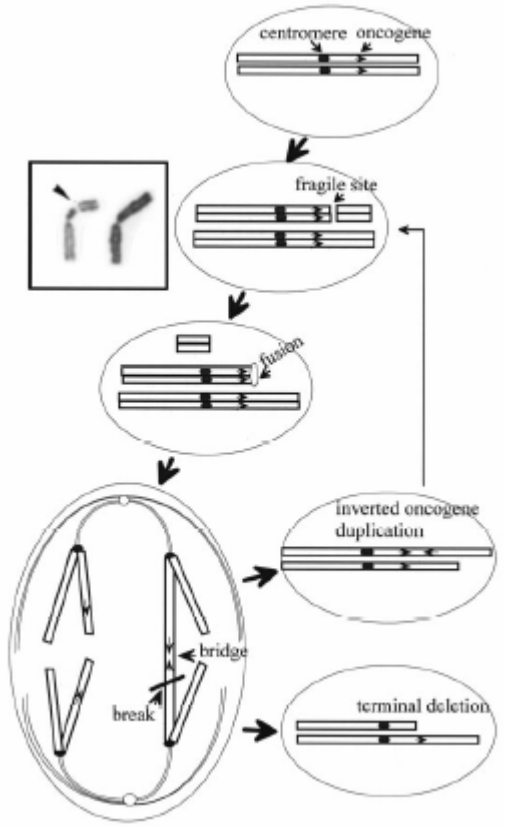

Figure 1.4 Gene amplification according to the breakage-fusion-bridge model (Figure from Schwab 1999)

Classical examples in human tumours involve the gene amplification of the Epidermal Growth Factor *(ERBB),* Rat Sarcoma oncogen *(RAS)* and v-myc myelocytomatosis viral oncogene homolog *(MYC)* families. Drug resistance is frequently induced by amplification of drug target genes, e.g. culture cells selected for resistance to N-(phosphonacetyl)-L-aspartate frequently amplify *CAD*. Dihydrofolate Reductase (*DHFR)* gene amplification conducts to methotrexate resistance, a DHFR enzyme inhibitor used in the treatment of various malignancies. Likewise, the *BCR-ABL* fusion gene encodes a mutant tyrosine kinase, which promotes the pathogenesis of chronic myeloid leukemia. Table 1.1 shows some of the genes frequently amplified in human malignancies and their location in the genome.

Table 1.1 Oncogenes frequently amplified in human malignancies

Oncogene	Function	Tumour type	Localisation
ABL	Protein tyrosine kinase	Chromic myologenous leukaemia	9q34
CCND1	Cyclin D1	Breast cancer, oesophageal and head and neck cancer, bladder cancer	11q13
CDK4	Cyclin dependent kinase 4, phosphorylation of pRb	Sarcoma, glioblastoma	12q13-q15
C-MYC	DNA-binding protein	Breast cancer, ovarian cancer, carcinoma of the uterine cervix, squamous cell lung carcinoma, osteosarcoma, gastric cancer, colon cancer	8q24
ERBB1/EGFR	Epidermal growth factor receptor	Glioma, head and neck squamous cell carcinoma, lung carcinoma, breast cancer	7p12
ERBB2/HER2	Growth factor receptor	Breast cancer, ovarian cancer, gastric and oesophageal cancer	17q11-q12
FGF4	Fibroblast growth factor 4	Kaposi sarcoma, breast cancer, oesophageal carcinoma	11q13.3
GLI	Zinc finger protein	Glioma	12q13-q15
HMGIC	Architectural transcription factor	Sarcoma	12q13-q15
INT2	Fibroblast growth factor 3	Kaposi sarcoma	11q13
KRAS2	GTPase	Adrenocortical tumours, giant cell carcinoma of the lung, breast cancer, ovarian cancer, gastric carcinoma	12p12
MDM2	Binding and inactivation of p53	Sarcoma, glioblastoma	12q13-q15
MYB	DNA-binding protein	Pancreatic cancer, leukaemia, colon carcinoma, melanoma	6q22-q24
N-MYC	DNA-binding protein	Neuroblastoma, retinoblastoma small-cell lung carcinoma, astrocytoma, rhabdomyosarcoma	2p24
NRAS	GTPase	Breast cancer, lung carcinoma, head and neck squamous cell carcinoma	1p13
SAS	Transmembrane protein	Sarcoma	12q13-q15

Traditionally, genomic aberrations have been studied using cytogenetics. Conventional chromosome analysis based on banding was one of the first techniques used to study the cellular aspects of heredity, chromosome structure and causes of disease. Later, Gall and L. Pardue (1969) reported the use of DNA-RNA hybridisation to localize the genes encoding ribosomal RNA, giving birth to *in situ* hybridisation.

Fluorescence *in situ* hybridisation (FISH) was introduced by Rudkin and Stollar (1977). The technique used fluorescently labelled antibodies that recognized specific DNA-RNA hybrids. Now, DNA or RNA sequences (probes) are hybridised to a target such as metaphase chromosomes, interphase nuclei or extended chromatin fibres. However, there have been huge advances in FISH-based techniques as new technology is being developed. Comparative Genomic Hybridisation (CGH) is an example of that.

CGH was developed by Kallioniemi (1992) in order to overcome the difficulties in preparation of high-quality metaphase spreads from solid tumours. This technique is still used to detect and map DNA copy number changes throughout the genome. However, the

microarray technology makes it possible to use mapped clones instead of metaphase chromosomes.

Other molecular genetic technologies used in chromosome aberration analysis are also being widely used; High-throughput analysis of loss of heterozygosity (LOH), restriction landmark genome scanning (RLGS) and representional difference analysis (RDA) are to be mentioned.

1.3 Sarcomas

Cancers are classified according to the tissue from which they arise. Cancers arising from connective or supportive tissue or muscle cells are termed sarcomas (See figure 1.5).

Sarcomas are rare malignant tumours and account for approx 1% of all human cancers. The patient's average age for is 60-65 years old, nevertheless sarcoma can also develop in children and youngsters, accounting for 10% of all cancers in young people (Bjerkehagen and Myklebost 2005).

Sarcomas are generally derived from the mesenchymal tissue. The mesenchyme is a loose network of cells within the mesoderm, one of the three primary germ layers created at an early stage in the embryo. They give rise to as connective tissue, bone, cartilage, and other structures and systems, e.g. blood cells, smooth muscle cells, circulatory system, etcetera (Alberts et. Al. 2002).

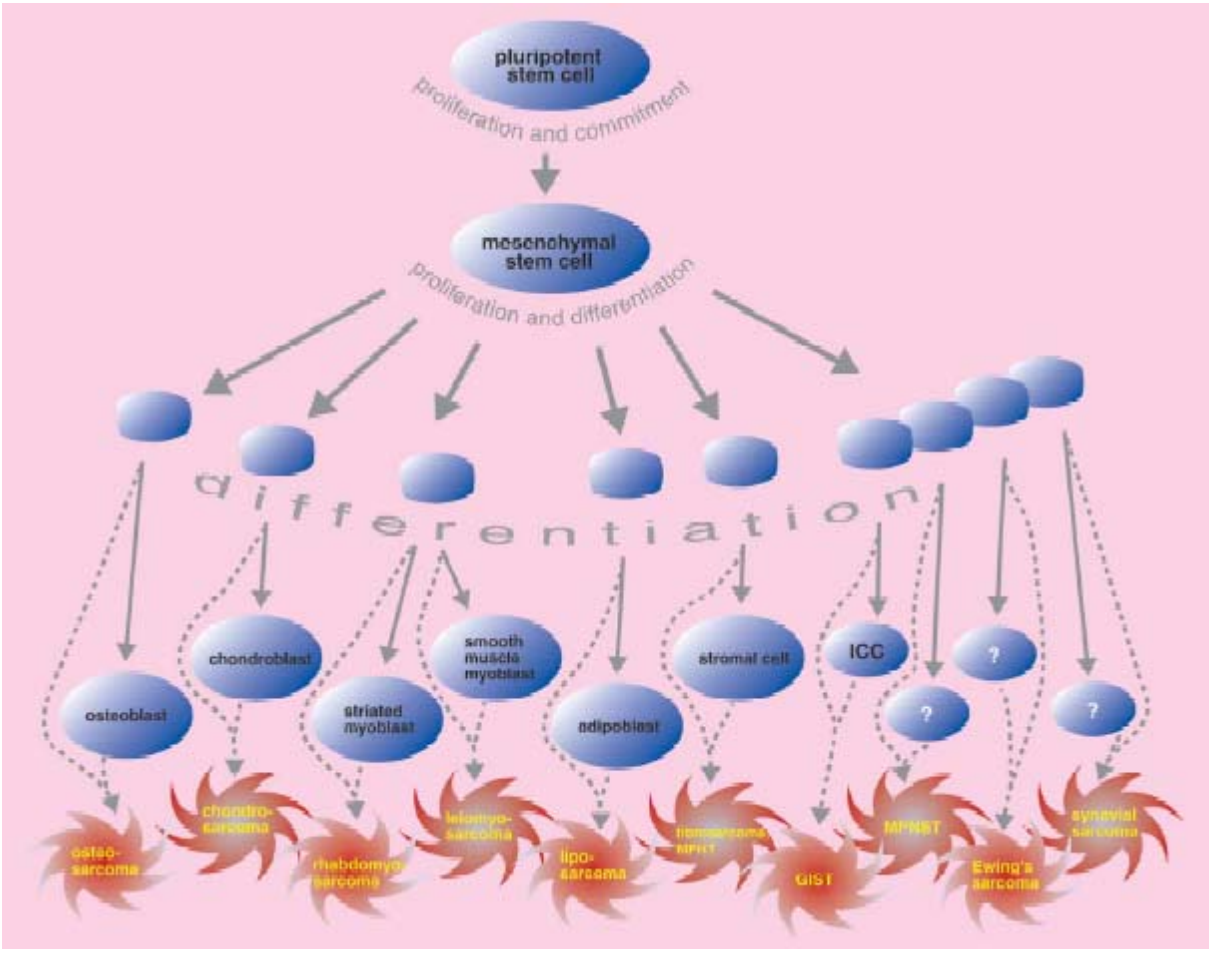

Figure 1.5 Pathways of sarcoma development from tissue of mesenchymal origin (Figure from Mackall et al. 2002)

Sarcomas can be classified in sarcomas of the bone and soft tissue sarcomas. Table 1.2 shows some of the different subtypes of sarcomas.

Table 1.2 Most common Sarcoma types and the tissue they resemble

Sarcoma Type	Normal Tissue
Osteosarcoma	Bone cells
Liposarcoma	Fat tissue
Fibrosarcoma	Fibrous tissue
Rabdomyosarcoma	Striated muscle tissue
Leiomyosarcoma	Smooth muscle tissue
Synovial sarcoma	Joints
Malignant Peripheral Nerve Sheath	Peripheral nerves sheath
Kaposi's sarcoma, Angiosarcoma and Hemangiopericytomas	Blood or lymphatic vessels
Malignant Fibrous Histiocytoma	Fibrous tissue

Most sarcomas are sporadic but some are associated to genetic predisposition syndromes or environmental exposure. As an example, patients with retinoblastoma are at an increased risk of developing sarcomas later in life. *TP53* mutations in Li-Fraumeni syndrome, *NF1* in neurofibromatosis type I and *c-KIT* in gastrointestinal stromal tumours (GISTs) can all lead to

different types of genetic predisposition to sarcoma (See Helman and Meltzer 2003 for review).

Different diagnostic criteria are used when classifying sarcoma types. Traditionally, sarcoma classification was based on histology and pathology, but now genetic techniques are widely used. This is the result of new knowledge on molecular alterations which are present in some subtypes of sarcomas.

1.3.1 Genetics of Sarcoma

Sarcomas can be divided into two main types by their genetic aberrations. One group has simple near-diploid karyotypes with few chromosome rearrangements, whereas the other has complex karyotypes with severe disturbance in genomic stability (Helman L. and Meltzer P., 2003).

The first group contains disease-specific chromosome translocations. These translocations create fusion genes that are related to growth-factor signalling cascades, having dramatic effects on the pathways they affect. In Ewing`s sarcoma, gene fusions between the *EWS* and *FLI1* gene are common. The *EWS* gene is also translocated in myxoid/round cell liposarcoma and extraskeletal myxoid chondrosarcoma. The *SYT* gene in synovial sarcoma is also involved in gene fusions with the genes *SSX1* and *SSX2* (Bjerkehagen and Myklebost 2005). Other chromosomal changes as inversions, deletions and insertions are also present in this group of sarcomas.

The second group contains aneuploid tumours that show complex chromosomal rearrangements but not simple reciprocal translocations. Many of these aberrations result in copy number changes, such as amplification of proto-oncogene-containing regions. Malignant Fibrous Histiocytoma (MFH), Osteosarcoma and Leiomyosarcoma (LMS) are examples of sarcomas with complex karyotypes.

Some gene alterations seen in other types of cancer are also being found in sarcomas. This is the case of the *TP53* gene. p53-regulation of the cell cycle and cell death can be repressed by high production of MDM2, a protein blocking p53. *MDM2* is often amplified in many types of sarcoma although never in tumours where p53 is inactive by mutation. Another important

gene is *RB1*, coding for the retinoblastoma protein pRB. *RB1* is often inactivated by higher production of different proteins, e.g. *CDK4*. This gene codes for a kinase capable of inactivating pRB by phosphorylation by heightened production. In other cases, this kinase can be highly activated because the gene coding of the p16-protein is deleted and therefore is unable to repress *CDK4* (Bjerkehagen and Myklebost 2005).

Although some genes involved in sarcoma development have been identified, they are few in the spectra of genes controlling different pathways in the cells leading to tumour progression. The use of modern molecular biochemical techniques, DNA microarray in particular, makes the approach to target genes easier and, by understanding how they affect the cell we can learn more about the nature behind sarcoma development.

1.4 Gastrointestinal Stromal Tumours

Gastrointestinal stromal tumours (GIST) are uncommon tumours that usually occur in middle-aged or older persons. These tumours are found in the gastrointestinal tract (mostly stomach and intestine) and account for 1-3% of all gastrointestinal malignancies.

Gastrointestinal stromal tumours (GISTs) were thought to be of a heterogeneous group of mesenchymal tumours, based on the finding that some of these tumours were very similar to smooth muscle tumours; they were known as leiomyoblastoma, leiomyoma or leiomyosarcoma. However, is it thought that GIST is derived from the interstitial cells of Cajal (ICC). The ICC are a network of unique, innervated cells that are situated between the peripheral nervous system and the smooth muscle cells of the gastrointestinal (GI) tract. They develop from mesenchymal cell precursors that give rise to both ICC and true smooth muscle cells. Their principal function is to serve as pacemaker cells responsible for generating rhythmic contractions of the GI tract involved in digestion and peristalsis (See Tornillo et al. 2005 for review).

Both ICC and GIST show diffuse strong c-KIT (CD117) and CD34 protein expression, therefore GIST have been suggested to be tumours of the ICC (See Sandberg and Bridge 2002 for review). It could thus be argued that GIST is not a mesenchymal tumour, but of neurogenic origin. However, this is also the case for e.g. malignant peripheral nerve sheath tumours (MPNST), but both subtypes are still regarded as sarcomas.

c-KIT (KIT) also known as CD117, is a type III receptor tyrosine kinase, that is involved in the development and maintenance of ICC. Binding of KIT ligand, also known as stem cell factor causes KIT dimerization and autophosphorylation through phosphorylation of critical tyrosine residues. This activation leads to the phosphorylation of other signal transduction proteins; many of them have kinase activity, resulting in modulation of cellular behaviour including proliferation, chemotaxis and apoptosis (See Rubin 2006 for review).

However, KIT is not the only target of mutation in GIST. The platelet derived growth factor receptor A gene *(PDGFRA)* is also mutated in some GIST. PDGFRA is a member of the same family of receptor tyrosine kinases as KIT, and they are thus very similar. A tyrosine kinase inhibitor, Imatinib mesylate has been used to target KIT, PDGFRA and PDGFRB. This drug is widely used for the treatment of metastatic unresectable GIST, but its efficiency in primary GIST is not yet certain (See Tornillo et al. 2005 for review).

GIST has complex karyotype changes. Losses of chromosome 14, 22q, 1p, 9p or 11p are the most common cytogenetic findings (Bardi et al. 1992; Sreekantaiah et al. 1993; El-Rifai et al. 2000; Heinrich et al. 2003 and more). High-level DNA amplification at 3q26-q29, 5p and 8q22-q24 (el-Rifai et al. 1996) as well as gain at 19q13 have also been observed (Knuutila et al. 1998).

1.5 Leiomyosarcoma

Leiomyosarcoma is a malignant tumour resembling smooth muscle tissue, relatively rare, accounting for 10% of the soft tissue sarcomas. It usually occurs in middle-aged or older persons, although it can also occur in young adults and children. LMS has also become the second leading malignancy of children with human immunodeficiency virus (HIV) infection or other immunodeficiency diseases (Sandberg A, 2005). Leiomyosarcomas arise often in the retroperitoneum, but they can also develop in the uterus, gastrointestinal system and extremities. The cause of leiomyosarcoma is still unknown (Fletcher CDM. et. Al., 2002).

This type of soft tissue sarcoma has a complex karyotype, and no consistent aberrations have been noted. Cytogenetic findings show frequent gain of chromosome region 1q21-31, and loss of 3p21-23, 8p21-pter, 13q12-13 and 13q32-qter. The variation in these regions is also common among subtypes. Previous studies by CGH shows gain of material from

chromosomes 1, 15, 17, 19, 20, 22 and X and loss from 1q, 2, 4q, 9p, 10, 11q, 13q and 16, and have identified regions of amplification in 1q21, 5p14-pter, 12q13-15, 13q31, 17p11 and 20q13 (Fletcher CDM 2002).

From previous studies, some genes have been related to LMS progression. The loss of chromosome 13 material showed association to the Retinoblastoma 1 gene (*RB1*). The Rb-cyclin D pathway involving *RB1, CDKN2A, CCNDI, CDK4* and *CCND3* shows abnormalities in LMS. Amplification at different regions of the genome suggests candidate genes including *MDM2, GLI* and *SAS* at 12q13-15, the *FLF* and *PRUNE* genes at 1q21, and the critical region involved in Smith-Magenis syndrome at 17p11.2 (Fletcher CDM 2002).

KIT expression can be used to differentiate GIST from LMS and other types of soft tissue sarcoma. Approximately 70-80 % of the mutations in GISTs occur in exon 11. Most LMS lacks mutations of exon 11 of *c-KIT*, although studies in uterine LMS have occasional reported expression of KIT (Caudell et al. 2005). This mutation makes the tumour responsive to imatinib mesylate therapy, although mutations in other exons maybe causing low or no response to the treatment. The lack of expression of KIT, CD34, cytokeratin, myoglobin, or neural markers may be used diagnostically to distinguish LMS from other tumours.

1.6 Aims of this study

In sarcomas, alterations in DNA copy number are frequently seen. The goal of this study was to identify target genes that could be important for leiomyosarcoma development and/or progression by determining novel areas for amplification and deletion in a panel of human leiomyosarcomas using array comparative genomic hybridization.

2. Materials and methods

A list of solutions and reagents can be found in the appendix A.

2.1 Tumour material

In this project, tumor samples from 20 human sarcomas were used. 19 initially diagnosed leiomyosarcoma samples (LMS), some of which were later revised to GIST, and one malignant fibrous histiocytoma sample (MFH), later revised to LMS (Table 2.1). The anonymous and frozen samples were obtained from the Norwegian Radium Hospital's biobank.

Table 2.1 Summary of the 20 human sarcomas studied in this project, their diagnosis and location.

Sample	Sample origin	Age/Sex	Diagnosis		Location
			Initial	Revised	
GIST1	Prim	73/M	LMS	GIST	Abdomen
GIST2	Rec	52/M	LMS	GIST	Small bowel
GIST3	Rec	61/F	LMS	GIST	Small bowel
GIST4	Rec	47/M	LMS	GIST	Rectum
GIST5	Met	53/M	LMS	GIST	Liver
GIST7	Prim	74/M	LMS	GIST	Abdomen
GIST8	Prim	70/M	LMS	GIST	Stomach
LMS1*	Prim	59/F	LMS	LMS	Retroperitonum
LMS3	Prim	72/F	LMS	LMS	Retroperitonum
LMS5x	Prim	46/F	LMS	LMS	Uterus
LMS7	Prim	71/F	LMS	LMS	Thigh
LMS10	Prim	67/F	LMS	LMS	Retroperitonum
LMS12[4]	Prim	67/M	LMS	LMS	Retroperitonum
LMS17	Prim	59/F	LMS	LMS	Uterus
LMS18	Prim	46/F	LMS	LMS	Uterus
LMS21	Prim	31/F	LMS	LMS	Retroperitonum
LMS23	Prim	72/F	LMS	LMS	Thigh
LMS24	Prim	66/F	LMS	LMS	Perineum
LMS25*	Met	59/F	LMS	LMS	Arm
LMS28	Prim	82/M	MFH	LMS	Knee

GIST: Gastrointestinal stromal tumour
LMS: Leiomyosarcoma
MFH: Malignant fibrous histiocytoma
x: Xenograft
Prim: Primary tumour
Met: Metastasis
F: Female
M: Male
* Sample from the same patient. LMS1 primary tumour and LMS25 arm metastasis.

2.1.1 Isolation of genomic DNA

The DNA was isolated from most of the samples using the following protocol.

The equipment (mortar and pestle) was pre-cooled in liquid N_2. Frozen tissue was placed in the mortar with liquid N_2 and it was grinded to powder. The powder (in N_2) was transferred to a 50 ml tube, leaving the cap open until most of the N_2 had evaporated. The tube was immersed in N_2 to maintain low temperature.

4 ml Lysis buffer A was added to the powder and then it was placed at -70°C for 30 minutes. The suspension was thawed at 37°C, then 4 ml Lysis buffer B were and 100 μg/ml Proteinase K were added; diluting 1:100 from 10 mg/ml stock = 80 μl. The suspension was placed in the incubator at 37°C over night, with gentle rocking of the sample (Orbital Incubator, Gallenkamp).

Then, DNA was extracted from the suspension by using first a phenol-chloroform extraction. Phenol (~pH 8) was thawed, and an equal volume was added to the solution. The solution was mixed gently by inversion and later centrifuged (Sorvall® RC 5C Plus SS-34 rotor) for 7-8 minutes: 172 x g, at room temperature. The upper phase (containing the DNA) was transferred to a new 50 ml tube, avoiding the interphase. The same procedure was repeated one time, keeping the upper phase.

An equal volume of phenol-chloroform-isoamylalcohol was added to the solution (25:24:1) and mixed gently by inversion. The same procedure was repeated one time, keeping the upper phase.

An equal volume of chloroform-isoamylalcohol was added to the solution (24:1) and mixed gently by inversion. The upper phase was transferred to a 30 ml centrifuge tube, and then 1/10 of the volume of NaAc 3M pH 5.2 was added. 10 ml isopropanol was added and mixed gently, then another 10 ml, and again mixed gently. The solution was centrifuged for 40-60 minutes: 11951 x g at 4°C (Sorvall® RC 5C Plus SS-34 rotor).

The pellet was washed once in 1 ml 70% (v/v) ethanol, followed by a 5-minute spin at 20198 x g at 4°C (Sorvall® RC 5C Plus SS-34 rotor) and it was dissolved in a suitable amount of 1x

TE. The 1xTE volume depends on the size of the pellet: 0.15-0.2 ml at first, then increasing if necessary to dissolve more DNA. The sample was stored at 4°C, and it was later quantified by using a spectrophotometer or fluorometer (Picogreen® dsDNA Quantitation Kit, Molecular Probes)

2.2 Array Comparative Genomic Hybridisation

2.2.1 Theory

Conventional nucleic acid hybridisation is the pairing of complementary DNA strands to produce DNA-DNA hybrids (or DNA-RNA hybrids). If a double-stranded DNA is subjected to heat, the complementary strands will separate. When these single strands are cooled slowly down, they will reunite to form again a double-stranded molecule (See Magliano D 2001 for review)

In CGH, total genomic DNA is isolated from test (tumour DNA) and reference sample (normal diploid sample). The two DNA samples are differentially labelled, and hybridised to normal human metaphase chromosomes where DNA sequences from both sources will bind to different genomic locations to be distinguished (See Pinkel and Albertson 2005 for review) In that way, CGH detects and maps DNA copy-number differences throughout the genome (See figure 2.1)

The ratios of test and reference fluorescence along the chromosomes are quantified using digital image analysis. Gains and amplifications in the test DNA are identified as chromosomal regions with increased fluorescence ratios, whereas losses and deletions result in a reduced ratio. Ratios are normalized so that the modal value is 1.0 on a linear scale or 0.0 on a logarithmic scale.

Conventional CGH is unable to detect balanced chromosomal translocations, inversions and whole- genome ploidy changes. In addition, because of the limited resolution of metaphase chromosomes, alterations smaller than 5-10 Mb cannot be detected using conventional CGH (See Oostlander et al. 2004 for review) The necessity for higher resolution led to the development of microarray-based CGH. In microarrays, the target isn't metaphase chromosome but a large number of mapped clones spotted onto a glass slide. In this project,

slides containing 4549 genomic clones were used. The slides were provided by the Norwegian Microarray Consortium; they cover the whole genome at a resolution of 1 Mb.

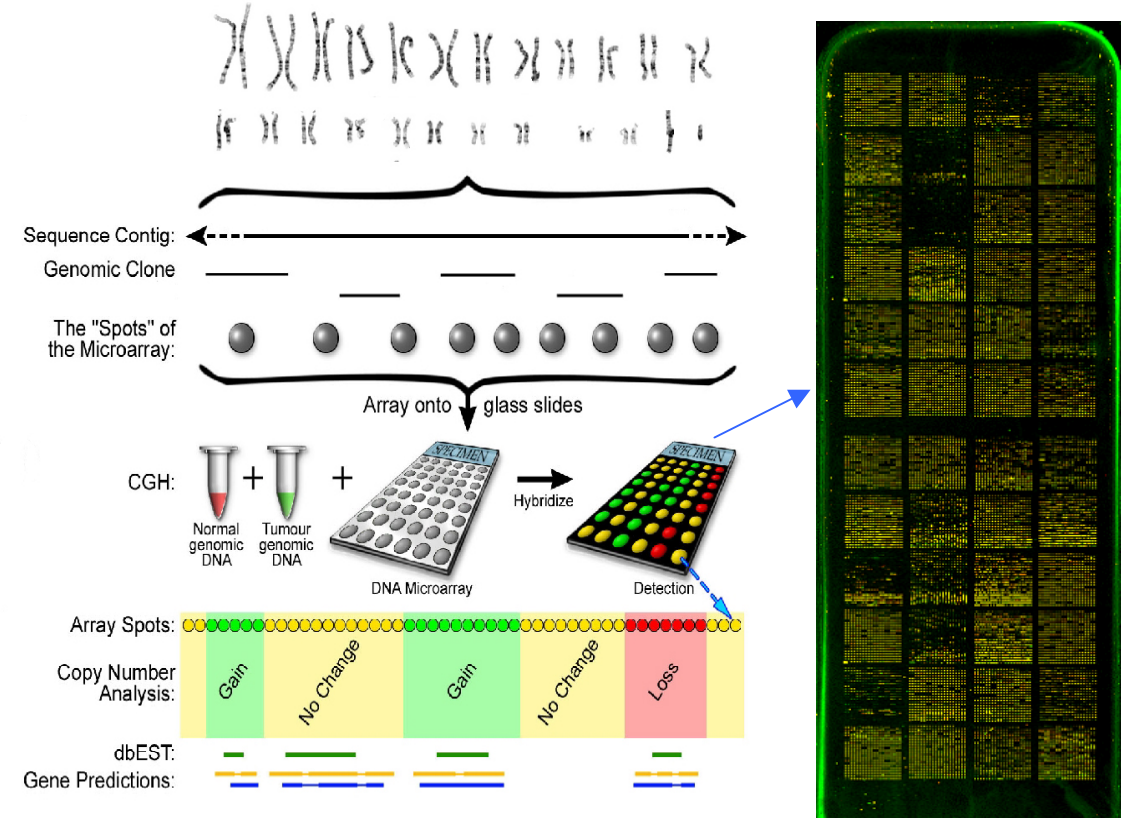

Figure 2.1 Overview of array-based comparative genomic hybridisation. Genomic DNA samples from sample and a control subject are individually labelled with fluorescent dyes to a DNA microarray consisting of genomic clones with known location. Relative levels of copy number changes are measured according to the fluorescence intensity for each probe. Green spots will represent gain, yellow no change and red loss in DNA copy number.

The clones used in the genomic microarrays are artificial chromosomes (bacterial and P1) provided by the Wellcome Trust Sanger Institute. The BAC clones used belonged to the RPCI-11 library, whereas the PAC clones were from the RPCI-1, -3, -4 and –5 libraries. The location of each BAC and PAC was based on sequence alignment from the search tool BLAST (http://www.ebi.ac.uk/blast/) from the European Bioinformatics Institute.

Additional RPCI-11 clones belonging to chromosome 1, 12, 17, X and Y from the Cancer Chromosome Aberration Project (CCAP) (Kirsch and Ried 2000) and the VGC mapped BAC library (Cheung et al. 1999) were incorporated, as well as 575 Caltech clones from the OncoBAC clone collection. The clones were arrayed in quadruplicate onto amine-binding slides (CodeLink, Amersham) using a MicroGrid II arrayer (BioRobotics).

2.2.2 Labelling and setup for array CGH

Digestion of genomic DNA

Each sample was prepared as follows, using a 1.5-ml microcentrifuge tube:

Component		Amount
1 µg genomic DNA (Sample/Reference)		1 µg
10x Dpn II buffer	(100 mM NaCl, 50 mM Bis Tris-HCl, 10 mM MgCl2, 1 mM dithiothreitol New England Biolabs®)	5 µl
Dpn II	(20 U/ µl; New England Biolabs®)	3 µl
Sterile water		to 50 µl

The samples incubated at 37°C overnight, and then it follows a purification step by using the QIAquick® PCR purification kit (QIAGEN®)

Labelling of genomic DNA

Typical array CGH procedures use 300 ng to 3 µg of test DNA. In this project, 0.5 µg DNA was used in the labelling reaction; a random primer labelling was followed with the purpose of amplifying DNA for the hybridisation (BioPrime® Array CGH Genomic Labelling System, Invitrogen®). Each sample was prepared as follows, using a 1.5 ml microcentrifuge tube:

Component		Amount
0.5 µg digested DNA (Sample/Reference)		0.5 µg
2.5x Random Primer Mix	(125 mM Tris-HCl (pH 6.8), 12.5 mM $MgCl_2$, 25 mM 2-mercaptoethanol, 750 µg/ml oligodeoxyribonucleotide primers (random octamers) Invitrogen®)	40 µl
Sterile water		to 46.5 µl.

The samples were denatured for 10 min at 100°C, placed in ice/water and spun down. Then, 10 µl 10x dCTP Nucleotide Mix (containing 1.2 mM dATP, dGTP and dTTP and 0.6 mM dCTP in 10 mM Tris (pH 8.0), 1mM EDTA; Invitrogen®) were added followed by 1.5 µl 1.0

mM Cy3™ (Sample) or 1.5 µl 1.0 mM Cy5™ (Reference) –dCTP (PerkinElmer) and 2 µl Exo-Klenow Fragment (40 U/ µl Invitrogen). The samples were incubated at 37°C overnight.

The reaction was stopped by adding 10 µl Stop Buffer (0.5 M EDTA pH 8.0) to each tube, then vortexed and spun down. Then, it follows a purification step by using the MicroSpin™ G-50 Columns (Amersham Pharmacia Biotech)

Probe preparation

Unlabeled competitor (Cot-1) DNA was included to suppress the hybridisation of the repetitive sequences in the genomes, so that the unique sequences could be measured. The fluorescent nucleotides are sensitive to photobleaching so it was very important to minimize exposure of the fluorescent nucleotides and labelled DNA to light.

Each sample was prepared as follows, using a 1.5 microcentrifuge tube:

Component	Amount
Cy3™ labelled DNA (Sample)	~100 µl
Cy5™ labelled DNA (Reference)	~100 µl
Cot-1 DNA (1 µg/µl; Invitrogen™)	135 µl
3M NaAc pH 5.2	37 µl
96% Ethanol (-20°C)	850 µl

The solution is placed at -80°C for 30-40 min, and then spun at 20985 x g for 30 min (4°C; IEC Micromax RF) The supernatant is discarded, letting the pellet air-drying for about 5 min and 84 µl of MMI solution is added. Let it stand for 10 min at RT.

The solution was later vortexed and placed at 70°C for 2-3 min to facilitate dissolving. After that, 24 µl 20% SDS and 4 µl yeast tRNA (100 µg/ µl; Invitrogen™) are added to the solution. The solution was placed at 72°C for 10 min (denaturing) and then at 37°C for 1h. The solution was applied to the hybridization chamber.

Hybridization and wash

In this study, there was used a Hybarray™ (PerkinElmer) hybridization station. This system automates hybridization and post-hybridization washes.

The hybridization was performed at 37°C with agitation for 42-46 hours. After hybridisation, the station washed the slides with three different solutions; First a solution of 50% formamide, 2X SSC at 48°C; 2X SSC/0.1% SDS at 48°C; and finally a PN buffer (0.1mol/l NaH_2PO_4/Na_2HPO_4 pH 8, 0.1% NP-40) at 25°C. For all three solutions, the hybridisation station washed for five cycles; each cycle has a flow time of 20 sec and a hold time of 40 min.

Scanning and analysis

The slides were taken out of the station and rinsed briefly manually 2 times in 0.05X SSC. The slides were spun at 1000 rpm for 3 min before being scanned using the Agilent G2565BA DNA microarray scanner (Agilent Technologies). The software GenePix Pro 6.0 was used to analyse the data (Axon Laboratories); Spots were automatically segmented and manually adjusted when necessary. Local background was subtracted, and the fluorescent intensities and ratio of the two dyes were calculated for each spot.

GenePix files were exported to M-CGH, a MATLAB toolbox designed to filter, normalise and visualise microarray data (Wang et al. 2004). Empty and manually flagged spots, and with intensities lower than the background in both channels as well as net signal intensities below local background plus twice the standard deviation of the background were excluded.

Log_2-transformed ratios were normalised using a global intensity depended algorithm (LOWESS) (Cleveland et al. 1976) and then they were combined in a text file. Calculating the mean and the standard deviation of the quadruplicated spots in the array assesses the quality of the spot reproducibility. The clones with standard deviation larger than 0.2 and ratios based on a single measurement were eliminated. The mean ratios of the replicate spots were exported to a text file (Wang et al. 2004).

Missing values were imputed via a K-nearest Neighbour algorithm normalisation using Significance Analysis of Microarrays (SAM) (Tusher et al. 2001). All samples were clustered

by J-Express (Dysvik and Jonassen 2001) with average linkage (WPGMA) as the cluster method and Pearson correlation as the metric distance.

The identification of significant copy number changes was performed by CGH-Explorer v. 2.52 (Lingjaerde et al. 2005). Analysis of copy number errors (ACE) was performed using a false discovery rate of 0.0000 in LMS and GIST separately. Chromosomal regions showing gains or losses in at least four of 13 LMSs (30%) and three of seven GISTs (42%) were considered for identifying minimal recurrent regions of alteration. Because information about the sex of some of the patients was missing before the array CGH was done, only autosomes (non-sex chromosomes) were taken into account in this study.

To identify chromosomal regions differing significantly in DNA copy number between LMSs and GISTs, a t-test analysis was performed using SAM. A list of genomic clones showing differences in copy number between the two groups was generated.

2.3 Artificial Chromosomes

Artificial chromosomes are DNA molecules assembled in vitro from defined constituents, capable of accepting selected fragments DNA, and replicating the resulting hybrid when it is introduced into living cells. They also guarantee the stable maintenance of large DNA fragments with the properties of natural chromosomes, because they have a lower recombination frequency and susceptibility to DNA shearing forces (See Roosen G 2002 for review).

In this project, the Bacterial Artificial Chromosomes (BACs) and P1 phage-derived Artificial Chromosomes (PACs) were used in FISH and array CGH slides. BACs are vectors based on the *E. coli* fertility plasmid (F factor), which is normally present at one to two copies per cell. This is essential for the stability of the cloned inserted. PACs are hybrids of P1 bacteriophage and BAC vectors. They are also of low copy number.

2.3.1 Isolation of BAC clones

A bacterial colony was inoculated to 8 ml 2X TY-medium pH 7.4 supplemented with 20 ug/ml Chloramphenicol in a 15 ml snap-cap tube. The tube was incubated overnight shaking

at 300 rpm at 37°C (GFL Shaking Incubator 3031, GFL). The tube was centrifuged at 1076 x g for 10 min (Sorvall® RC 5C Plus SS-34 rotor).

The supernatant was discharged, and the pellet resuspended in 0.3 P1 solution. The solution was transferred to an Eppendorf tube, and 0.3 ml P2 solution was added. The tube was gently mixed and was left at room temperature for 5 min. The appearance of the suspension changed from very turbid to almost translucent.

Slowly, 0.3 ml P3 solution was added to the tube. A thick white precipitate of the protein and *E. Coli* DNA was formed. The tube was left on ice for 5 min, and then centrifuged at 9327 x g for 10 min (4°C, IEC Micromax RF). The supernatant was transferred to a new Eppendorf tube containing 0.8 mL ice-cold isopropanol, avoiding to take any white precipitated material.

The tube was mixed by inversion a few times and again placed on ice for 5 min. The tube was centrifugated at 9327 x g for 15 min (4°C, IEC Micromax RF) The supernatant was removed and 0,5 ml 70% EtOH was added. The tube was inverted to wash the DNA pellet, and was centrifugated for 5 min (4°C, IEC Micromax RF)

The supernatant was removed and the pellet was air-dried at room temperature until it turned from white to translucent. The pellet was resuspended overnight in 60 ul 1X TE. The concentration was measured by fluorescence (PicoGreen® dsDNA Quantitation Reagent, Molecular Probes) and the purity by gel electrophoresis.

2.4 Fluorescence in Situ Hybridisation

FISH is a protocol used to detect specific nucleic acid sequences directly on the chromosomes of a karyotype. This technique is based in sequence-specific annealing of denatured nucleic acid strands; fluorescently labelled probes and the target chromosomes are denatured making the DNA single stranded. Complementary sequences in the probe will reanneal, and stringent washes will remove non-specifically bound probe. After that, the fluorescent signal can be observed at the site of hybridisation in a fluorescent microscope. The method is described in figure 2.2 (See Gole 2001 for review).

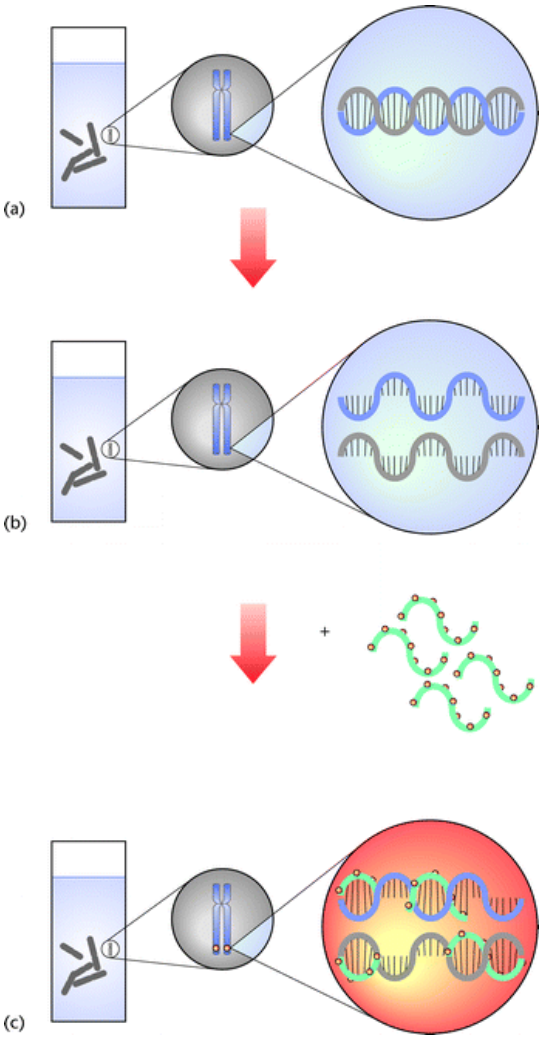

Figure 2.2 Principles of Fluorescence *in Situ* Hybridisation. Fluorescently labelled probes and the target chromosomes are denatured making the DNA single stranded. Complementary sequences in the probe will reanneal and the fluorescent signal can be detected (Figure from Gole 2001).

FISH has been used for many different purposes; one of them is the analysis of interphase cells for detection of numerical anomalies. A hybridisation control is usually done on metaphase cells: If the probes give a single signal at the expected chromosomal position, the interphase cells can be tested with the same hybridisation conditions, and will give specific signals according to how many copies there are in the cell nucleus (See KH A Choo 2001 for review). In cancer cytogenetics, amplification and deletion of genes can be visualized by FISH. A normal probe is seen as two signals in a nucleus. More signals means aneuploidy or amplification of the targeted region e.g an oncogene.

2.4.1 FISH Procedure

Sample Slide Preparation

Interphase slides from the samples LMS1,-5x,-10 and -25 were prepared.

The equipment (mortar and pestle) was pre-cooled in liquid N_2. The tissue was placed in the mortar with liquid N_2, and was grinded into powder. The powder (in N_2) was transferred to a 15 ml tube, leaving the cap open until most of the N_2 had evaporated. The tube was immersed in N_2 to maintain low temperature.

A solution of 1:3 acid acetic/methanol solution was prepared and 20 ml were added to the tube, and mixed by inversion. The tube was centrifuged 10 min at 172 x g at room temperature (Sorvall® RC 5C Plus SS-34 rotor). The supernatant was removed, and the pellet was resuspended in 60% acid acetic. The amount depending on the pellet size.

The solution was dripped (two drops) to pre-warmed glass slides (Superfrost Color, Menzel; 50°C) and dried at the same temperature. The slides were stored at -20°C.

Probe labelling

Each clone (RP11-12H18,-471L13,-219A15 and 121A13) was prepared as follows, using a 0.5-ml GeneAmp (Applied Biosystems) tube as follows:

Component	Amount
1 µg BAC DNA	1 µg
10x dNTP mix (0.2 mM each dCTP, dGTP, dTTP; 0.1 mM dATP; 0.1 mM biotin-14-dATP 500 mM Tris-HCl (pH 7.8); 50 mM $MgCl_2$; 100 mM β-mercaptoethanol 100 µg/ml nuclease-free BSA, Invitrogen)	5 µl
Enzyme mix (0.5 U/µl DNA Polymerase I; 0.007 U/µl DNase I; 50 mM Tris-HCl (pH 7.5) 5 mM magnesium chloride; 0.1 mM phenylmethylsulfonyl fluoride; 50% (v/v) glycerol; 100 µg/ml nuclease-free BSA, Invitrogen)	3 µl
Sterile water	to 50 µl

The tube was placed in a PCR-machine (Programmable Thermal Controller, MJ Research) in a programme of 16°C for 90 min, 70°C for 10 min and finally at 0°C. The concentration of the sample should be 20 ng/ µl. 5 µl DNA product was separated in a 1% agarose gel (90 V Power PAC 300, Bio Rad) stained with ethidium bromide (10 mg/ml). A photo of the gel was taken by Gel Doc 1000 Bio Rad.

Slide Preparation

The slides were placed in 70% Ethanol for about one hour at 4°C and then air-dried. The metaphase slides were also put in a 1:3 acetic acid /methanol solution for about half hour and then air-dried. A solution of 0.4 mg/ml pepsin in 0.01 M HCl was prewarmed to 37°C, and 1 ml was added to the slide. The slide was incubated for 10 min at 37°C in a humid atmosphere, and later washed three times in 1xPBS (Phosphate-buffer saline) for 5 min at RT.

The slides were incubated for 10 min at RT in a 1% formaldehyde/1%PBS solution, washed for 5 min in 1XPBS, dehydrated in ethanol series; 70, 90, 96 and 100% ethanol about 1 min each and air-dried.

Then, 100 µl DF (70% deionized formamide/2XSSC) was added to the slide and then covered with a cover slip. The slide was denatured at 70°C; metaphase: for 1.5 min and interphase for 2 min and put in cold 70% ethanol. The slide was again dehydrated in ethanol series and air-dried.

Hybridisation

Each probe (RP11-12H18, -471L13, -219A15 and 121A13) was prepared as follows, using a 1.5 microcentrifuge tube:

Component	Amount
Labelled DNA probe	200-400 ng
Cot-1 DNA (1 µg/µl; Invitrogen™)	10 µl
3M NaAc pH 5.2	1/10x of the volume
100% Ethanol (-20°C)	2.5x of the volume

The solution was placed at -80°C for 30-40 min, and then spun at 20986 x g for 30 min (4°C; IEC Micromax RF). The supernatant is discarded, letting the pellet air-dry for about 5 min. Later, 16 µl of MMI solution (Dextran sulphate, formamide, SSC) is added, and let stand for 10 min. The solution was denatured 10 min at 80°C, then 3 min on ice/water and finally at 37°C for about half an hour.

The solution was added to the slide and covered with a cover slip. The cover slip was sealed with rubber cement, and was placed at 37°C in a humid atmosphere overnight.

Detection

The following morning, the rubber cement was removed, and the slide washed in a pre-warmed 50% formamide/2XSSC solution for 5 min, 3 times; then 0.1XSSC for 5 min, 3 times; and in the end, it was washed in TNT solution (Tris HCl, NaCl, Tween).

In FISH, the probes bound to a target can be detected by two manners commonly used – indirect or direct labelling. For indirect labelling, probes are labelled with modified nucleotides that contain a binding molecule with affinity for a protein receptor, whereas direct labelling uses the incorporation of nucleotides that contain directly a bound fluorophore. In this project, the ligand incorporated to the probes was biotin. The Cy3-Avidin/TNB antibody (Tris HCl, NaCl, blocking reagent) binds to the biotin conjugated nucleotides. DAPI (4.6-diamino-2-phenylindole) was used for fluorescent counterstaining.

Blocking solution (TNB, 100 µl) was added to the slide and covered with a cover slip. The slide incubated at RT for 30 min. A solution of 1:50 Cy3-Avidin/TNB antibody was made and centrifugated for 3 min at 15762 x g (4°C; IEC Micromax RF). Of this solution 100 µl were added to the slide and covered with a cover slip. The slide incubated 30 min at 37°C in a humid atmosphere. The slide was washed in TNT 3 times for 5 min, dehydrated with ethanol series and air-dried. Later, 19 µl Vectashield® (DAPI, Vector Laboratories) was added to the slide. The slide was stored at 4°C.

Fluorescent signals were detected by a fluorescence microscope (Axioskop, Zeiss). The microscope has selective filters, for which different fluorochromes can be excited and

observed. The DAPI filter was used to localize the nucleus, while the DAPI/Cy3 filter was used to identify the probe signals in the nucleus (Cy3-Avidin antibody signals).

3. Results

3.1 Array Comparative Genomic Hybridisation

3.1.1 Hierarchical clustering of tumours

Initially 19 tumour samples diagnosed as LMSs were analysed using array CGH. GenePix Pro 6.0 was used to analyse the data (Axon Laboratories). GenePix files were later exported to M-CGH (MATLAB), where data was filtered and normalised. Log_2-transformed ratios were normalised using a global intensity algorithm and they were combined in a text file. The samples were later hierarchical clustered by J-Express.

Two well-defined main clusters were identified by clustering. After pathological revision of samples, it was revealed that all samples within one of these main clusters were re-classified to GISTs. All the samples in the other main cluster were LMSs, including one sample previously classified as MFH (Now LMS28) added to the study (See figure 3.1)

The LMS cluster was further divided into two subclusters. The only noteworthy difference between the subclusters was the anatomic location of the tumour. All LMSs of uterine origin clustered together, along with one LMS of retroperitoneal origin (Figure 3.1B). LMS1 and LMS25, primary and metastasis sample from the same patient, were the most closely related samples of the tumour panel.

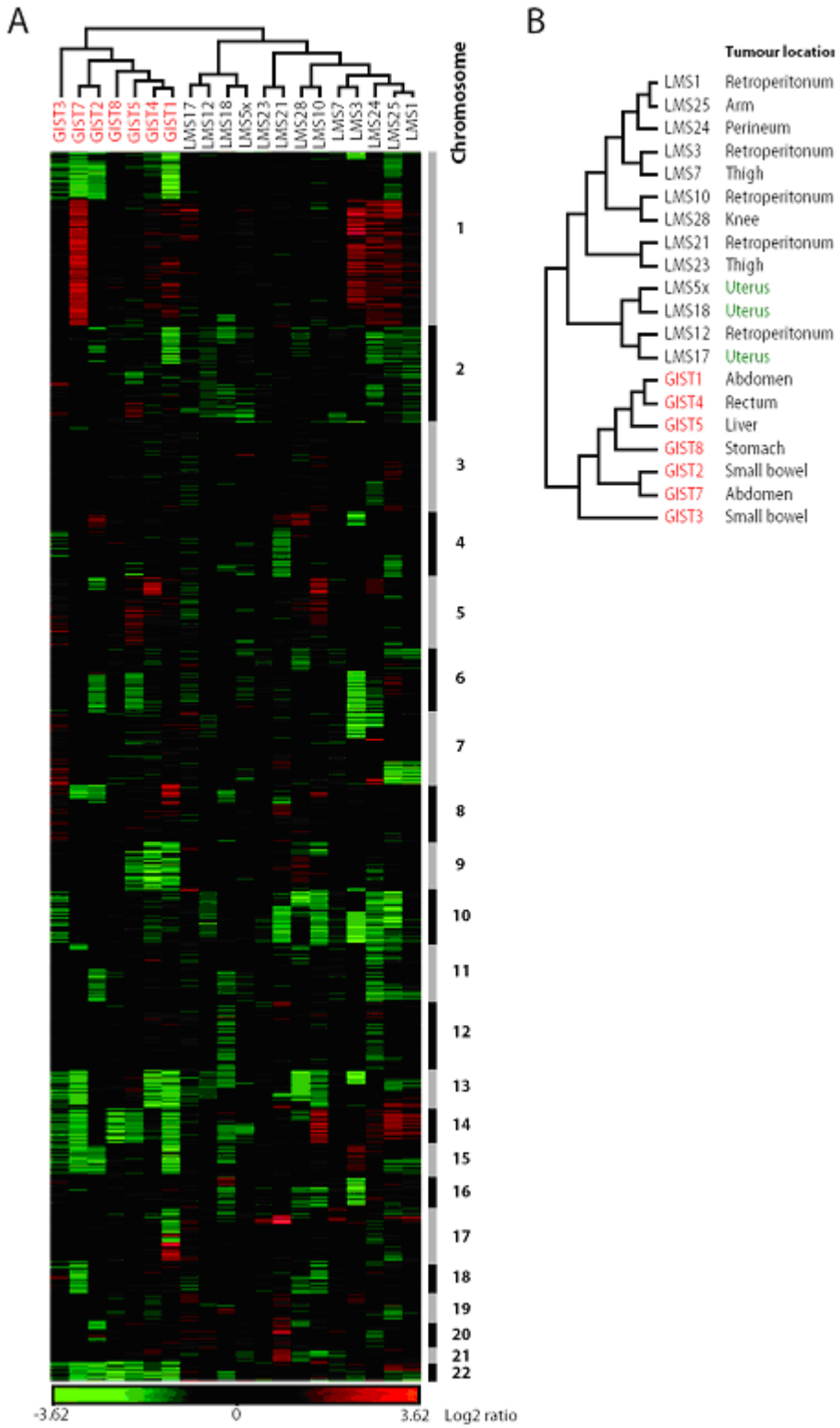

Figure 3.1 Hierarchical clustering of 20 human sarcomas by J-Express. Gain and loss of DNA copy number is seen as red and green respectively. **A** Two sample groups (LMS and GIST) are easily visualized in this cluster. **B** Tumour location in all samples

3.1.2 Genetic Alterations in LMS

CGH-Explorer was used for the statistical identification of gained and deleted region in all sarcoma samples. This program is used for visualization and statistical analysis of array CGH data. The algorithm, analysis of copy number errors (ACE) in CGH-Explorer was used for detecting copy number errors in the data set originated from GenePix and M-CGH. This algorithm also computes false discovery rates (FDR) of the data set. This means that the program calculates different rates of false significant values and shows how it could possibly affect the results. In this project, the FDR was restricted to 0.0000.

Analysis by ACE showed many genetic alterations in the LMS samples. Identification of minimal recurrent regions of alteration was done by considering gain or losses only when present in at least four of the 13 LMS samples (30%). There were 30 minimal recurrent regions with both gains and/or losses. From the 30 regions, 19 showed losses in DNA copy number while 15 showed gain in copy number. The most frequent losses were observed in 10q21.3 and 13q14.2-q14.3 whereas most frequent gains in region 17p13.1-p11.2. Results are summarized in Table 3.1.

Table 3.1 Minimal recurrent regions altered in leiomyosarcomas (n=13)

Cytoband	Aberration	Start clone	End clone	Size [Mb]	Frequency
1p36.32-p36.21	Loss	RP1-37J18	RP4-636F13	7.9	5/13
1q21.1-q23.2	Gain	RP11-277L2	RP11-550P17	9.9	6/13
1q23.2-q23.3	Gain	RP11-517F10	RP11-404F10	0.5	5/13
1q23.3-25.1	Gain	RP11-572K18	RP5-1198E17	11.3	6/13
2p25.1-p21	Loss	RP11-83M8	RP11-27C22	35.9	8/13
2p14-p13.1	Loss	RP11-263L17	RP11-1P9	8.8	7/13
2q24.1q31.2	Loss	RP11-552E1	RP11-250N10	19.1	7/13
2q37.1-q37.2	Loss	RP11-52C8	RP11-556H17	9.4	7/13
3p12.3-p12.1	Gain	RP11-16M12	RP11-447J13	7.9	6/13
4q31.3-qtel	Loss	RP11-259G7	CTC-963K6	36.2	5/13
5p13.2-pcen	Gain	CTD-2291F22	RP11-269M20	14.7	6/13
6p25.2-p22.3	Loss	RP1-136B1	RP11-289M23	21.2	6/13
6q14.1-q23.3	Loss	RP11-173D14	RP11-95M15	58.6	4/13
7p22.3-p13	Loss	RP11-510K8	RP4-647J21	43.4	4/13
7q31.33-qtel	Loss	RP5-902E20	CTB-3K23	35.0	4/13
9q21.13-q31.3	Gain	RP11-563H8	RP11-202G18	39.8	5/13
10q21.3	Loss	RP11-161L14	RP11-778O10	0.8	10/13
11p15.5-p15.4	Loss	RP11-295K3	RP11-438N5	2.4	5/13
11q22.1-q24.1	Loss	RP11-49M9	RP11-166D19	21.9	6/13
13q14.2-q14.3	Loss	RP11-305D15	RP11-40A8	2.7	10/13

14q12-q21.1	Gain	RP11-30H9	RP11-138H18	9.1	6/13
14q21.2-21.3	Gain	RP11-565J15	RP11-58E21	3.6	6/13
14q31.3-q32.2	Gain	RP11-300J18	RP11-76E12	9.7	6/13
15q11.2-q12	Gain	RP11-289D12	RP11-446P9	3.5	5/13
16q21.2-q22.1	Loss	RP11-452G23	RP11-354N7	20.2	6/13
17p13.2-p13.1	Loss	RP11-243K12	RP11-186B7	1.4	7/13
17p11.2	Gain/Amp	RP11-524F11	RP1-162E17	1.9	7/13
18q11.2-qtel	Loss	RP11-535A5	CTC-964M9	57.9	4/13
21q21.1-q22.11	Loss	RP1-152M24	RP1-245P17	18.4	5/13
22q13.1-q13.33	Loss	CTA-228A9	CTA-722E9	11.4	6/13

Some alterations involved a whole chromosome arm; the whole q arm of chromosome 1 was gained in four of the samples and three minimal recurrent gained regions were identified, 1q21.1-q23.2 (9.9 Mb) in 6 samples; 1q23.2-q23.3 (0.5 Mb) in 5 samples; and 1q23.3-1q25.1 (11.3 Mb) in 6 samples. Gain was also common in the q arm of chromosome 14, involving the regions 14q12-q21.1 (9.1 Mb), 14q21.2-q21.3 (3.6), 14q31.3-q32.3 (9.7 Mb). Chromosome 18 was also target for aberrations of the whole q arm; the region 18q11.2-qtel (57.9 Mb) was deleted within four samples out of 13.

Chromosome 2 was also a frequent target for deletion. Alterations of at least one segment in this chromosome were seen in 10 of the samples; four minimal recurrent regions were found in the p arm as well as two regions in the q arm. Other regions targeted by deletions were 4q31.3-qtel (36.2 Mb), 6q14.1-q23.3 (21.2 Mb), 7p22.3-p13 (43.3 Mb), 7q31.33-qtel (35.0 Mb), 13q14.2-q14.3 (2.7 Mb), 16q21.2-q22.1 (20.2 Mb), 17p13.2-p13.1 (1.4 Mb), 18q11.2-qtel (57.9 Mb), 21q21.1-q22.11 (18.4 Mb) and 22q13.1-q13.33 (11.4), all showing a frequency of at least four out of 13 samples.

Although deletions were more common than gains in LMSs, some regions, in addition to those above, showed increased DNA copy number. The regions 3p12.3-p12.1 (7.9 Mb), 5p13.2-pcen (14.7 Mb), 9q21.13-q31.3 (39.8 Mb), 12p11.22-p11.21 (2.3 Mb), 15q11.1-q12 (3.5 Mb), 15q25.1-q26.3 (21.2 Mb) and 20q11.21-q13.33 (32.1 Mb) were gained in at least five out of 13 samples. A frequency plot of gains and losses for LMSs is shown in Figure 3.2, as well as a representative copy number profile for this type of tumours (LMS23).

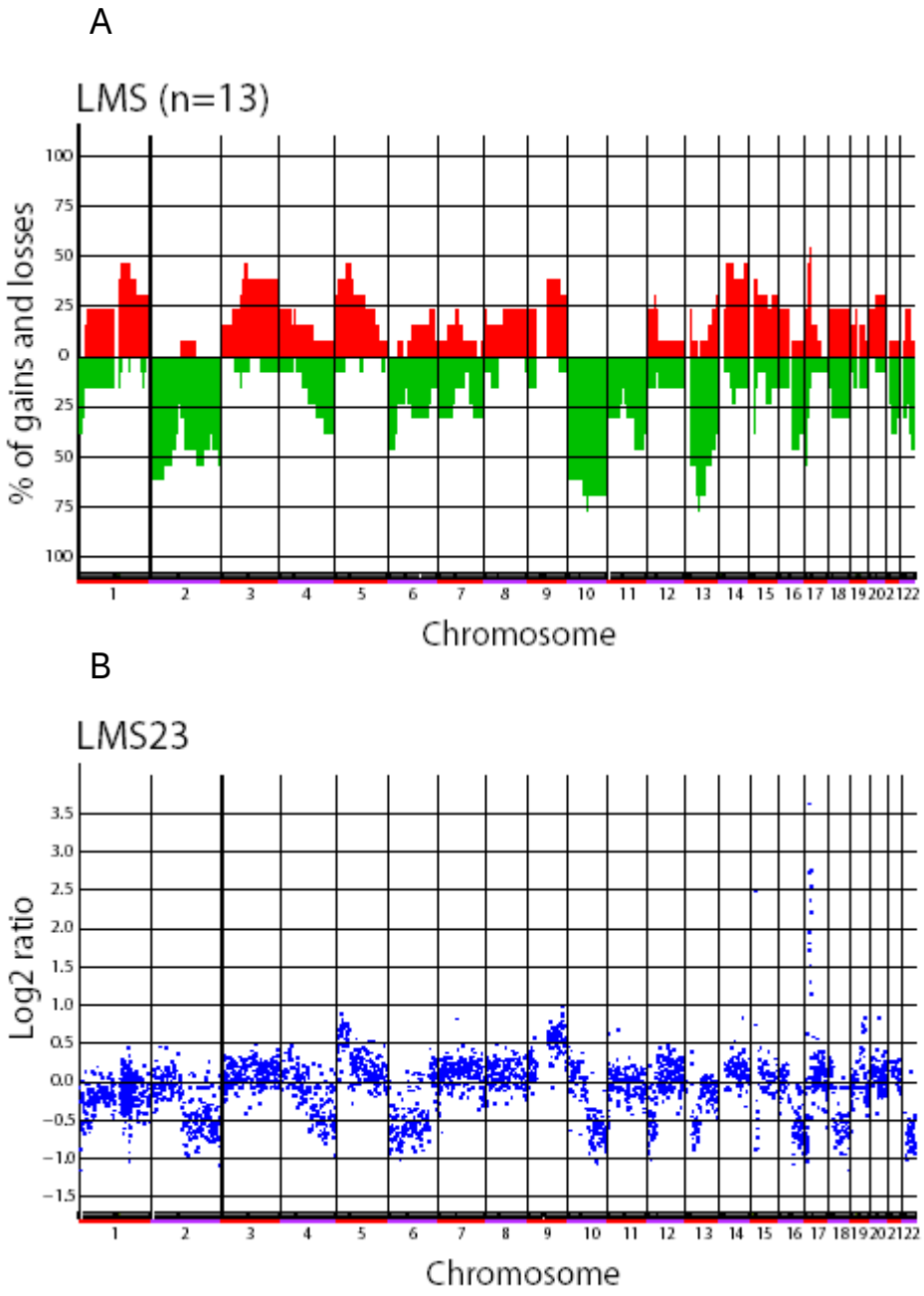

Figure 3.2 CGH Explorer plots showing **A** ACE plot of gains (red) and losses (green) from chromosome 1 to 22 of all LMS samples. The plot shows the frequency of gained and lost clones from the array CGH data **B** Copy number (Log_2 ratio) profile for LMS23 from chromosome 1 to 22.

Chromosome 17 showed regions with loss, but also high levels of amplification. The region 17p13.2-p13.1 showed in many LMS samples (7/13) loss of DNA copy number, although the region 17p13.1-p11.2 showed most frequent gains in the sample panel (7/13). The minimal recurrent region of gain was limited to 1.9 Mb in 17p11.2, where three samples showed high-level amplification (See fig. 3.3)

The region 17p13.1-p11.2 covered 12.7 Mb, and was represented by 23 BACs and PACs, starting with RP11-404G1 and ending with RP11-121A13. In order to narrow down the list of candidate target genes for this amplification, the array CGH results were compared to the expression levels in six of the LMSs analysed, and one additional sample (LMS29) (Data not shown. Francis, Namløs, Myklebost, Nilbert *et al*., unpublished). According to Ensembl, 172 genes are located within the amplified region.

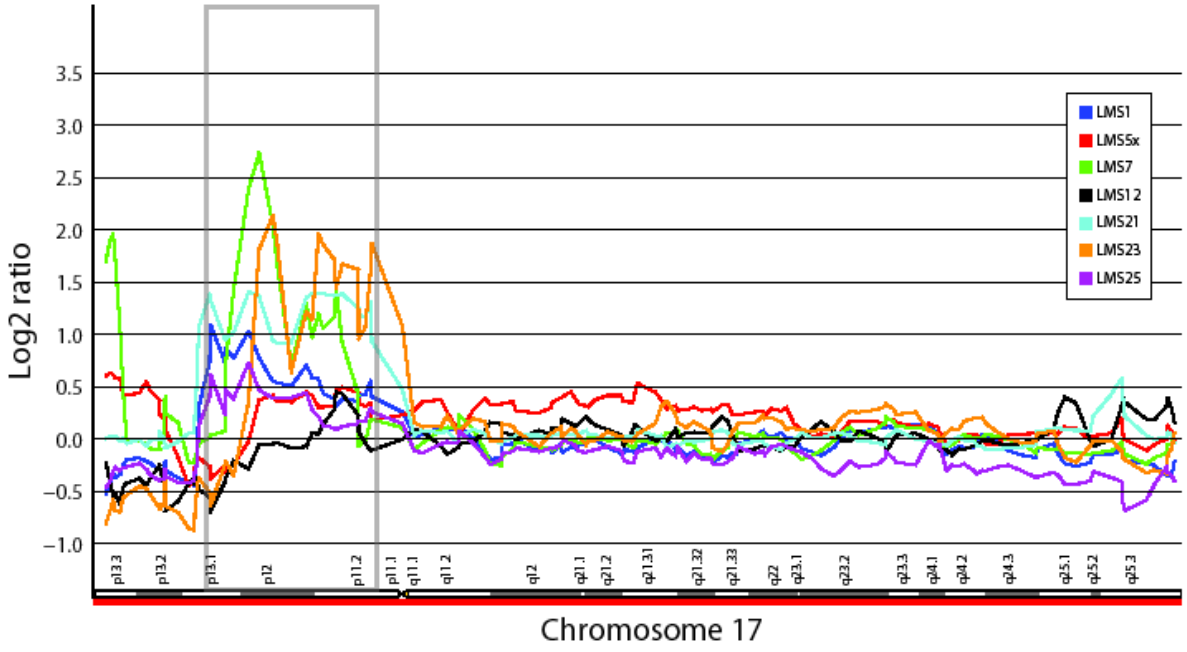

Figure 3.3 shows the copy number profile along chromosome 17 for the seven LMS samples. The region 17p13.1-p11.2 is selected within the square and showed increased copy number. The region 17p13.3-p13.1 showed loss of copy number, also visualized in this figure.

Nine genes located within the amplified region showed increased expression (log_2 ratio >1) relative to the soft tissue sarcoma median in two or more of the seven LMSs analysed. The expression level of these genes is shown in Figure 3.4. Microfibrillar-associated protein 4 (*MFAP4*) was over-expressed in four LMSs, whereas aurora kinase B (*AURKB*) and sterol regulatory element binding transcription factor 1 (*SREBF1*) were over-expressed in three LMSs. Aldehyde dehydrogenase 3 family member A2 (*ALDH3A2*), mitogen-activated protein kinase 7 (*MAPK7*) and serine hydroxymethyltransferase 1 (soluble) (*SHMT1*) were all over-expressed in two LMSs. In addition, three genes with no known function showed increased expression; *FLJ10847* in three LMSs and *LOC201158* and *LOC220594* in two LMSs.

SREBF1 showed the highest level of expression, being more than 16-fold higher expressed in LMS1 and -3 compared to the soft-tissue sarcoma median. Six of the genes; *SREBF1*, *SHMT1*, *LOC220594*, *MAPK7*, *MFAP4* and *FLJ10847* are located within the minimal recurrent region of amplification in 17p11.2 identified after ACE analysis.

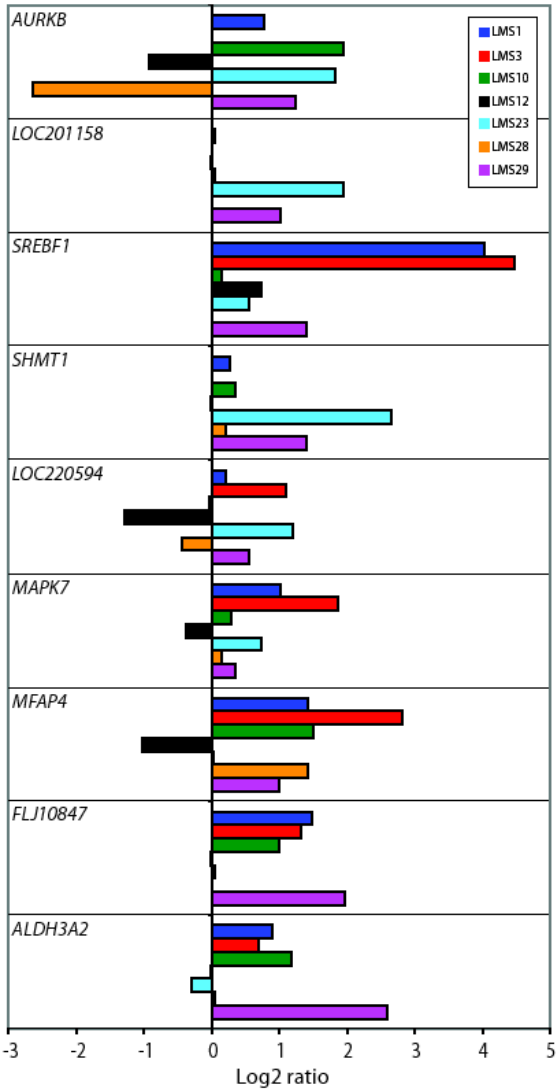

Figure 3.4 Levels of gene expression for seven LMS samples (Francis, Namløs, Myklebost, Nilbert *et al.*, unpublished)

Analysis of one primary tumour and its metastasis

After pathological revision (See table 1.3) it was revealed that LMS25 was a metastasis of a primary retroperitoneal tumour LMS1, meaning the two samples came from the same patient. The samples clustered together in J-Express (See figure 3.1A) and when compared by CGH-Explorer, they showed very similar copy number profiles except for a few regions. (See figure 3.5)

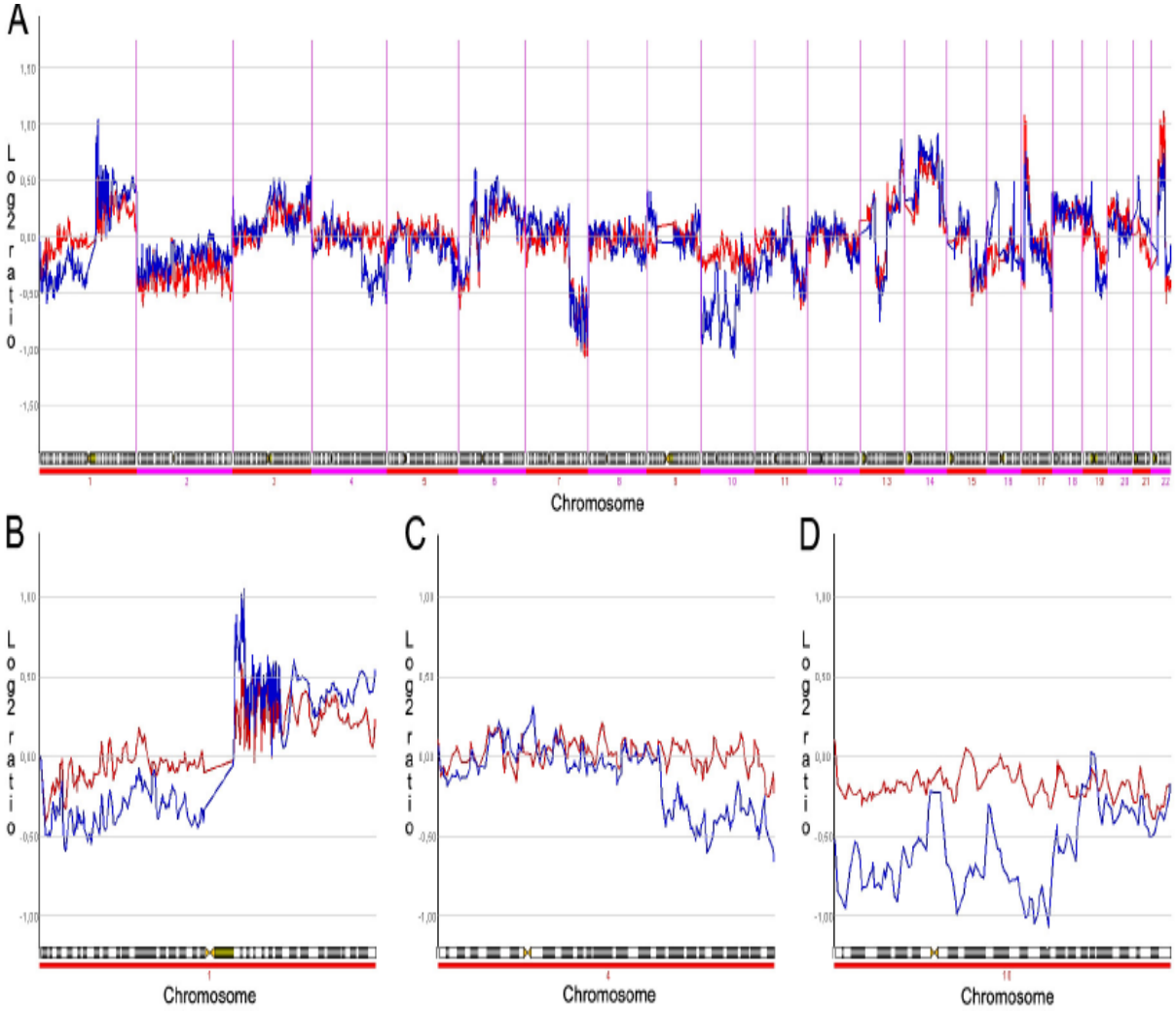

Figure 3.5 CGH Explorer log_2 ratio plots showing **A** All autosomes for a primary tumour (LMS1) and its metastasis (LMS25) visualized in red and blue respectively. Also detailed plots of some of the altered chromosomes differentiating LMS1 and LMS25 **B** In Chromosome 1, LMS25 presented loss of large parts of the p arm **C** In Chromosome 4 a region of 63.77 Mb is loss in LMS25 **D** Regions of chromosome 10 were also lost in the same sample.

The ACE algorithm was used to identify regions differing from LMS1 and LMS25. Loss of DNA was seen in the p arm of chromosome 1. The regions 4q28.1-qtel (63.77 Mb); 10q21.2-21.3 (17.56 Mb); 11p15.5-p12 (42.49 Mb); and 17q12-qtel (50.09 Mb) were deleted in LMS25 although these regions were not altered in LMS1. 13q12.11-q13.3 (19.47 Mb) was gained in LMS1 but normal in LMS25.

In chromosome 16 and 19, both gains and losses of DNA were identified. The regions 16p13.11-p11.2 (16.66 Mb) and 19p13.3-p12 (23.55 Mb) were gained in LMS25, while the regions 16q22.2-qtel (17.88 Mb) and 19q12-q13.43 (30.88 Mb) were deleted in the same

sample. These results are summarized in Table 3.2. Note that the regions listed are those differentiating LMS1 from LMS25 but the type of aberration is listed in relation to LMS25.

Table 3.2 Aberrations differentiating LMS25 from LMS1 from CGH-Explorer (ACE threshold)

Cytoband	Aberration	Size [Mb]	Start clone	End clone
1ptel-pcen	Loss	139.57 Mb	RP4-785P20	RP3-365I19
4q28.1-qtel	Loss	63.77 Mb	RP11-282B13	CTC-963K6
6p21.32-p21.1	Gain	13.37 Mb	RP5-1077I5	RP11-546O15
9p24.3-p21.3	Gain	21.82 Mb	GS1-77L23	RP11-149I2
10q21.2-21.3	Loss	17.56 Mb	RP11-532F4	RP11-314J18
11p15.5-p12	Loss	42.49 Mb	RP11-496I9	RP11-108L12
13q12.11-q13.3	Loss	19.47 Mb	RP11-76K19	RP11-131F1
16p13.11-p11.2	Gain	16.66 Mb	RP11-82O18	RP11-388M20
16q22.2-qtel	Loss	17.88 Mb	RP11-70E3	RP4-597G12
17q12-qtel	Loss	50.09 Mb	RP1-29G21	RP11-567O16
19p13.3-p12	Gain	23.55 Mb	CTD-3113P16	RP11-359H18
19q12-q13.43	Loss	30.88 Mb	CTC-459F4	GS1-1129C9

3.1.3 Genetic Alterations in GIST

The most frequent aberration observed was loss of the whole or parts of chromosome 22, seen in all tumours with a minimal recurrent region in 22q12.2-q13.31 (17.8 Mb). Minimal recurrent regions are summarized in Table 3.3.

Table 3.3. Minimal recurrent regions altered in gastrointestinal stromal tumours (n=7)

Cytoband	Aberration	Start clone	End clone	Size [Mb]	Frequency
1p36.32-p13.1	Loss	RP4-785P20	RP11-27K13	114.1	4/7
4ptel-q13.2	Gain	CTC-36P21	RP11-211G17	67.2	3/7
5p15.33-q35.3	Gain	CTD-2265D9	RP11-451H23	177.5	3/7
8p23.3-pcen	Gain	RP11-338B22	CTD-2115H11	43.0	4/7
9p21.3	Loss	RP11-149I2	RP11-468C2	3.2	3/7
9q13-q34.2	Loss	RP11-274B18	RP11-153P4	65.3	3/7
13q12.11-q33.2	Loss	RP11-76K19	RP11-406G20	86.5	4/7
14q11.2-q32.33	Loss	RP11-84C10	RP11-417P24	85.1	5/7
15q13.2-qtel	Loss	RP11-38E12	CTB-154P1	71.8	4/7
17q22-qtel	Gain	RP11-429O1	GS1-50C4	30.0	3/7
18qcen-qtel	Gain	RP11-296E23	CTC-964M9	58.8	3/7
20q12-q13.12	Gain	RP4-600E6	RP1-138B7	4.1	3/7
22q12.2-q13.31	Loss	RP1-76B20	LL22NC03-75H12	17.8	7/7

Most frequent gains were seen in 8p23.3-pcen (43 Mb) in four out of seven samples. Other alterations were gains in 4ptel-q13.2 (67.2 Mb), 5p15.33-q35.3 (177.5 Mb), 17q22-qtel (30.0

Mb), 18q (58.8 Mb) and 20q12-q13.12 (4.1 Mb), all observed in three out of seven samples, the lower frequency limit selected for this analysis.

Another frequent alteration was loss of chromosome 14, where one copy of the entire chromosome was lost in five out of seven samples. In addition, three chromosomal regions were lost in four out of seven samples; 1p36.32-p13.1 (114.1 Mb); 13q12.11-q33.2 (86.5 Mb); and 15q13.2-qtel (71.8 Mb), whereas 9q13-q34.2 (65.3 Mb) was lost in three out of seven tumours. A frequency plot of gains and losses for GISTs is shown in Figure 3.6, as well as a representative copy number profile for this type of tumours (GIST1).

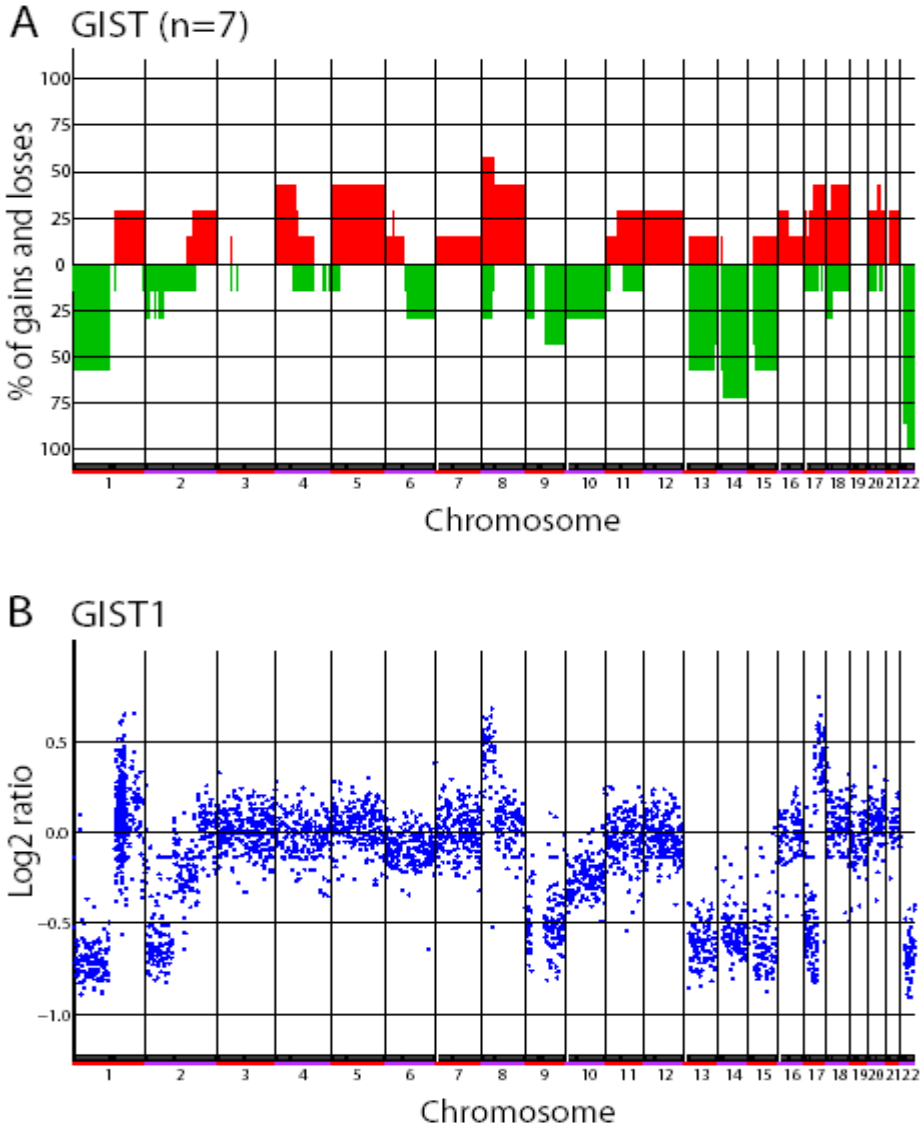

Figure 3.6 CGH Explorer plots showing **A** ACE plot of gains (red) and losses (green) from chromosome 1 to 22 of all GIST samples. The plot shows the frequency of gained and lost clones from the array CGH data **B** Copy number (Log_2 ratio) profile for GIST1 from chromosome 1 to 22.

3.1.4 Copy number changes distinguishing GISTs from LMSs

Significance Analysis of Microarrays (SAM) was applied to determine regions of the genome that can differentiate LMSs and GISTs by means of DNA copy number changes. SAM uses a modified t-test to find significantly differing genes, or genomic clones between two microarray datasets. This analysis led to 178 genomic clones that are significantly differentially altered between LMSs and GISTs. The 178 clones identified, almost exclusively, four chromosomal regions; in 1p, 9q, 14q and 22q.

Between 1p36.11 and p13.1, over 60 % of the genomic clones (58/95) were identified as lost. SAM also selected 29 of 91 clones (32 %) between 9q21.11 and 9q34.3, both regions significantly lost in GISTs compared to LMSs. A large proportion of the clones representing the chromosomal segments 14q11.2-q32.33 and 22q11.32-q13.31 were lost in GIST, 47/91 (52 %) and 40/43 (93 %) clones, respectively. SAM showed that these segments were lost in GIST compared to LMS. Figure 3.7 shows the genomic areas that are significantly different in copy number between GISTs and LMSs.

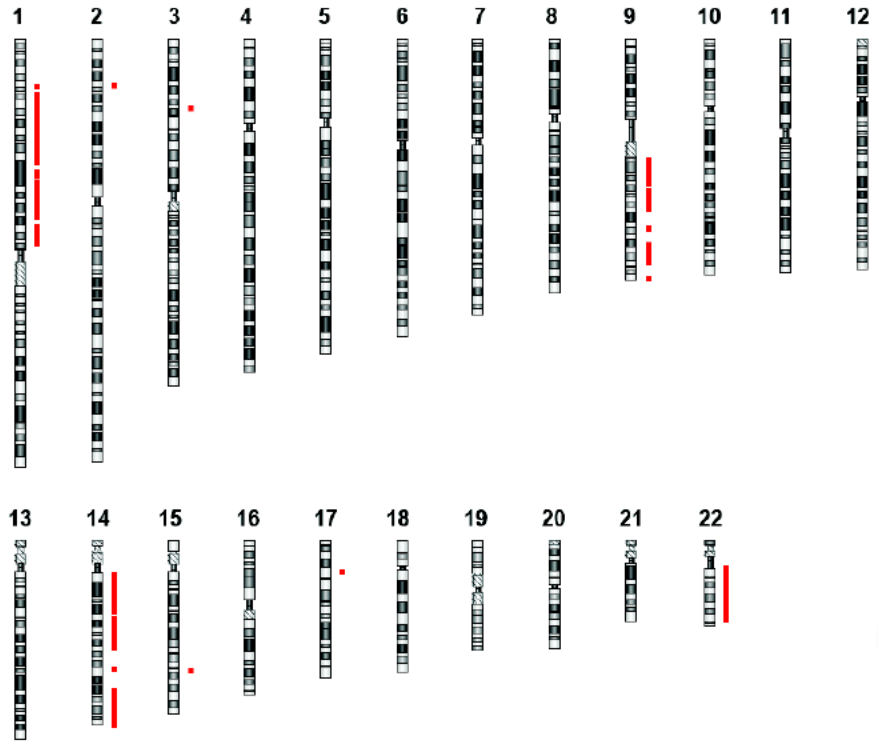

Figure 3.7 In red, genomic areas differing significantly in copy number between GISTs and LMSs.

3.2 Fluorescence in Situ Hybridisation

From the array-CGH results, the region 17p13.1-p11.2 (12.7 Mb) showed the highest level of amplification in the LMS samples. 23 BACs and PACs, starting with RP11-404G1 and ending with RP11-121A13, represented the region. To confirm these results, four clones within the 23 were used as probes in FISH experiments: RP11-12H18, RP11-471L13, RP11-219A15 and RP11-121A13.

Probes from 17q11.2 (centromeric) and 17q25.3 (telomeric) regions, were also tested. These probes, located within normal regions of chromosome 17q (based on array CGH analysis), could give information on ploidy of 17.

The four BACs were first hybridised to metaphase slides to assay the specificity of the clones to the chromosome 17. Figure 3.8 (B&C) shows photo of metaphase chromosomes hybridised with RP11-121A13 and RP11-471L13; the remaining clones showed equal specificity and similar signals (data not shown).

Three LMS samples, LMS1, LMS10 and LMS25 were tested with the four clones. The clones were hybridised to interphase nuclei from the LMS samples. From the array CGH data, LMS1 and LMS25 showed DNA amplification in 17p13.1-p11.2, while LMS10 had normal copy numbers in that region.

Hybridisation to normal cells would usually give two signals, one from each chromosome. In case of amplification, more signals can be seen, and the number reflects the level of amplification in the sample. 10 or more signals show high-level amplification, while three to nine are scored as moderate amplification. Figure 3.8 (D&E) shows photos of interphase nuclei hybridised with one of the clones.

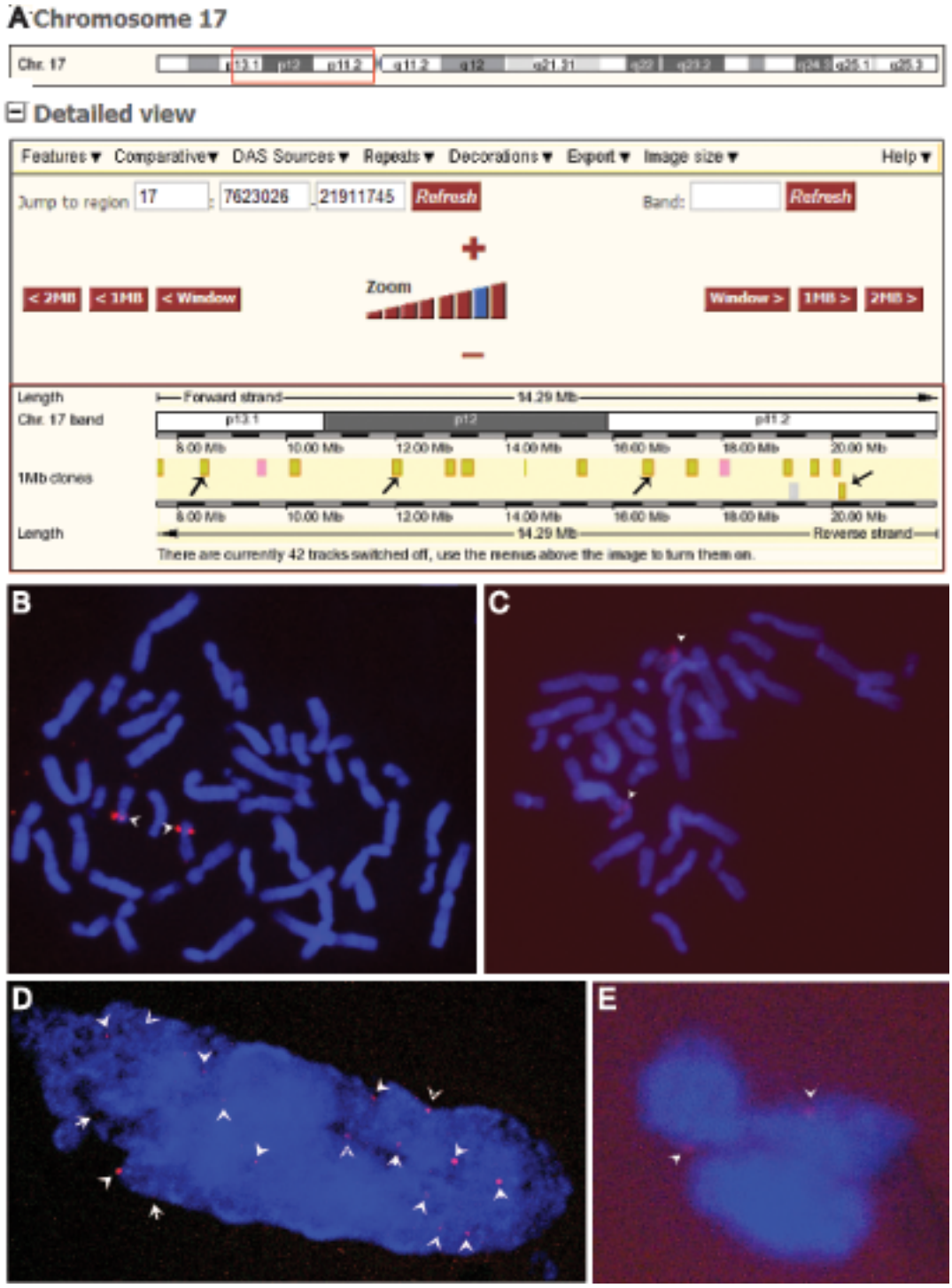

Figure 3.8 FISH experiments. **A** Overview of the chromosome 17 and the gained region 17p13.1-p11.2. The four arrows are pointing the clones tested, from left to right: RP11-12H18, RP11-471L13, RP11-219A15 and RP11-121A13. Specificity of the four clones to the chromosome 17 were also tested (only two are shown here) **B** RP11-121A13 **C** RP11-471L13. Photos of RP11-121A13 hybridisation to LMS25 are shown in D & E. **D** Photo shows high-level amplification in this nucleus (>10 signals are counted). **E** Normal cell

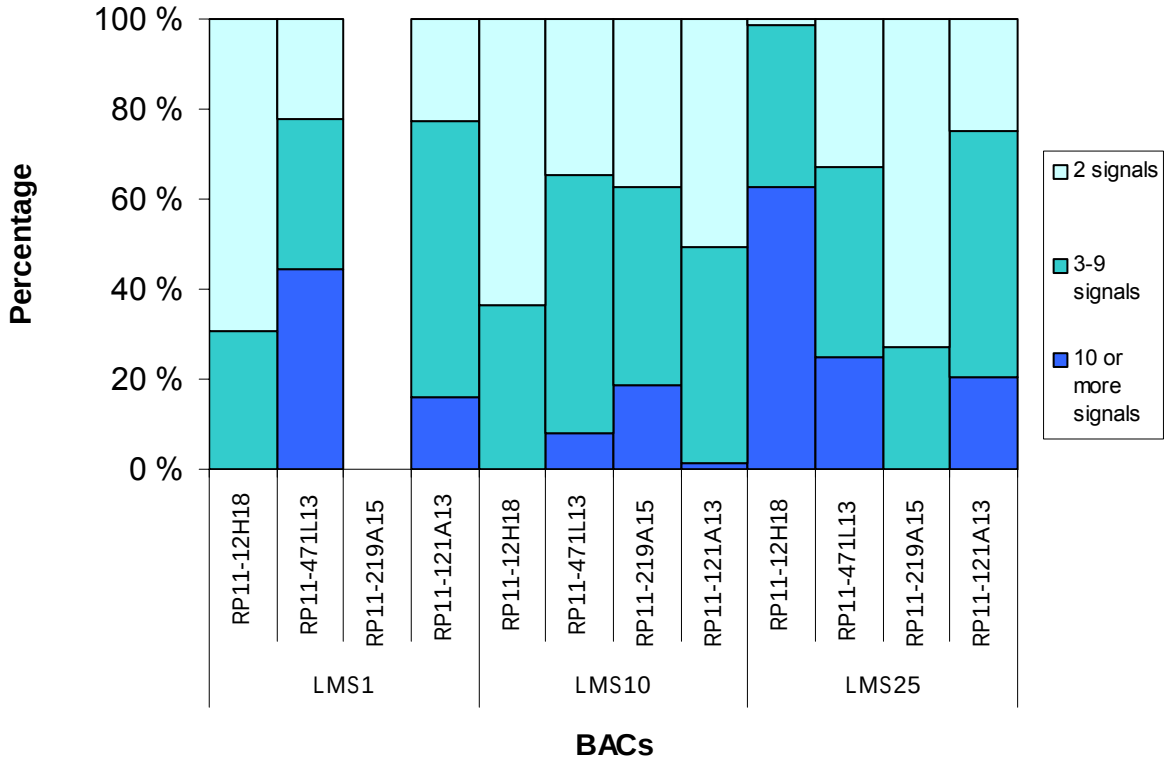

Figure 3.9 Percentage of amplification of four clones tested in LMS1, -10 and –25. 2 signals represent normal DNA copy number, 3-9 a moderate amplification while 10 or more represent high level amplification

The signals were counted and summarised as shown in Figure 3.9. Most cells show normal or moderate amplification, although normal stroma cells (tumour heterogeneity) contribute to this percentage. The sample showing highest amplification is LMS25; with three of four clones showing at least 20% high amplification, especially the clone RP11-12H18 with more than 60%.

LMS10 was expected to show a normal profile, but moderate amplification was seen in all four clones and high amplification in three of them, although the fraction was low compared to LMS25.

In LMS1, RP11-471L13 was the most amplified clone with over 40% followed by RP11-121A13. Data for RP11-219A15 and probes in centromeric and telomeric regions of 17q, is missing because of technical problems. These probes were set up many times but still did not show good results.

4. Discussion

In order to identify genes associated with human diseases, a combination of disciplines is needed; genome mapping, sequencing and computational tools, as well as refined molecular techniques. Cancer is to a large extent caused by epigenetic and structural chromosomal aberrations, inheritable at the cellular level. In sarcomas, chromosome aberrations have been analysed by different methods, among them chromosome banding, FISH, multicolour FISH (M-FISH), spectral karyotyping (SKY) and conventional CGH.

The initial aim of this project was to discover novel cancer related genes by identifying recurrent DNA copy number changes in a panel of 20 human sarcoma samples (LMS and GIST). Tumour samples were analysed for DNA copy number changes using array CGH. Chromosomal regions showing gains or losses in at least 30% of the tumours were considered as recurrent regions of alteration. We detected an amplicon in 17p13.1-p11.2 and used FISH to validate our findings.

4.1 Microarray analyses of DNA copy number variation

Array CGH makes it possible to measure DNA copy number at high resolution and sensitivity. Using BAC and PAC clones with known chromosomal location spotted on a glass slide, it was possible to localise in which regions of the genome the aberrations took place and whether a gain or a deletion of chromosomal material was present. Furthermore, the BACs and PACs used in this study were all from the genome project, and their sequence was precisely defined. Upon mapping, it was thus easy to identify which genes were included in the detected amplicons.

Except for validation purposes, array CGH is much more efficient than clone-by-clone testing of copy numbers by southern blotting or FISH experiments, often time requiring because of the need to select and optimize which probes to use, the technical difficulties of the procedures and the laborious manual scoring. Array CGH decreases substantially the labour involved as the whole genome can be analysed in a single experiment. The implementation of whole-genome amplification protocols is also bringing array CGH closer to clinical use, enabling the analysis of small numbers of cells such as those obtained from thin-needle biopsies. Another advantage of this method is its rapidity and amenity to automation, as it can

be performed on any sample from which genomic DNA can be extracted, and thus, does not require cell culture (Salman et al. 2004).

Although array CGH has been used mostly in cancer research, its application in the study of many other diseases is increasing exponentially. Compared to traditional cytogenetic techniques, array CGH offers higher resolution opening the possibility to discover aberrations or other submicroscopic aneuploidies. Such results would otherwise be missed by chromosomal banding or by FISH analysis with a limited set of probes.

Although this technology is sensitive, fast and specific, it also has its limitations. Array CGH is impaired to identify balanced chromosomal abnormalities, such as balanced translocations, inversions and intragenic rearrangements. Ploidy changes are also not detected, and array CGH cannot distinguish cytogenetic variation in mixed populations of cells. To detect the mosaicism of a sample, the use of FISH is recommended; as the analysis is done at the single cell level and a large number of nuclei can be screened.

Resolution of the array will be given by the genomic distance between probes. The slides used in this study cover the whole genome at a resolution of 1 Mb, meaning that there are probes distanced approximately every 1 Mb. This has the probability of missing small genetic alterations that fall between two probes. Currently, tiling-path arrays using 32800 BAC probes has decreased the resolution to less than 80 Kbp. In order to improve the resolution further, different types of probes can be used; fosmids, PCR products or oligonucleotides. These new generations of arrays offer the possibility of detecting very small alterations down to 100 bp (Davies et al. 2005).

It is also important to take into account copy number polymorphisms (CNPs), a novel type of genetic diversity between individuals. Although the identification of genome-wide large-scale CNPs is virtually untouched, several studies have ascertained their importance in health and disease. Lin et. al., identified a region (276 bp) of chromosome 22q13 that was deleted not only in 47% of ovarian cancer cell lines but also in 18% of constitutional DNA samples from healthy individuals (Lin et al. 2000). Another study reported a homozygous deletion (102 bp) on chromosome 8p12-21 in biliary and pancreatic tumours (Ryu et al. 2001). Both studies concluded that the identified regions might represent normal human genetic variation rather than cancer-associated aberrations (Buckley et al. 2005). Therefore in order to correctly

interpret genomic data, it is important to distinguish abnormal lesions from normal CNPs (Sebat et al. 2004). However, one can overcome this problem using normal DNA from the same patient as reference, when available. However, this is more demanding than using standard DNA pools, as was done here because the experimental variation will increase.

The analysis of the array CGH data requires software capable to filter and normalise according to statistical criteria. These types of data transformation must be carefully applied, because they can have a profound effect on the results. The purpose of filtering is to exclude poor-quality spots. Quality scores are generated when spots are recognised (gridding), segmented, and their intensity extracted. Softwares like GenePix automatically flags spots of poor quality, but it is often necessary to adjust the gridding and/or manually flag spots. Low signal intensity spots also contribute to data noise. In order to eliminate weak signals spots with intensities lower than the background in both channels and net signal intensities below local background plus twice the standard deviation of the background were excluded.

Normalisation of the microarray data is also important as it removes systematic errors by balancing the fluorescence intensities of the two labelling dyes. Differences in quantity of starting DNA, dye-labelling efficiencies and heat and light sensitivities are among the reasons why data must be normalised (Quackenbush 2002). Some methods for calculating normalization factor include median global normalisation, global intensity depended (also known as global locally weighted scatterplot smoothing, LOWESS) and print-tip LOWESS.

In spite of careful normalisation, $\log_2$-ratio values can have a systematic dependence on intensity; this can be seen as deviation from zero for low-intensities. Global intensity depended algorithm (LOWESS) (Cleveland et al. 1976) was favoured in this study because it can remove those intensity-dependent effects in the $\log_2$-ratio values, and it could be applied on the entire data set (Quackenbush 2002). This type of normalisation applies to slides with fluorescent images that do not suffer from spatial effects. When there are significant differences affecting regions of the slide separately, suggesting a geographical effect, print-tip LOWESS should be considered (Leung and Cavalieri 2003).

Missing values can also affect the analysis of microarray data. This can occur for diverse reasons such as scratches or dust over the array feature, or low intensity. The latter problem may be more serious, as certain probes may give poorer signal due to low DNA amount or

less efficient representation by the degenerate PCR primers used to amplify it. Thus the data used for imputation in those cases is less reliable, and may affect a larger part of the data set. Before methods for replacement of missing values were developed, missing $\log_2$-transformed data was often replaced by zeros, however analysis methods such as clustering and self-organising maps required more accurately estimation of missing values (Troyanskaya et al. 2001).

Common methods for missing value imputation include filling in least squares estimates, iterative analysis of variance methods, randomised inference methods and likehood-based approaches. In this work, missing values were permuted by K Nearest Neighbours algorithm (KNN). The KNN-method selects the clones with values similar to the clone of interest to impute missing values. This is done comparing both the values of the same clone in the neighbouring samples and the neighbouring values of the clones in the same sample. This method is very accurate, showing only 6-26% average deviation from the true values depending on the type of data and fraction of values missing (Troyanskaya et al. 2001).

It is important to mention that different statistical methods may produce different (but usually overlapping) sets of significantly gained or deleted regions. Therefore, it is important to choose the right transformation method according to the sample and different criteria to study (Quackenbush 2002).

Although the results from array CGH data can be difficult to interpret, it is possible to compare and validate them with other techniques. In this project, array CGH results from some of the samples were integrated with gene-expression data. Finding that a gene is over-expressed when its copy number is elevated supports its functional role in cancer.

Identification of deleted regions and genes related to them can be complex. In some cases, a decrease in expression caused by deletion of a single copy of a gene contributes to tumour development. In the case of tumour suppressor genes, function is lacking because of deletion of all copies of a gene or deletion of one copy and mutation or epigenetic alterations of the other, or alteration of one copy and replacement of the other by a duplicate of the altered copy (Pinkel and Albertson 2005). There is also the case that, only a single working copy of a gene is present (with the other copy inactivated by hereditary mutation or another mechanism), and

this single functional copy of the gene does not produce enough of gene product leading to an abnormal or disease state (Griffith 2005).

4.1.1 Array CGH as a classification tool

Initially 19 tumour samples diagnosed as LMSs were analysed using array CGH. The data resulting from this analysis was filtered and normalised. The samples were clustered using J-Express.

In this study, hierarchical clustering was used to determine if copy number changes could be used to classify samples based on their phenotypic, or clinical characteristics. In this approach, standard statistical algorithms are used to arrange the tumour samples according to their similarity. The output is displayed graphically, grouping the samples according to patterns of gain or loss in DNA copy number. This approach was preferred because it is an unsupervised method, meaning it does not take into account sample classification therefore prior knowledge about the sample is not needed (Eisen et al. 1998; Quackenbush 2002). It is important to mention that only the tumour samples were clustered, thus keeping genomic clones in chromosomal order to maintain the positional relation.

A disadvantage of hierarchical clustering is that each sample or clone can only be placed in one relation, whereas clearly there may be relations in several directions. Other algorithms, such as self-organising maps and K means clustering, can be used to predefine the number of groups (Nilbert et al. 2004)

Supervised methods depend on prior knowledge about samples in order to search for correlations with disease state. Also there has to be sufficient numbers of samples with known classification in a training set to calibrate the method and in this study, the set of samples is small making the unsupervised method a better alternative to cluster the data set (Ringner et al. 2002).

By using hierarchical clustering, two well-defined main clusters were identified. Histological revision of the samples using current pathological criteria showed that seven of the initial 19 LMS samples would today be scored as GIST, confirming that the two clusters separated LMS and GIST. Histopathology, immunohistochemical staining (particulary CD117) and

other clinical parameters has recently helped to differentiate these tumours. Therefore, in this study, array CGH had the power to differentiate the two different groups based on DNA copy number changes. This work shows that changes in DNA copy number reflect the underlying biology for these tumours, and that this information can be used for their classification.

4.1.2 Gene alterations in LMS

Analysis by ACE showed many genetic alterations in the LMS samples. Identification of minimal recurrent regions of alteration was done by considering gain or losses in at least four of the 13 LMS samples (>30%).

In LMSs, 19 recurrent regions of loss and 11 of gain were identified. The most frequent aberrations observed were loss in 10q and 13q, in 10 out of 13 samples. This has been reported in several studies as the most common genomic alterations for LMS (El-Rifai et al. 1998; Otano-Joos et al. 1998; Derre et al. 2001; Hu et al. 2001; Wang et al. 2003). The minimal recurrent region of loss in 10q21.3 contains only one gene reported in the literature; *CTNNA3* (alpha-T-catenin) is involved in the organization of the actin cytoskeleton and in cell adhesion. Functional assays in alpha-catenin-deficient carcinoma cells (e.g. prostate and colon cancer) showed restoration of cadherin-mediated cell–cell adhesion. That indicates that *CTNNA3* is necessary for the formation of stretch-resistant cell–cell adhesion complexes (Ewing et al. 1995; Janssens et al. 2001 and more). Several studies show that aberrations in the catenin/cadherin pathway is close related to tumour aggressiveness and metastasis (Clairotte et al. 2006; Robles-Frias et al. 2006 and more)

Deletion of the 13q14-34 region was found at least in five of 13 LMS samples. This region includes the *RB1* and *DBM* gene loci. Alteration of the *RB1* tumour suppressor gene has frequently been seen in a number of cancers, including leiomyosarcoma and other sarcomas (Stratton et al. 1989). Another locus contained in the same region of chromosome 13 has been suggested to contain genes acting as tumour suppressor genes. This locus is composed by the genes *KCNRG*, *FAM10A4* and *DLEU7,* involved in B-cell chronic lymphocytic leukemia and prostate cancer development (Sossey-Alaoui et al. 2002; Ivanov et al. 2003; van Everdink et al. 2003; Hammarsund et al. 2004).

Loss of DNA copy number of 2p and 10q, as well as gain in 1q and 17p has been frequently detected in high-grade tumours, while loss of 13q is an early event in development. All these regions may be associated with more aggressive behaviour and shorter survival time (El-Rifai et al. 1998; Wang et al. 2003; Hu et al. 2005). The region 1q21.1-q23.3 is frequently amplified in sarcomas and other types of cancer as osteosarcoma, breast and ovarian cancer among others (Forus et al. 2001). The *APOA2* and *PPIAL4* genes are known target of this region (Meza-Zepeda 2003; Kresse et al. 2005).

The region 17p13.1-p11.2 was gained in at least seven out of 13 samples with four samples showing high levels of amplification (log2 ratio >1). 17p has been reported in the literature as a common region for high amplification in LMS (El-Rifai et al. 1998; Levy et al. 2000; Otano-Joos et al. 2000 and more) and is also frequently amplified in osteosarcomas and retinoblastoma (Forus et al. 1995; Tarkkanen et al. 1995; Atiye et al. 2005 and more).

The minimal recurrent region in 17p13.1-p11.2 of approximately 1.9 Mb contained many genes that are usually co-amplified. This makes it difficult to narrow down the search for the possible "target genes", assumed to "drive" the amplification. One way to address this problem is to analyse many samples, looking for a minimal common amplicon. Another way is to look for expression levels, as an amplified gene should be consistently over-expressed in amplified samples. Both of these methods usually lead to more than one candidate gene in each amplicon.

Gene expression for 70 of the 172 genes contained in the region 17p13.1-p11.2 has been analysed using cDNA microarrays in a parallel study by Francis, *et al.,* (unpublished). Expression data from seven samples were used to narrow down the list of candidate genes, unfortunately only three of the samples (LMS1, -12 and –23) analysed by array CGH, had increased copy number of the region. Over-expression was also seen in samples that were not amplified (LMS3 and -10). Those genes are probably over-expressed by other mechanisms than amplification. Nine genes located within the amplified region showed increased expression ($\log_2$ ratio >1) relative to the soft tissue sarcoma median in two or more of the seven LMSs analysed.

The *MFAP4* gene was over-expressed in four LMSs; *MFAP4* could be involved in calcium-dependent cell adhesion or intercellular, and is commonly deleted in Smith-Magenis

syndrome and might have a role in neuroblastoma development (Zhao et al. 1995; Hienonen et al. 2005).

AURKB was over-expressed in three LMSs; this gene is involved in regulation of the cleavage of polar spindle microtubules and is a key regulator for the onset of cytokinesis during mitosis (Kimura et al. 1998), consequently it has been related to lung, pancreatic, thyroid and prostate cancer development (Kokkinakis et al. 2005; Nikiforov 2005; Smith et al. 2005; Chieffi et al. 2006).

SREBF1 was over-expressed in three LMSs. The gene *SREBF1* regulates the transcription of genes for sterol biosynthesis and the LDL receptor gene (Yokoyama et al. 1993) and has been shown to be up-regulated in prostate cancer during progression to androgen independence (Ettinger et al. 2004), as well as transcriptional regulation of fatty acid synthesis in colorectal cancer (Otano-Joos et al. 2000)

ALDH3A2 was over-expressed in two of the samples. Aldehyde dehydrogenase isozymes are thought to play a major role in the detoxification of aldehydes generated by alcohol metabolism and lipid peroxidation. This gene product catalyzes the oxidation of long-chain aliphatic aldehydes to fatty acid. Mutations in the gene cause Sjogren-Larsson syndrome (De Laurenzi V 1996).

MAPK7 is involved in the transcriptional activation of factors modulating expression of genes required for cell proliferation and survival (Zhou et al. 1995). Together with *SHMT1,* a gene coding for a folate-metabolizing enzyme which catalyses the reversible conversion of serine into glycine (Garrow et al. 1993) were found over-expressed in two of the LMS samples. MAPK7 is a member of the MAP kinase subfamily of the serine/threonine protein kinase family and has been related to many malignancies including sarcoma. *SHMT1* has been associated to adult acute lymphocytic leukaemia and malignant lymphoma (Skibola et al. 2002; Hishida et al. 2003).

Based on the functions of some these genes, it is possible to related them to LMS biology. Some of genes can control cell proliferation, something that can be linked to tumour development and metastasis. *AURKB,* key regulator of cytokinesis during mitosis can affect

both proliferation and the amount of chromosomal DNA in the cells; its disruption could lead to aneuploidy or other aberrations in LMS.

High level amplification of 17p11~12 has also been reported in previous studies of osteosarcomas (Forus et al. 2001; Atiye et al. 2005 and more) suggesting that multiple amplicons in the 17p region are involved in the development of sarcomas. It is also important to account that other target genes not identified within this study could be responsible for alterations in this region.

Two interesting samples: LMS1 and LMS25

In this study, array CGH was also used to compare a primary tumour, LMS1 to its metastasis, LMS25. The patient, from whom this sample was taken from, suffered multiple recurrences and metastasis during a period of about 20 years. It is interesting to see how the DNA copy number profiles of the two samples were very similar, supporting that the two tumours were clonally related, although they differ in some regions of the genome.

Two important aberrations differing between LMS25 and LMS1 were loss of large parts of chromosome 10 and 13q12.11-q13.3. Although in this study, 10p was also altered, loss of 10q has been proposed as a potential marker for clinical diagnosis and prognosis in early studies because of its association with metastases and large tumours, meaning aggressive behaviour of LMS. That is also the case of regions in the q arm of the chromosome 13 (Hu et al. 2001; Hu et al. 2005). Other aberrations differentiating the two samples were losses in the p arm of chromosome 1, 4q28.1-qtel, 11p15.5-p12, 16q22.2-tel, 17q12-qtel and 19q12-q13.43 and gains in 6p21.32-p21.1, 9p24.3-p21.3, 16p13.11-p11.2 and 19p13.3-p12.

Loss of tumour suppressor genes in deleted regions may be responsible for activation or inactivation of other oncogenes. The restoration or loss of chromosome regions in metastases can be explained by clonal evolution in tumour development. The metastasis will have selective advantages over their primary tumour with each carcinogenic mutation. Genetic modifications will then evolve extensively in the tumour cells, in parallel to tumour progression, meaning higher number aberrations will accumulate over time (late-stage tumours will have greater alterations than their precursors). Although it is important to consider that the changes on the magnitude of the aberrations could indicate different tumour

heterogeneities in the two samples. This means different populations of cancer/normal cells could be enriched in one tumour compared to the other.

4.1.3 Genetic Alterations in GIST

Loss of DNA was more frequent in GISTs, seven recurrent regions of loss were determined compared to six regions of gain. Some of the results of the copy number analysis can be compared to the literature. That is the case for the lost regions in 14q11.2-q32.33 and 22q12.2-q13.31. Previous studies, shown that loss of chromosome 14q and 22q are common, followed by the loss of regions in 1p, 9p and 15q (el-Rifai et al. 1996; Debiec-Rychter et al. 2001; Heinrich et al. 2003 and more). Deletion of 14q is often seen in benign GISTs indicating that this change is an early event in GIST tumorigenesis, also deletion in 15q was postulated as a genetic marker distinguishing high-risk GISTs from low-risk GISTs in the same study (Heinrich et al. 2003). Loss of 9p in GISTs has been associated with aggressive behaviour and correlation with loss of *p16Ink4a* tumour supressor gene, frequently loss in many types of cancer. (Schneider-Stock et al. 2003; Ricci et al. 2004; Sabah et al. 2004)

Gain of 5p and 20q was observed in three out of seven tumours. This as well as gain in 17q and loss of 13q has been associated with aggressive and metastatic GISTs (El-Rifai et al. 2000; Debiec-Rychter et al. 2001). Recurrent gain of 17q and loss of 13q were observed in three and four out of seven samples respectively. In general, increased number of genetic changes correlates often with aggressive and malignant behaviour in GISTs.

4.1.4 Patterns Distinguishing GIST from LMS

As shown above, hierarchical clustering of array CGH data can be used to differentiate two histologically similar sarcoma groups. In order to identify the specific regions that differ in copy number between GISTs and LMSs significance Analysis of Microarrays (SAM) was applied. The analysis identified 178 genomic clones that were significantly differently altered between LMSs and GISTs. The 178 clones group almost exclusively into four chromosomal regions; in 1p36.11-p13.1, 9q21.11-9q34.3, 14q11.2-q32.33 and 22q11.32-q13-31 All these regions were preferentially but not exclusively lost in GISTs compared to LMSs.

Most common regions differencing GISTs from LMSs from the literature include regions in chromosome 1, 14, 15 and 22 (El-Rifai et al. 1998; Knuutila et al. 1998; Sarlomo-Rikala et al.

1998). These regions could be potential markers for differential clinical diagnosis. This distinction is important because GIST show a poor response to chemotherapy and radiotherapy, often used in LMS treatment but have positive response to imatinib mesylate a new designer drug for kinases (Tornillo et al. 2005).

4.2 Validation of Array CGH analyses by FISH

FISH is a powerful technique, capable of detecting a specific DNA target (in this case, the nuclei of interphase cells or normal metaphase chromosomes) in fixed normal or tumour samples. Genetic alterations can be detected at single-cell level making it possible to discern cells with normal copy number and cells with gains or losses of DNA. Searching for candidate genes only by FISH, can be time-consuming because many probes have to be selected, tested and optimised. Combining array CGH as a screening tool and later FISH to narrow down and validate areas, makes it less demanding to identify target candidate genes.

In this project, FISH was used to validate copy number alteration detected by array CGH in three LMS samples. Array CGH showed 23 BACs and PACs representing a region of amplification in 17p13.1-p11.2. From this region, four BACs showing amplification were used as probes. Probes from 17q11.2 (centromeric) and 17q25.3 (telomeric) regions, were also tested. After several attempts, these probes did not provide any satisfactory result in metaphase chromosomes. These probes, located within normal regions of chromosome 17, could give information on ploidy of 17q in the samples.

All tumour samples are heterogeneous, consisting normal and sometimes multiple populations of neoplastic cells. Array CGH analysis determines the average copy number of the tumour and does not give information on heterogeneity. On the other hand, FISH is capable of distinguishing this relation, because single cells are being investigated and the information on different populations can be obtained. All the samples showed a minimum of 20% of normal cells (2 signals/nuclei) but also a fraction that showed moderate amplification up to 50% (3-9 signals/nuclei) and high amplification, up to 60% (more than 9 signals/nuclei). In order to examine the heterogeneity of the tumour, a high number of nuclei (>100) must be counted, but in some of the samples high number were difficult to obtain, especially LMS1. As explained before, the slides were dripped with cell suspension (in acetic acid, the amount depended on pellet size), but it was difficult to estimate the dilution of the cell suspension to

avoid “flooding” the slides with cells. Under the microscope, cells would be seen as clumps, making signal counting difficult.

High background was also a problem when counting the number of signals in each nucleus. According to Henegariu et. al.(2001) background is mainly influenced by: i) the efficacy of the competitor DNA in blocking the repetitive sequences of the probes, ii) the hybridisation temperature, when lowered non-specific binding it could increase, iii) hybridisation time and probe amount, iv) the stringency of the post-hybridisation washes and v) incomplete RNA removal. For future attempts the quality of the slides, and either the amount of competitor DNA and probe have to be adjusted cautiously for each BAC. It is well known, however that some BACs just won’t give satisfactory results from FISH.

Because of the low amount of nuclei counted for some of the probes, it is difficult to compare the samples to each other. LMS10 was expected to have normal DNA copy number but FISH results showed also moderate amplification and some nuclei with high amplification in RP11-471L13 and RP11-219A15. The heterogeneity in the sample can explain why the results in this case were different. Normal cells in the tumours could be the reason why the array CGH analysis underestimated the amplification levels and a subpopulation of cells with high amplification was possibly undetected.

LMS1 and LMS25 were expected to have gain in DNA copy number according to array CGH data, and this agrees with the results. All four clones showed amplification in three LMSs indicating the possibility that they contain target genes involved in cancer development and/or progression.

The clones tested in this project contained both unknown genes and some genes previous cited in the literature. RP11-471L13 contains the Mitogen-Activated Protein Kinase Kinase 4 gene (*MAP2K4)* involved in activation of JUN kinases, signal transduction and also the Ras pathway and hepatogenesis. *MAP2K4* is also one of the most consistently mutated genes across tumour types (Xin et al. 2004; Koed et al. 2005; Hickson et al. 2006) although only one study has reported high-level amplification (van Dartel et al. 2003) RP11-121A13 contain the Spectrin Domain with Coiled-Coils 1 gene (*SPECC1).* This gene has been reported as highly expressed in tumour cells and for being a fusion partner to the gene PDGFRB in juvenile myelomonocytic leukaemia (Hammarsund et al. 2004; Morerio et al. 2004)

Because only four clones were analysed by FISH, it cannot be excluded that other clones within the amplicon could also contain interesting “target” genes. There is also the probability of alterations being missed because they fall between probes. To improve the attempt overlapping probes can be used to narrow down interesting areas and identify candidate genes.

5. Conclusions and future prospects

Most of previous studies on copy number alterations used conventional CGH and cytogenenics in order to identify gained or deleted regions. In this study, a panel of sarcoma sample were analysed by array CGH, a technology with high resolution that makes possible the identification of narrow regions with copy number changes.

The limit of conventional CGH is its resolution. Large altered chromosome arms or regions have been reported in literature, but few use the array CGH technology in their analyses. In this project, DNA microarray slides covering the whole genome at a resolution of 1 Mb were used trying to identify smaller altered regions.

In this study it was possible to narrow down previously reported regions as well as identification of novel segments. Most frequent aberrations in LMSs, were narrow down to loss of 10q21.3 and 13q14.2-q14.3 and gains in 5p13.2-pcen, 9q21.13-q31.3, among others. Novel regions of gain in 3p12.3-p12.1 and large parts of 14q; losses in 6p25.2-p22.3 and q14.1-q23.3 as well as 11p15.5-q24.1, 21q21.1-q22.11 and 22q13.1-q13.33 were identified by array CGH.

In GIST, the most frequent aberrations were could be narrowed down to 1p36.32-p13.1, 13q12.11-q33.2, 9q13-q34.2, 14q11.2-q32.2, 15q13.2-qtel and 22q12.2-q13.31, among others. Novel regions as 4ptel-q13.2, 8p23.3-pcen and 18qcen-qtel, also loss in 9q13-q34.2 could also be identified.

Until recently, GISTs were classified among smooth muscle tumours, and they were diagnosed often as LMS. In this project, array CGH was also used to test whether this technology could be used to differentiate these two kinds of sarcoma. Hierarchical clustering of the samples could discriminate LMS from GIST. These two groups could also be differentiated by their patterns of copy number alterations; the four chromosomal regions: in 1p36.11-p13.1, 9q12.11-q34.4, 14q11.2-q32.33 and 22q11.32-q13.31.

These results highlight the use of array CGH to distinguish histological similar tumours. This technique has also the potential to be implemented in clinical diagnostics. As mentioned in the introduction, it is important by means of the treatment to be followed. However, further

validation with a larger panel of samples could also bring information about those novel regions. Knowledge on the alterations affecting LMS and GIST can tell us more about the biology of these types of tumours. Some of the altered regions could also be used as biological markers in the prediction of tumour development and progression.

Two LMS samples, LMS1 and –25, a primary tumour and its metastasis, are samples that will be followed closely. Because the patient, whom the sample was taken from, has had recurrent metastases during a period of about 20 years, it is possible samples from others biopsies had been spared. Array CGH analysis of those samples and comparison on alterations, could lead to regions and candidate genes important for leiomyosarcoma development and progression.

The region in chromosome 17p13.1-p11.2 was also gained in LMSs, therefore it is further investigated in this study. Four BAC probes within this region were chosen and tested in two of the high-amplified samples and one normal sample. They all showed gain of DNA copy number. High background and noise was a problem when counting the signals, therefore alternative and milder techniques for cell fixation and pre-treatment of the slides should be considered. Also, in order to achieve representative data, a higher number of nuclei for each sample must be counted. Clones overlapping this region could be tested to assure the identification of candidate genes.

Only two genes previously cited in the literature were contained in the clones tested by FISH; those were *MAP2K4* often mutated in many tumour types and *SPECC1* involved in juvenile myelomonocytic leukaemia. Because only four BACs were analysed by FISH, it cannot be excluded that other parts of the amplicon could also contain interesting “target” genes. To improve this attempt other amplified clones could be tested and/or overlapping clones could be used to map that region.

In LMS, some interesting genes were target of amplification in the 17p13.1-p11.2 region and were also by expression arrays found to be over-expressed. They include genes related to cancer and other diseases. The most over-expressed genes include *MFAP4* possible involved in neuroblastoma development and *AURKB* involved in different types of cancer. It is possible other genes went missing because the limited information on gene expression. A study of a more complete panel of genes could have given even more information on

amplified and over-expressed genes in the data set and better discriminate the candidate genes.

After validation with a larger panel of samples in array CGH and gene expression, functional analyses of a limited set of consistently amplifies and over-expressed genes, such as over-expression in cell lines and/or siRNA to knocking down the candidate genes could be done to it, to better identify the “real” driver genes.

References

Alberts B, J. A., Lewis J, Raff M, Roberts K and Walter P (2002). Molecular Biology of the Cell. New York Garland Publishing.

Albertson, D. G., C. Collins, et al. (2003). "Chromosome aberrations in solid tumors." Nat Genet **34**(4): 369-76.

Alitalo, K., R. Winqvist, et al. (1984). "Aberrant expression of an amplified c-myb oncogene in two cell lines from a colon carcinoma." Proc Natl Acad Sci U S A **81**(14): 4534-8.

Atiye, J., M. Wolf, et al. (2005). "Gene amplifications in osteosarcoma-CGH microarray analysis." Genes Chromosomes Cancer **42**(2): 158-63.

Bardi, G., B. Johansson, et al. (1992). "Recurrent chromosome aberrations in abdominal smooth muscle tumors." Cancer Genet Cytogenet **62**(1): 43-6.

Bjerkehagen, B. and O. Myklebost (2005). "Molekylaerbiologisk diagnostikk ved bein- og blotvevssvulster." Tidsskrift for Den Norske Laegeforening **125**(23): 3286-9.

Buckley, P. G., K. K. Mantripragada, et al. (2005). "Copy-number polymorphisms: mining the tip of an iceberg." Trends Genet **21**(6): 315-7.

Caudell, J. J., M. T. Deavers, et al. (2005). "Imatinib mesylate (gleevec)--targeted kinases are expressed in uterine sarcomas." Applied Immunohistochemistry & Molecular Morphology **13**(2): 167-70.

Cheung, V. G., H. L. Dalrymple, et al. (1999). "A resource of mapped human bacterial artificial chromosome clones." Genome Res **9**(10): 989-93.

Chieffi, P., L. Cozzolino, et al. (2006). "Aurora B expression directly correlates with prostate cancer malignancy and influence prostate cell proliferation." Prostate **66**(3): 326-33.

Clairotte, A., I. Lascombe, et al. (2006). "Expression of E-cadherin and alpha-, beta-, gamma-catenins in patients with bladder cancer: identification of gamma-catenin as a new prognostic marker of neoplastic progression in T1 superficial urothelial tumors." Am J Clin Pathol **125**(1): 119-26.

Cleveland, W. S., B. Kleiner, et al. (1976). "Robust statistical methods and photochemical air pollution data." J Air Pollut Control Assoc **26**(1): 36-8.

Dang, C. V. (1999). "c-Myc target genes involved in cell growth, apoptosis, and metabolism." Mol Cell Biol **19**(1): 1-11.

Davies, J. J., I. M. Wilson, et al. (2005). "Array CGH technologies and their applications to cancer genomes." Chromosome Res **13**(3): 237-48.

De Laurenzi V, R. G., Hamrock DJ, Marekov LN, Steinert PM, Compton JG, Markova N, Rizzo WB. (1996). "Sjogren-Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene." Nat Genet **12**(1): 52-7.

Debiec-Rychter, M., J. Lasota, et al. (2001). "Chromosomal aberrations in malignant gastrointestinal stromal tumors: correlation with c-KIT gene mutation." Cancer Genet Cytogenet **128**(1): 24-30.

Derre, J., R. Lagace, et al. (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): 211-5.

Dysvik, B. and I. Jonassen (2001). "J-Express: exploring gene expression data using Java." Bioinformatics **17**(4): 369-70.

Eisen, M. B., P. T. Spellman, et al. (1998). "Cluster analysis and display of genome-wide expression patterns." Proc Natl Acad Sci U S A **95**(25): 14863-8.

El-Rifai, W., M. Sarlomo-Rikala, et al. (2000). "DNA sequence copy number changes in gastrointestinal stromal tumors: tumor progression and prognostic significance." Cancer Res **60**(14): 3899-903.

El-Rifai, W., M. Sarlomo-Rikala, et al. (1998). "DNA copy number changes in gastrointestinal stromal tumors--a distinct genetic entity." Ann Chir Gynaecol **87**(4): 287-90.

El-Rifai, W., M. Sarlomo-Rikala, et al. (1998). "DNA copy number changes in development and progression in leiomyosarcomas of soft tissues." Am J Pathol **153**(3): 985-90.

el-Rifai, W., M. Sarlomo-Rikala, et al. (1996). "DNA copy number losses in chromosome 14: an early change in gastrointestinal stromal tumors." Cancer Res **56**(14): 3230-3.

Esteller, M. (2006). "Epigenetics provides a new generation of oncogenes and tumour-suppressor genes." Br J Cancer **94**(2): 179-83.

Ettinger, S. L., R. Sobel, et al. (2004). "Dysregulation of sterol response element-binding proteins and downstream effectors in prostate cancer during progression to androgen independence." Cancer Res **64**(6): 2212-21.

Ewing, C. M., N. Ru, et al. (1995). "Chromosome 5 suppresses tumorigenicity of PC3 prostate cancer cells: correlation with re-expression of alpha-catenin and restoration of E-cadherin function." Cancer Res **55**(21): 4813-7.

Feuk, L., A. R. Carson, et al. (2006). "Structural variation in the human genome." Nat Rev Genet **7**(2): 85-97.

Fletcher CDM, U. K., Mertens F, red. (2002). World Health Organization Classification of Tumours Pathology and Genetics Tumours of Soft Tissue and Bone. Lyon, IARC Press.

Forus, A., M. L. Larramendy, et al. (2001). "Dedifferentiation of a well-differentiated liposarcoma to a highly malignant metastatic osteosarcoma: amplification of 12q14 at all stages and gain of 1q22-q24 associated with metastases." Cancer Genet Cytogenet **125**(2): 100-11.

Forus, A., D. O. Weghuis, et al. (1995). "Comparative genomic hybridization analysis of human sarcomas: II. Identification of novel amplicons at 6p and 17p in osteosarcomas." Genes Chromosomes Cancer **14**(1): 15-21.

Gall, J. G. and M. L. Pardue (1969). "Formation and detection of RNA-DNA hybrid molecules in cytological preparations." Proc Natl Acad Sci U S A **63**(2): 378-83.

Garrow, T. A., A. A. Brenner, et al. (1993). "Cloning of human cDNAs encoding mitochondrial and cytosolic serine hydroxymethyltransferases and chromosomal localization." J Biol Chem **268**(16): 11910-6.

Gole, L. (2001). Fluorescence in situ Hybridisation. Chichester, John Wiley & Sons.

Griffith, D. J., et. al. (2005). Introduction to Genetic Analysis. New York, W.H. Freeman.

Hammarsund, M., M. M. Corcoran, et al. (2004). "Characterization of a novel B-CLL candidate gene--DLEU7--located in the 13q14 tumor suppressor locus." FEBS Lett **556**(1-3): 75-80.

Hartwell L., H. L., Goldberg M. L., Reynolds A. E., Silver L. M., Veres R. C. (2000). Genetics From Genes to Genomes. USA, Mc Graw Hill.

Heinrich, M. C., C. L. Corless, et al. (2003). "PDGFRA activating mutations in gastrointestinal stromal tumors." Science **299**(5607): 708-10.

Helman, L. J. and P. Meltzer (2003). "Mechanisms of sarcoma development." Nature Reviews. Cancer **3**(9): 685-94.

Henegariu, O., N. A. Heerema, et al. (2001). "Improvements in cytogenetic slide preparation: controlled chromosome spreading, chemical aging and gradual denaturing." Cytometry **43**(2): 101-9.

Hickson, J. A., D. Huo, et al. (2006). "The p38 kinases MKK4 and MKK6 suppress metastatic colonization in human ovarian carcinoma." Cancer Res **66**(4): 2264-70.

Hienonen, T., H. Sammalkorpi, et al. (2005). "A 17p11.2 germline deletion in a patient with Smith-Magenis syndrome and neuroblastoma." J Med Genet **42**(1): e3.

Hishida, A., K. Matsuo, et al. (2003). "Associations between polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and susceptibility to malignant lymphoma." Haematologica **88**(2): 159-66.

Hu, J., V. Khanna, et al. (2001). "Genomic alterations in uterine leiomyosarcomas: potential markers for clinical diagnosis and prognosis." Genes Chromosomes Cancer **31**(2): 117-24.

Hu, J., U. N. Rao, et al. (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): 20-7.

Ivanov, D. V., T. V. Tyazhelova, et al. (2003). "A new human gene KCNRG encoding potassium channel regulating protein is a cancer suppressor gene candidate located in 13q14.3." FEBS Lett **539**(1-3): 156-60.

Janssens, B., S. Goossens, et al. (2001). "alphaT-catenin: a novel tissue-specific beta-catenin-binding protein mediating strong cell-cell adhesion." J Cell Sci **114**(Pt 17): 3177-88.

Kallioniemi, A., O. P. Kallioniemi, et al. (1992). "Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors." Science **258**(5083): 818-21.

KH A Choo, J. M. C., S M Cutts, A W Lo (2001). In Situ Hybridisation. Chichester, John Wiley & Sons.

Kimura, M., Y. Matsuda, et al. (1998). "Identification and characterization of STK12/Aik2: a human gene related to aurora of Drosophila and yeast IPL1." Cytogenet Cell Genet **82**(3-4): 147-52.

Kirsch, I. R. and T. Ried (2000). "Integration of cytogenetic data with genome maps and available probes: present status and future promise." Semin Hematol **37**(4): 420-8.

Knudson, A. G., Jr. (1971). "Mutation and cancer: statistical study of retinoblastoma." Proc Natl Acad Sci U S A **68**(4): 820-3.

Knuutila, S., G. Armengol, et al. (1998). "Comparative genomic hybridization study on pooled DNAs from tumors of one clinical-pathological entity." Cancer Genet Cytogenet **100**(1): 25-30.

Koed, K., C. Wiuf, et al. (2005). "High-density single nucleotide polymorphism array defines novel stage and location-dependent allelic imbalances in human bladder tumors." Cancer Res **65**(1): 34-45.

Kokkinakis, D. M., X. Liu, et al. (2005). "Modulation of cell cycle and gene expression in pancreatic tumor cell lines by methionine deprivation (methionine stress): implications to the therapy of pancreatic adenocarcinoma." Mol Cancer Ther **4**(9): 1338-48.

Kresse, S. H., J. M. Berner, et al. (2005). "Mapping and characterization of the amplicon near APOA2 in 1q23 in human sarcomas by FISH and array CGH." Mol Cancer **4**: 39.

Lengauer, C., K. W. Kinzler, et al. (1998). "Genetic instabilities in human cancers." Nature **396**(6712): 643-9.

Leung, Y. F. and D. Cavalieri (2003). "Fundamentals of cDNA microarray data analysis." Trends Genet **19**(11): 649-59.

Levy, B., T. Mukherjee, et al. (2000). "Molecular cytogenetic analysis of uterine leiomyoma and leiomyosarcoma by comparative genomic hybridization." Cancer Genet Cytogenet **121**(1): 1-8.

Li, F. P., J. F. Fraumeni, Jr., et al. (1969). "Cancer mortality among chemists." J Natl Cancer Inst **43**(5): 1159-64.

Lin, H., E. S. Pizer, et al. (2000). "A frequent deletion polymorphism on chromosome 22q13 identified by representational difference analysis of ovarian cancer." Genomics **69**(3): 391-4.

Lindahl, T. (1996). "Genetic instability in cancer. Introduction." Cancer Surv **28**: 1-2.

Lingjaerde, O. C., L. O. Baumbusch, et al. (2005). "CGH-Explorer: a program for analysis of array-CGH data." Bioinformatics **21**(6): 821-2.

Mackall, C. L., P. S. Meltzer, et al. (2002). "Focus on sarcomas." Cancer Cell **2**(3): 175-8.

Magliano D, L. W., Choo A (2001). Nucleic Acid: Hybridization. Chichester, John Wiley & Sons.

McClintock, B. (1942). "The Fusion of Broken Ends of Chromosomes Following Nuclear Fusion." Proc Natl Acad Sci U S A **28**(11): 458-463.

Meza-Zepeda, L. A. (2003). Molecular Genetics of Human Sarcomas: A comprehensive study of two novel amplified regions in 1q21 and 12q14. Faculty of Medicine. Oslo, University of Oslo. **Dr. Philos**.

Morerio, C., M. Acquila, et al. (2004). "HCMOGT-1 is a novel fusion partner to PDGFRB in juvenile myelomonocytic leukemia with t(5;17)(q33;p11.2)." Cancer Res **64**(8): 2649-51.

Myllykangas, S. and S. Knuutila (2006). "Manifestation, mechanisms and mysteries of gene amplifications." Cancer Lett **232**(1): 79-89.

Nikiforov, Y. E. (2005). "Editorial: anaplastic carcinoma of the thyroid--will aurora B light a path for treatment?" J Clin Endocrinol Metab **90**(2): 1243-5.

Nilbert, M., L. A. Meza-Zepeda, et al. (2004). "Lessons from genetic profiling in soft tissue sarcomas." Acta Orthop Scand Suppl **75**(311): 35-50.

Oostlander, A. E., G. A. Meijer, et al. (2004). "Microarray-based comparative genomic hybridization and its applications in human genetics." Clin Genet **66**(6): 488-95.

Otano-Joos, M., G. Mechtersheimer, et al. (1998). "[Analysis of chromosome copy number changes in leiomyosarcoma through molecular cytogenetic methods]." Verh Dtsch Ges Pathol **82**: 207-9.

Otano-Joos, M., G. Mechtersheimer, et al. (2000). "Detection of chromosomal imbalances in leiomyosarcoma by comparative genomic hybridization and interphase cytogenetics." Cytogenet Cell Genet **90**(1-2): 86-92.

Pinkel, D. and D. G. Albertson (2005). "Array comparative genomic hybridization and its applications in cancer." Nat Genet **37 Suppl**: S11-7.

Pinkel, D. and D. G. Albertson (2005). "Comparative genomic hybridization." Annu Rev Genomics Hum Genet **6**: 331-54.

Pinson, B., E. M. Brendeford, et al. (2001). "Highly conserved features of DNA binding between two divergent members of the Myb family of transcription factors." Nucleic Acids Res **29**(2): 527-35.

Quackenbush, J. (2002). "Microarray data normalization and transformation." Nat Genet **32 Suppl**: 496-501.

Ricci, R., V. Arena, et al. (2004). "Role of p16/INK4a in gastrointestinal stromal tumor progression." Am J Clin Pathol **122**(1): 35-43.

Ringner, M., C. Peterson, et al. (2002). "Analyzing array data using supervised methods." Pharmacogenomics **3**(3): 403-15.

Roberts, P. (2001). Cancer Cytogenetics. Chichester, John Wiley & Sons.

Robles-Frias, A., R. Gonzalez-Campora, et al. (2006). "Robinson cytologic grading in invasive ductal carcinoma of the breast: correlation with E-cadherin and alpha-, beta- and gamma-catenin expression and regional lymph node metastasis." Acta Cytol **50**(2): 151-7.

Roosen G, V. F., Vos J-M (2002). Artificial Chromosomes. Chichester, John Wiley & Sons.

Rubin, B. P. (2006). "Gastrointestinal stromal tumours: an update." Histopathology **48**(1): 83-96.

Rudkin, G. T. and B. D. Stollar (1977). "High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence." Nature **265**(5593): 472-3.

Ryu, B., J. Song, et al. (2001). "Frequent germline deletion polymorphism of chromosomal region 8p12-p21 identified as a recurrent homozygous deletion in human tumors." Genomics **72**(1): 108-12.

Sabah, M., R. Cummins, et al. (2004). "Loss of heterozygosity of chromosome 9p and loss of p16INK4A expression are associated with malignant gastrointestinal stromal tumors." Mod Pathol **17**(11): 1364-71.

Salman, M., S. C. Jhanwar, et al. (2004). "Will the new cytogenetics replace the old cytogenetics?" Clin Genet **66**(4): 265-75.

Sandberg, A. A. and J. A. Bridge (2002). "Updates on the cytogenetics and molecular genetics of bone and soft tissue tumors. gastrointestinal stromal tumors.[erratum appears in Cancer Genet Cytogenet 2002 Sep;137(2):156]." Cancer Genetics & Cytogenetics **135**(1): 1-22.

Sarlomo-Rikala, M., W. El-Rifai, et al. (1998). "Different patterns of DNA copy number changes in gastrointestinal stromal tumors, leiomyomas, and schwannomas." Hum Pathol **29**(5): 476-81.

Schneider-Stock, R., C. Boltze, et al. (2003). "High prognostic value of p16INK4 alterations in gastrointestinal stromal tumors." J Clin Oncol **21**(9): 1688-97.

Schwab, M. (1999). "Oncogene amplification in solid tumors." Seminars in Cancer Biology **9**(4): 319-25.

Sebat, J., B. Lakshmi, et al. (2004). "Large-scale copy number polymorphism in the human genome." Science **305**(5683): 525-8.

Skibola, C. F., M. T. Smith, et al. (2002). "Polymorphisms in the thymidylate synthase and serine hydroxymethyltransferase genes and risk of adult acute lymphocytic leukemia." Blood **99**(10): 3786-91.

Smith, S. L., N. L. Bowers, et al. (2005). "Overexpression of aurora B kinase (AURKB) in primary non-small cell lung carcinoma is frequent, generally driven from one allele, and correlates with the level of genetic instability." Br J Cancer **93**(6): 719-29.

Sossey-Alaoui, K., E. Kitamura, et al. (2002). "Characterization of FAM10A4, a member of the ST13 tumor suppressor gene family that maps to the 13q14.3 region associated with B-Cell leukemia, multiple myeloma, and prostate cancer." Genomics **80**(1): 5-7.

Speicher, M. R. (2005). Chromosome. Chichester, John Wiley & Sons.

Sreekantaiah, C., J. R. Davis, et al. (1993). "Chromosomal abnormalities in leiomyosarcomas." Am J Pathol **142**(1): 293-305.

Stratton, M. R., S. Williams, et al. (1989). "Structural alterations of the RB1 gene in human soft tissue tumours." Br J Cancer **60**(2): 202-5.

Tarkkanen, M., R. Karhu, et al. (1995). "Gains and losses of DNA sequences in osteosarcomas by comparative genomic hybridization." Cancer Res **55**(6): 1334-8.

Tornillo, L., G. Duchini, et al. (2005). "Patterns of gene amplification in gastrointestinal stromal tumors (GIST)." Laboratory Investigation **85**(7): 921-31.

Troyanskaya, O., M. Cantor, et al. (2001). "Missing value estimation methods for DNA microarrays." Bioinformatics **17**(6): 520-5.

Tusher, V. G., R. Tibshirani, et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A **98**(9): 5116-21.

van Dartel, M., S. Leenstra, et al. (2003). "Infrequent but high-level amplification of 17p11.2 approximately p12 in human glioma." Cancer Genet Cytogenet **140**(2): 162-6.

van Everdink, W. J., A. Baranova, et al. (2003). "RFP2, c13ORF1, and FAM10A4 are the most likely tumor suppressor gene candidates for B-cell chronic lymphocytic leukemia." Cancer Genet Cytogenet **146**(1): 48-57.

Wang, J., L. A. Meza-Zepeda, et al. (2004). "M-CGH: analysing microarray-based CGH experiments." BMC Bioinformatics **5**: 74.

Wang, R., J. C. Titley, et al. (2003). "Loss of 13q14-q21 and gain of 5p14-pter in the progression of leiomyosarcoma." Mod Pathol **16**(8): 778-85.

Xin, W., K. J. Yun, et al. (2004). "MAP2K4/MKK4 expression in pancreatic cancer: genetic validation of immunohistochemistry and relationship to disease course." Clin Cancer Res **10**(24): 8516-20.

Yokoyama, C., X. Wang, et al. (1993). "SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low density lipoprotein receptor gene." Cell **75**(1): 187-97.

Zhao, Z., C. C. Lee, et al. (1995). "The gene for a human microfibril-associated glycoprotein is commonly deleted in Smith-Magenis syndrome patients." Hum Mol Genet **4**(4): 589-97.

Zhou, G., Z. Q. Bao, et al. (1995). "Components of a new human protein kinase signal transduction pathway." J Biol Chem **270**(21): 12665-9.

Appendix A: Solutions

EDTA, 0.5 M pH 8.0
186.1 g disodium EDTA◦2H2O
800 ml d H2O
Stir vigorously on a magnetic stirrer
Adjust the pH to 8.0 with NaOH
d H2O to 1000 ml
Sterilize by autoclaving

1M Tris-HCl pH 8.0
400 ml dd H_20
60.56 g TRIS
Adjust to pH 8.0 with 10M HCl
Dilute to 500 ml in volumetric flask

Lysis buffer A
10 mM Tris-HCl pH 8
0.1 M NaCl
1 mM EDTA

Lysis Buffer B
10 mM Tris-HCl pH 8
0.1 M NaCl
1 mM EDTA
2% Sodium Sarcosyl
100 µg/ml Proteinase K

Sodium Acetate, NaAc 3M pH 5.2
408 g NaAc◦3 H_2O
800 ml dH_2O
Adjust pH with glacial acetic acid
dH_2O to 1000 ml
Sterilize by autoclave

1X TE
10 mM Tris-HCl pH 7.5
1 mM EDTA

MMI
50% deionised formamide
10% dextran sulphate
2X SSC, pH 7
Warm to 70°C to help dissolving

20X SSC
175 g NaCl
88.2 g $Na_3Citrate◦dH_2O$
800 ml DEPC-dH_2O
Adjust pH to 7.0 with HCl
dH_2O to 1000 ml
Sterilize by autoclave

2X TY-Medium pH 7.4
16 g Tryptone
10 g Yeast Extract
5 g NaCl
1 l dH_2O
Sterilize by autoclave

P1
15 mM Tris, pH 8
10 mM EDTA
100 ug/ml RNase A

P2
0,2 N NaOH
1% SDS

P3
3M KOAC, pH 5.5

1% Agarose Gel
50 ml 1XTEA
0,5 g Agarose
3 ul Ethidium Bromide

Washing Solution 1
250 ml Formamide
50 ml 20X SSC
200 ml dH_2O

Washing Solution 2
50 ml 20X SSC
2.5 ml 20X SDS
450 ml dH_2O

PN-buffer
50 ml 1M Na_2HPO_4
2.5 ml 1M NaH_2PO_4
525 µl Nonidet P-40 (Octylphenoxy, Polyethoxyethanol)

1X TNT
100 ml 1M Tris pH 7.5
30 ml 5M NaCl
5 ml10% Tween
865 ml dH_2O

1X PBS
2.7 mM KCl
1.4 mM KH2PO4
137 mM NaCl
4.3 mM Na2PO4
200 mg KCl
200 mg KH2PO4
8.0 g NaCl
1.15 g Na2PO4·7H2O
800 ml dH_2O
Adjust pH to 7.2 with NaOH
dH_2O to 1000 ml
Sterilize by filtration through a 0.2-0.45 µm filter

Appendix B: Overview of the clone set in 17p13.3-p11.2

List of clones covering the gained region 17p13.3-p11.2 containing the interesting amplicon studied in this. The four clones tested by FISH are underlined in red. Figure is from Ensembl v39 Human Cytoview (www.ensembl.org)

Features in set 1Mb clone set in Chromosome 17 7623026 - 21911745

SeqRegion	Start	End	Name	Well name	Sanger	EMBL Acc	FISH	Centre	State
17	7563870	7733801	RP11-404G1	NONSC12G5	bA404G1	AC025335		WIBR	FinishAc
17	8408778	8574309	RP11-12H18	NONSC27B3	bA12H18	AC011061		WIBR	FinishAc
17	9466335	9611536	RP11-208F13	NONSC8C11	bA208F13	AC026855		WUGSC	Phase1Ac
17	10055064	10253595	RP11-401O9	NONSC29A3	bA401O9	AC005291		WIBR	FinishAc
17	11938351	12108967	RP11-471L13	NONSC14B9	bA471L13	AC005244		WIBR	FinishAc
17	12920204	13085706	RP11-488L1	NONSC14E8	bA488L1	AC005303		WIBR	FinishAc
17	13211602	13416771	RP11-388F14	NONSC28H7	bA388F14	AC005375		WIBR	FinishAc
17	14364543	14365215	RP1-27J12	NONSC24A1	dJ27J12	AC006578		WIBR	FinishAc
17	15321328	15510169	RP11-385D13	NONSC12D5	bA385D13	AC005838		WIBR	FinishAc
17	16526746	16718786	RP11-219A15	NONSC8F2	bA219A15	AC022596		WIBR	FinishAc
17	17338154	17521462	RP11-524F11	NONSC29E5	bA524F11	AC020558		WUGSC	FinishAc
17	17940393	18114679	RP11-189D22	NONSC7H10	bA189D22	AC020567		WUGSC	Phase1Ac
17	19126535	19251679	RP1-162E17	NONSC31A7	dJ162E17	AL160492		MPIMG	FinishAc
17	19211694	19360796	CTB-1187M2	Cancer_3C11	bK187M2				
17	19614077	19748613	RP11-78O7	nonsc43B7	bA78O7	AC015726		WIBR	FinishAc
17	20013343	20134139	RP5-836L9	NONSC24F12	dJ836L9	AC004702		WIBR	FinishAc
17	20112029	20230301	RP11-121A13	nonsc43C12	bA121A13	AC008088		WIBR	FinishAc