Norwegian University of Life Sciences

Master's Thesis 2024 30 ECTS
Faculty of Chemistry, Biotechnology and Food Science

Characterization of Neoplasm-Specific Chromosomal Aberrations and Their Molecular Consequences in Lipomas: Identification of the *HMGA2::LINC00535 Fusion Transcript*

Inger Jean Cooper Kjønnøy
Biotechnology

ABSTRACT

Acquired genetic alterations are known to drive neoplastic transformation and tumor development. On a molecular level, chromosomal aberrations can lead to the formation of fusion transcripts, which may play a role in tumorigenesis. These aberrations and their gene products are in many cases characteristic of neoplasms and are used to subclassify tumors. They are often valuable biomarkers for diagnostic, prognostic and therapeutic purposes. The aim of this project was to increase the knowledge of chromosomal aberrations and their molecular consequences in lipomas by investigating two lipomas showing rearrangements of the 8q21-q22 and 12q14 chromosomal bands in their karyotypes. Since the breakpoints of both rearrangements map on the same chromosomal bands, it was thought that these aberrations could form the same gene products, hence this was investigated in this project.

Output from previous total RNA sequencing of the two lipomas were investigated for fusion transcripts, where the ones involving sequences from both 8q21-q22 and 12q14 chromosomal bands were of interest. Reverse Transcription Polymerase Chain Reaction (RT-PCR) and Sanger sequencing were performed to verify these transcripts and to localize the breakpoint of the fusions more precisely. Verification of the rearrangements at the genome level was performed through Fluorescent In Situ Hybridization (FISH). Furthermore, array Comparative Genomic Hybridization (aCGH) was used as an exploratory method to screen the genomes for imbalanced cryptic aberrations and thereby gain further knowledge of chromosomal aberrations in these lipomas.

Aberrations involving the 8q22 and 12q14 chromosomal bands were found to be recurrent in lipomas and led to the formation of an *HMGA2::LINC00535* fusion transcript, which has not been previously identified. If translated into a protein, a truncated High-Mobility Group AT-hook 2 (HMGA2) protein may be formed, which is already thought to play a role in the development of lipomas and other tumors. Furthermore, no imbalanced cryptic aberrations were identified in the two lipomas. Combining these findings with clinical information may contribute to distinguish benign lipomas more efficiently from malignant adipocytic tumors and thereby improve the subclassification of these tumors. This could further enhance the decision-making regarding diagnosis, prognosis and treatment of adipocytic tumors. However, further studies, involving other lipomas and adipocytic tumors, are needed to determine if the *HMGA2::LINC00535* fusion transcript is specific for lipomas.

SAMMENDRAG

Ervervede genetiske mutasjoner er kjent for å bidra til neoplastisk transformasjon og utviklingen av tumorer. På et molekylært nivå kan kromosomaberrasjoner lede til produksjonen av et fusjonstranskript, som kan være av betydning for tumorutvikling. Disse aberrasjonene, samt genproduktene de danner, er i mange tilfeller karakteristiske for neoplasmer og er brukt til å subklassifisere tumorer. De er derfor verdifulle biomarkører som blir brukt i diagnostiske, prognostiske og terapeutiske sammenhenger. Hensikten med dette prosjektet var å øke kunnskapen om kromosomaberrasjoner og deres molekylære konsekvenser i lipomer ved å undersøke to lipomer som viste strukturelle endringer i kromosombåndene 8q21-q22 og 12q14 i karyotypene sine. Siden bruddpunktet i begge kromosomendringene er på de samme kromosombåndene, var det tenkt at disse endringene kunne føre til samme genprodukt, noe som ble undersøkt i dette prosjektet.

Resultater fra tidligere RNA-sekvensering utført på total RNA fra de to lipomene ble undersøkt for fusjonstranskripter, hvor de som var av interesse involverte sekvenser fra kromosombåndene 8q21-q22 og 12q14. For å verifisere transkriptene og lokalisere bruddpunktet til fusjonene mer presist, ble polymerasekjedereaksjon med revers transkriptase og Sanger-sekvensering utført. Fluorescent In Situ Hybridization (FISH) ble utført for å verifisere de kromosomale endringene på gennivå. Videre ble array Comparative Genomic Hybridization (aCGH) benyttet til å screene genomene for kryptiske, ubalanserte strukturelle endringer. Dette ble gjort for å utforske andre mulige kromosomale endringer i lipomene.

Aberrasjoner som involverer de kromosomale båndene 8q22 og 12q14, viste seg å være tilstede i begge lipomene og har trolig ført til uttrykk av fusjonstranskriptet *HMGA2::LINC00535,* som ikke har vært identifisert tidligere. Dersom transkriptet hadde blitt oversatt til et protein, ville det mest sannsynlig blitt dannet et forkortet High-Mobility Group AT-hook 2 (HMGA2)-protein, som allerede er antatt å ha en betydning i utviklingen av lipomer og andre tumorer. Det ble ikke identifisert noen kryptiske, ubalanserte aberrasjoner i de to lipomene. Ved å kombinere disse funnene med klinisk informasjon, vil benigne lipomer kunne skilles mer effektivt fra maligne fett-tumorer, og dermed forbedre subklassifiseringen av slike svulster. Dette kan igjen føre til bedre vurderinger av diagnose, prognose og behandling av slike tumorer. Det bør likevel forskes mer på flere lipomer og andre fettsvulster for å vite om *HMGA2::LINC00535* fusjonstranskriptet er spesifikt for lipomer.

AKNOWLEDGEMENT

This thesis is the final work of my Master's degree in Biotechnology at the Norwegian University of Life Sciences (NMBU). It was conducted at the Section of Cancer Cytogenetics, Institute for Medical Informatics at The Norwegian Radium Hospital, Oslo University Hospital (OUS). I would like to express my gratitude to the people at the Section of Cancer Cytogenetics for giving me the opportunity to write my master's thesis within this interesting field of research. I would also like to thank you for the warm welcome and for making my time so memorable.

I would like to thank my main supervisor from OUS: Professor Francesca Micci for all the constructive feedback you have given regarding my thesis. I have learned a lot from you.

Especially thanks to my other supervisor at OUS: Marta Brunetti, PhD for the invaluable guidance throughout my time in the lab. You have made this process very motivating.

Also, thank you to Special Engineer Kristin Andersen for guidance regarding FISH analysis.

I would also like to thank my supervisor at NMBU: Professor Hilde-Gunn Opsahl Sorteberg, for the academic feedback regarding my thesis.

Finally, thank you to all my friends and family, especially Espen Hansen Høijord, who have motivated and supported me throughout this semester.

Oslo, May 2024

Inger Jean Cooper Kjønnøy

LIST OF ABBREVIATIONS

ABL1	*ABL proto-oncogene 1*
aCGH	array Comparative Genomic Hybridization
AT	Adenine-thymine
BCR	*Breakpoint Cluster Region*
BLAST	Basic Local Alignment Search Tool
BLAT	BLAST-Like Alignment Tool
bp	base pair
cDNA	complementary DNA
CML	Chronic Myeloid Leukemia
CNV	Copy Number Variation
C-terminal	Carboxy-terminal
ddNTPs	dideoxynucleotide triphosphates
DNA	Deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
DSB	Double-strand break
dsDNA	double-stranded DNA
FISH	Fluorescence In Situ Hybridization
HMGA2	*High-Mobility Group AT-hook 2 gene*
HMGA2	High-Mobility Group AT-hook 2 protein
HMGA2Tr	Truncated HMGA2
kb	Kilobase
LINC00535	*Long Intergenic Non-protein Coding RNA 535*
lncRNA	long non-coding RNA

LPP	*LIM Domain Containing Preferred Translocation Partner In Lipoma*
Mb	Megabase
miRNA	microRNA
NCBI	National Center for Biotechnology Information
ncRNA	non-coding RNA
NGS	Next-Generation Sequencing
nt	nucleotides
N-terminal	Amino-terminal
PCR	Polymerase Chain Reaction
qPCR	quantitative PCR
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RT	Reverse transcriptase
RT-PCR	Reverse Transcription PCR
RT-qPCR	quantitative Reverse Transcription PCR
SNP	Single Nucleotide Polymorphism
SSC	Saline-Sodium Citrate
ssDNA	single stranded DNA
T_a	Annealing temperature
TK	Tyrosine kinase
T_m	Melting temperature
UTR	Untranslated region
WGS	Whole Genome Sequencing

WHO	World Health Organization

TABLE OF CONTENTS

1 INTRODUCTION

1.1 CLASSIFICATION OF ADIPOCYTIC TUMORS

World Health Organization (WHO) classifies soft tissue tumors into different subgroups, and the classification is typically based on microscopic, immunophenotypic, cytogenetic and molecular methods (Antonescu et al., 2020). Combining these investigation methods is important as distinguishing benign and malignant tumors using solely traditional pathologic methods can be challenging and, as a consequence, may result in incorrect diagnosis and treatment (Mandahl & Mertens, 2015).

Adipocytic tumors consist of fat tissue and are classified, in accordance with the WHOs classification system of soft tissue tumors, as benign, intermediate or malignant tumors (Antonescu et al., 2020), as illustrated in Table 1. In this classification, examples of intermediate and malignant adipocytic tumors are atypical lipomatous tumors and well differentiated liposarcomas, respectively. Lipomas are classified as benign adipocytic tumors and was the focus of the project in this thesis. Gaining more knowledge about chromosomal aberrations and their molecular consequences in adipocytic tumors, including lipomas, may improve the subclassification and thereby the diagnostic accuracy of these neoplasms.

Table 1. WHO 2020 classification of adipocytic tumors (Antonescu et al., 2020).

Benign
Lipoma and lipomatosis
Lipomatosis of nerve
Lipoblastoma and lipoblastomatosis
Angiolipoma
Myolipoma of soft parts
Chondroid lipoma
Spindle cell/pleomorphic lipoma
Atypical spindle cell/pleomorphic atypical lipomatous tumor
Hibernoma
Intermediate
Atypical lipomatous tumor
Malignant
Well differentiated liposarcomas (lipoma-like, sclerosing, inflammatory)
Dedifferentiated liposarcoma
Myxoid liposarcoma
Pleomorphic liposarcoma
Myxoid pleomorphic liposarcoma

1.2 LIPOMA PATHOLOGY

Lipoma is a subcutaneous benign adipocytic tumor (Kosztyuova & Shim, 2017) and is the most common soft tissue neoplasm in humans (Mandahl & Mertens, 2015). The tumors are typically seen as encapsulated masses through medical imaging, and may be located in the regions of the neck, head, shoulder and back, and occasionally in the thighs (Antonescu et al., 2020). Furthermore, they can be found in internal organs, such as in the gastrointestinal tracts. The size of the tumors varies between 1-20 cm, however, they can be larger (Antonescu et al., 2020). Lipomas are often slow-growing and are usually noticed in adulthood (Kosztyuova & Shim, 2017). According to Kosztyuova & Shim (2017), lipomas can either occur as singular or multiple tumors, where multiple tumors are more frequently seen in men, while singular tumors are often seen in women. Treatment is usually not necessary unless the size of the tumor causes symptoms (Kosztyuova & Shim, 2017). In such cases, lipomas are removed through surgical excision (Antonescu et al., 2020).

1.3 CHROMOSOMAL ABERRATIONS

In cytogenetics, chromosomes and chromosomal changes are investigated (Kannan & Zilfalil, 2009). The accurate determination of the number of human chromosomes was established by Tjio and Levan as 46 in 1956 (Tjio & Levan, 1956 as cited in Kannan & Zilfalil, 2009). Additionally, chromosome preparation methods for leukocytes were established and adapted some years later, making it possible to describe alterations to the normal chromosome number and structure (Kannan & Zilfalil, 2009).

Chromosomal abnormalities, also known as chromosomal aberrations, can either be inherited or acquired throughout life (Heim & Mitelman, 2015). They are often categorized as structural or numerical alterations, where structural alterations refer to any changes to a specific part of a chromosome, while numerical alterations are changes to the number of chromosomes present (Jackson et al., 2018). Structural alterations are further divided into balanced and imbalanced aberrations, depending on whether a loss or a gain of the total genomic material in a cell has occurred or not (Heim & Mitelman, 2015). Alterations can be imbalanced due to gain or loss of chromosomal material caused by duplications or deletions, respectively, giving rise to Copy Number Variations (CNVs) (Pös et al., 2021). In 2004, Sebat et al. illustrated that CNVs were widespread in human genomes. Although they are a natural source of genetic variation and give rise to unique traits, they are also associated with some disorders and diseases, such as

Parkinson's disease, Alzheimer's disease and autism (Zhang et al., 2009). Balanced alterations can occur through inversions within a chromosome, or translocations or insertions between chromosomes (Heim & Mitelman, 2015). Some numerical and structural alterations are illustrated in Figure 1.

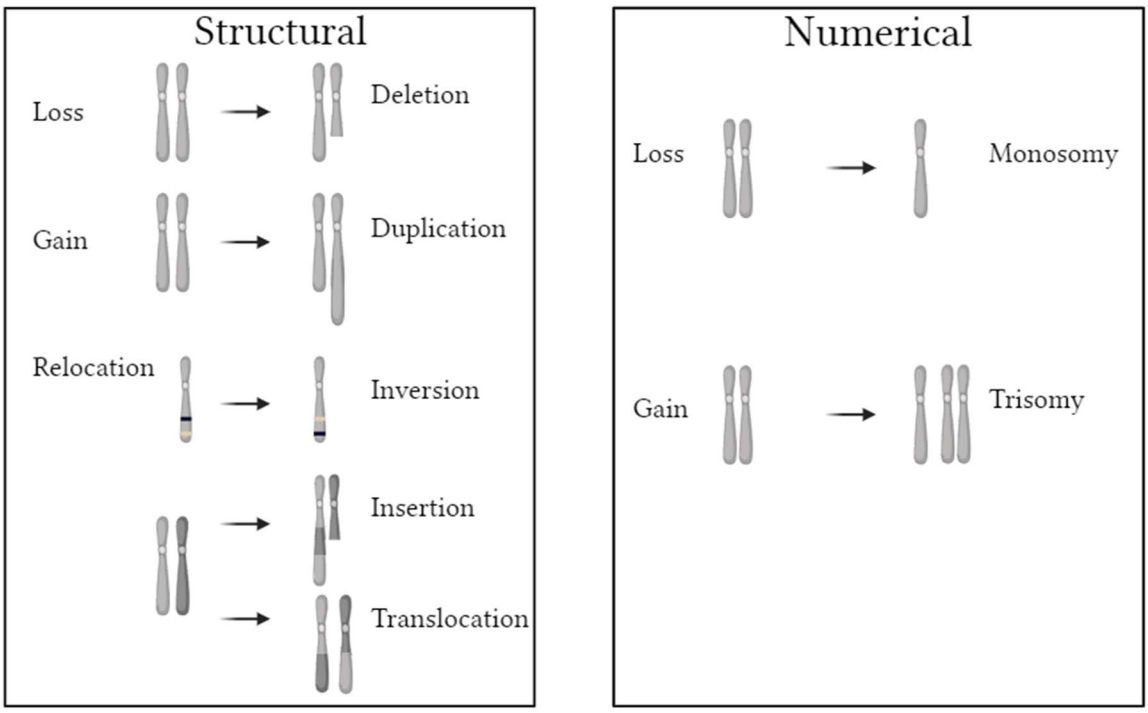

Figure 1. Structural and numerical chromosomal alterations, with loss, gain and relocation of genetic material. Created with BioRender.com.

Some chromosomal aberrations appear to be nonrandom and are associated with distinctive diseases (Mitelman & Heim, 2015). The first neoplasia-associated karyotypic abnormality to be discovered in humans was the "Philadelphia chromosome", which was discovered in 1960 in patients with Chronic Myeloid Leukemia (CML) (Nowell & Hungerford, 1960 as cited in Mitelman et al., 2007). The abnormality is seen in 90-98% of the CML patients (Sampaio et al., 2021). Since the discovery of the first neoplasia-associated karyotypic abnormality, other aberrations have been associated with a variety of neoplasms. As an example, two-thirds of investigated lipomas exhibit chromosomal aberrations, where some occur more frequently than others (Mandahl & Mertens, 2015).

Chromosomal aberrations can be classified as primary, secondary or noise. According to Heim & Mitelman (2015), primary aberrations occur in the earliest stages of tumorigenesis and are

essential for establishing a neoplasm, while secondary aberrations are found in cells that already carries a primary aberration and may be important for tumor progression. Furthermore, they describe noise as background aberrations with no identified consequences that are randomly distributed throughout the genome. Identifying chromosomal abnormalities that are pathogenically important for transformation of normal cells to neoplasms is of interest in cytogenetics (Heim & Mitelman, 2015). Although chromosomal aberrations are associated with disease, not all tumors exhibit chromosomal aberrations, and non-neoplastic cells from healthy people may have similar aberrations as those found in neoplasms (Heim & Mitelman, 2015). As an example, trisomy 7 is seen in both benign and malignant neoplasms, and also in normal tissue (Broberg et al., 2001). Since some chromosomal aberrations are found in both normal tissue and neoplasms, combining cytogenetic and molecular findings with the medical history of a patient is important to determine the pathological relevance of an aberration (Bäsecke et al., 2002).

As described in Mitelman & Heim (2015), the identification of the initial karyotypic abnormality associated with neoplasia sparked an interest in cancer cytogenetics and is now fundamental for identifying genes crucial for tumorigenesis. It is now known that most of the chromosomal rearrangements in cancer lead to deregulation of normal genes, oncogene activation or the creation of fusion genes (Mitelman & Heim, 2015).

1.4 FUSION GENES AND FUSION TRANSCRIPTS

Fusion genes are made up of two genes that are combined into a hybrid gene and often represent the primary event in tumorigenesis (Panagopoulos & Heim, 2021). More than 33 800 unique fusion genes are listed in the "Mitelman Database Chromosome Aberrations and Gene Fusions in Cancer" database (last updated 16th of January 2024) (Mitelman et al., 2024). According to Panagopoulos (2015), several fusion genes are found to be tumor-specific and are identified in mesenchymal tumors, carcinomas and hematological malignancies. Furthermore, he describes that fusion genes generated by chromosomal rearrangements are results of a deoxyribonucleic acid (DNA) double-strand break (DSB). If a DSB is not properly repaired by the cell's DNA repair mechanisms, DNA strands can be incorrectly rejoined and form a fusion gene (Panagopoulos, 2015). Some fusion genes are expressed as fusion transcripts. Fusion transcripts are either transcribed from fusion genes formed through chromosomal rearrangements, which is the most common mechanism, or created through abnormal transcription (Panagopoulos,

2015). Chromosomal rearrangements and abnormal transcription are presented in the following section.

In general, a fusion gene is created by the recombination of two genes via breakpoints within each gene, however, there are some exceptions. Breakpoints may occur between a gene and a region that is transcribed into non-coding ribonucleic acid (ncRNA), such as long ncRNA (lncRNA), microRNA (miRNA), small nucleolar RNA and circular RNA (Taniue & Akimitsu, 2021b). In these cases, it can be more correct to name them fusion transcripts, rather than fusion genes, although the terms are used interchangeably.

The numbers of identified fusion genes are increasing, as new technologies are introduced and more samples are studied. Identification of these fusions are essential for accurate subclassification and diagnosis of neoplasms (Panagopoulos, 2015).

1.4.1 Fusion Gene Generation Through Translocations and Inversions

Translocations and inversions are balanced chromosomal rearrangements and stand out as the most common genetic events leading to the formation of fusion genes (Panagopoulos & Heim, 2021).

Translocation involves the exchange of genetic fragments between chromosomes, and in some instances, they can involve almost entire chromosome arms (Annala et al., 2013). Fusion gene formation through translocation produce a chimera where the head (5'-end) of one gene is joined with the tail (3'-end) of another gene (Panagopoulos & Heim, 2022). As an example, the "Philadelphia chromosome" is generated through a translocation between chromosomes 9 and 22, leading to the t(9;22)(q34;q11) translocation (Sampaio et al., 2021). Molecular characterization performed in the 1980s revealed a fusion of the *Breakpoint Cluster Region* (*BCR)* and the *ABL proto-oncogene 1* (*ABL1)* gene, creating a *BCR::ABL1* fusion gene (Mitelman et al., 2007). The fusion creates a constitutively active tyrosine kinase (TK) protein, which inhibits apoptosis and stimulates cell division (Kang et al., 2016).

As described by Taniue & Akimitsu (2021b), inversions involve the flipping of chromosomal regions within a single chromosome. Furthermore, they describe that inversions can either be paracentric, where centromeres are not involved and the break is within one chromosome arm, or pericentric, where centromeres are involved and there is a break in each chromosome arm.

These fusions may involve the presence of fusion genes at both ends of the inversion (Annala et al., 2013).

1.4.2 Fusion Gene Generation Through Deletions, Insertions and Tandem Duplications

Fusion genes may also be formed by deletions, insertions or tandem duplications. According to Panagopoulos & Heim (2021), two genes or genomic regions can be fused through an interstitial deletion. They describe the principle as the same as for translocations, mentioned in Section 1.4.1, where the head of one gene is fused with the tail of another gene creating a chimera. In some cases, loss of gene loci between the fusion genes may occur (Panagopoulos & Heim, 2021). Submicroscopic deletions, less than 10 Megabases (Mb) (Bass, 2018), cannot be detected by cytogenetic analysis. However, they may be detected through Fluorescent In Situ Hybridization (FISH) or other molecular methods, such as array Comparative Genomic Hybridization (aCGH) or sequencing (Panagopoulos & Heim, 2021).

Insertions can either be intrachromosomal, where DNA fragments are transferred from one region to another within the same chromosome, or interchromosomal, also known as nonreciprocal nonmutual translocation, where it occurs between separate chromosomes (Taniue & Akimitsu, 2021b). In tandem duplications, a chromosomal region is duplicated and fused with a gene from the original region (Taniue & Akimitsu, 2021b).

1.4.3 Fusion Transcript Generation Through Abnormal Transcription and Splicing

Abnormal transcription may form chimeric transcripts, also known as readthrough transcripts (Panagopoulos, 2015). The independent genes are co-transcribed to form one RNA chimera and the intergenic region is spliced out as an intron. Run-off transcription emerges as the most likely mechanism driving the formation of these chimeras (Panagopoulos, 2015). The chimeric transcript may comprise sequences coding for protein domains from both of the original genes (Panagopoulos & Heim, 2021). Intergenic trans-splicing can also lead to the formation of a fusion transcript, where two separate pre-messenger RNAs form a chimeric RNA (Calabrese et al., 2020). This demonstrates that chromosomal rearrangements are not causing all cases of fusion transcript formations. However, in a study performed by Calabrese et al. (2020), 82% of

fusion transcripts were found to display evidence of genomic rearrangements, by comparing RNA sequencing (RNA-Seq) and Whole Genome Sequencing (WGS).

1.4.4 Consequences of Fusion Genes

As previously mentioned, fusion genes can arise from diverse types of chromosomal rearrangements. Consequences of fusion genes are diverse, and depends on the location of the breakpoint (Taniue & Akimitsu, 2021b). They can either be in-frame or out-of-frame. In-frame chimeras can be translated to transcripts and lead to the creation of chimeric proteins if two protein coding genes are fused together, whereas out-of-frame chimeras can be formed if the breakpoints are within protein coding regions, creating truncated transcripts (Panagopoulos & Heim, 2022).

Promoters can also be swapped, which can increase or reduce the normal expression of a gene (Panagopoulos, 2015). Often, one of the two genes involved encodes a transcription factor, and a hybrid gene will therefore create an abnormal transcription factor that can affect the regulation of downstream genes (Panagopoulos, 2015).

Chromosomal aberrations can create fusion genes encoding for fusion proteins, such as TK chimeras. As described in Panagopoulos (2015), normal TK proteins often consist of the carboxy-terminus where the TK domain is located, and an inhibitory domain at the amino-terminal (N-terminal). In a TK chimeric protein, the inhibitory domain in the N-terminal is replaced by sequences from a partner gene which, through a promoter, drives the expression of the TK domain (Panagopoulos, 2015). Fusion genes involving TKs can activate signaling pathways that are important for cell proliferation and apoptosis, and are found in various types of cancers (Taniue & Akimitsu, 2021b).

Tandem duplications are known to lead to overexpression and constitutively active kinases in neoplasms (Taniue & Akimitsu, 2021b). All these events may serve a role in tumor development.

1.5 IDENTIFICATION OF FUSION GENES AND FUSION TRANSCRIPTS

In recent years, fusion genes have been identified through Next-Generation Sequencing (NGS) methods, such as WGS or RNA-Seq, where short sequences of DNA or RNA are sequenced and aligned to a reference genome (Panagopoulos, 2015). As described by Panagopoulos

(2015), WGS can be used to identify fusion genes in the genome, regardless of whether they are transcribed or not. However, the relevance of a fusion gene relies on whether it produces a fusion transcript. RNA-Seq, on the contrary, only identifies transcribed regions and may be used to detect transcribed fusion genes as well as alternative gene spliced transcripts (Panagopoulos, 2015). In 2009, RNA-Seq was proven by Maher et al. to be an efficient method to detect fusion transcripts and has later been used in fusion transcript studies (Panagopoulos et al., 2023).

As described in Panagopoulos (2015), an RNA-Seq procedure starts with RNA extraction, followed by removal of DNA contamination. Sequentially, RNA is fragmented and reverse transcribed to complementary DNA (cDNA). Sequencing adaptors are then ligated, and fragment size is selected. The cDNA is then sequenced to produce short reads. Paired-end reads are generated if both ends of the cDNA are sequenced. Quality control is performed to remove poor-quality reads before novel transcripts are identified using bioinformatic tools (Panagopoulos, 2015). False positive and false negative findings are common when using fusion detection algorithms, and the findings should therefore be validated with other methods, such as FISH and Polymerase Chain Reaction (PCR), followed by Sanger sequencing (Carrara et al., 2013; Heim & Mitelman, 2015; Taniue & Akimitsu, 2021b).

1.6 CHROMOSOMAL ABERRATIONS AND FUSION GENES IN LIPOMAS

According to Mandahl & Mertens (2015), two-thirds of all lipomas have chromosomal aberrations. They further describe that structural rearrangements are the most common aberrations, where two-thirds of the karyotypes are balanced. For imbalanced karyotypes, partial losses are more common than gains and often occur in chromosomes 1, 6, 12 and 13, in increasing order of frequency (Mandahl & Mertens, 2015). In accordance with the WHO classification of soft tissue tumors, four chromosomal aberrations are often used to subclassify lipomas (Antonescu et al., 2020). These are rearrangement of chromosome region 12q13-15, loss of material from 13q, supernumerary ring chromosomes and rearrangement of chromosome segment 6p21. Their frequencies are two-thirds, 15%, 5% and 5%, respectively, in lipomas exhibiting rearrangements (Antonescu et al., 2020). The most frequent rearranged region in lipomas is the 12q13-15, and Mandahl & Mertens (2015) describe that some partner regions are seen more frequently than others, such as the 3q27-29, which is seen in 20% of

these cases. Other chromosomal bands that are less frequently involved are the 1p32-34, 2p22-24, 2q35-37, 5q32-34, 12p11-12 and 12q24 (Mandahl & Mertens, 2015).

The 12q13-15 chromosomal region encompasses the *High-Mobility Group AT-hook 2* (*HMGA2)* gene which is located in 12q14.3 (Mandahl & Mertens, 2015). The *HMGA2* gene is approximately 140 kilobases (kb) (Mansoori et al., 2021) and comprises five exons with a total combined length of 330 base pairs (bp) (Wang et al., 2021). Furthermore, *HMGA2* encodes for the High-Mobility Group AT-hook 2 (HMGA2) protein, which consists of 109 amino acids (Wang et al., 2021). High-mobility group proteins were first discovered in 1973 by Goodwin et al. (1973) and have later been divided into three subgroups based on their unique protein signatures (Zhang et al., 2019). The HMGA2 proteins have "adenine-thymine (AT)-hook" sequence motifs which can bind to AT-rich sites in minor grooves of B-form DNA (Mansoori et al., 2021). Subsequently, transcription may be influenced by conformational changes of the DNA, and the activity of specific genes may be altered and influence biological processes. *HMGA2* also encodes an acidic carboxy-terminal (C-terminal) domain containing regulatory parts that are important for the *HMGA2* expression and protein-protein interactions (Mansoori et al., 2021; Panagopoulos et al., 2023). An illustration of the *HMGA2* gene and the HMGA2 protein is shown in Figure 2. First three exons (exon 1-3) of the *HMGA2* gene encodes for the three AT-hooks, exon 4 encodes for a protein linker, while exon 5 encodes for the acidic C-terminal tail (Mansoori et al., 2021).

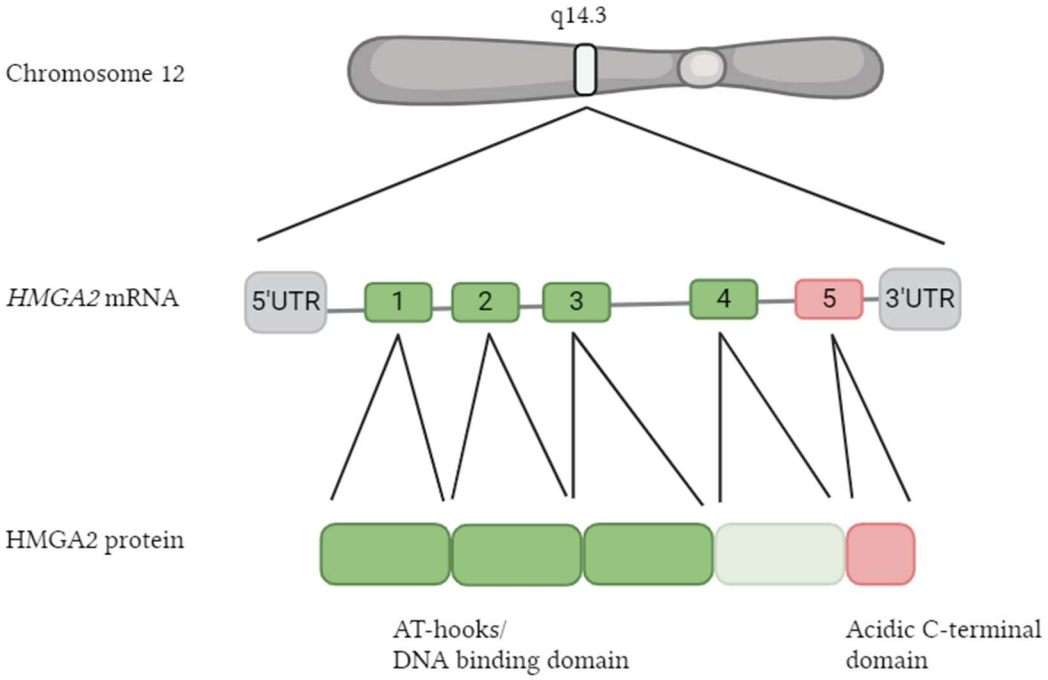

Figure 2. The five exons transcribed from the *HMGA2* gene (140 kb) (numbered 1-5) with their corresponding region in the HMGA2 protein. The AT-hooks (green) and the acidic C-terminal domain (red) in the HMGA2 protein is shown. Created with BioRender.com.

HMGA2 is expressed in mesenchymal stem cells, which are important for adipose tissue development, and may therefore be associated with lipoma development (Mandahl & Mertens, 2015). The *HMGA2* gene is silent in most tissues in adults but it is expressed during embryogenesis (Antonescu et al., 2020). In some lipomas, *HMGA2* is expressed and has shown to form fusion genes through translocations with partner genes (Antonescu et al., 2020). The "Mitelman Database Chromosome Aberrations and Gene Fusions in Cancer" database (last updated 16th of January 2024) displays that 29 unique fusion genes are identified in lipomas, whereas 21 of these involve the *HMGA2* gene (Mitelman et al., 2024). Although rare, chimeric HMGA2 proteins may be formed through such fusion events (Panagopoulos et al., 2023). In some lipomas, full-length HMGA2, involving exon 1-5, and truncated HMGA2 (HMGA2Tr), involving exon 1-3, are expressed (Mandahl & Mertens, 2015). As described by Panagopoulos (2015), HMGA2Tr proteins occur when rearrangements of the 12q13-15 chromosomal regions lead to an out-of-frame fusion of *HMGA2* with a partner gene. It results in a protein that consists of the three AT-hooks, lacking the C-terminal domain (Panagopoulos, 2015). It is suggested that full-length, truncated and fused HMGA2 serves a role in lipoma development (Antonescu et al., 2020).

1.7 FUSION GENES AS TARGETS FOR TUMOR TREATMENT

Fusion genes are known to be drivers in tumorigenesis in humans and play an important role in diagnostics, prognostics and treatment of various tumor types (Annala et al., 2013). According to Mitelman et al. (2007), it is thought that fusion genes account for 20% of human cancer morbidity. Some fusion genes tend to be tumor-specific and are therefore used as biomarkers to establish the diagnosis and prognosis of a patient (Taniue & Akimitsu, 2021b). In addition, the biomarkers can be used as targets in cancer treatment. Targeting fusion genes and transcripts that are cancer-specific is leading us closer to the goal of precision medicine, where tumor cells are targeted for treatment without affecting normal cells (Taniue & Akimitsu, 2021b).

Kinases and transcription factors are often involved in oncogenic fusion genes (Brien et al., 2019). Fusion genes that alter kinase activity are used as targets in cancer therapy and may be targeted by inhibitors. An example is the inhibition of kinase activity of the tyrosine kinase protein transcribed from the *BCR::ABL1* fusion gene in patients with CML, using imatinib (Taniue & Akimitsu, 2021b). According to Cohen et al. (2021), imatinib was approved by the U.S. Food and Drug Administration in 2001, and was the first drug to be approved of those designed to target a specific kinase. CML is no longer a rare leukemia, as patients survive longer with imatinib treatment, and it is estimated that more than 100 000 patients use imatinib daily to survive (Cohen et al., 2021). In a study performed with CML patients receiving imatinib for a period of 10 years, Hochhaus et al. (2017) found that the estimated survival rate was over 80%.

As previously mentioned, transcription factors are commonly involved in oncogenic fusion genes. In Taniue & Akimitsu (2021b), fused transcription factors are described to produce fusion proteins that can bind to the genome, change the transcription of genes and promote tumorigenesis. However, they have shown to be more difficult to use as targets in cancer therapy compared to kinases (Taniue & Akimitsu, 2021b).

It is important to remember that fusion genes may be present in normal tissue and are therefore not always ideal targets for tumor therapy (Taniue & Akimitsu, 2021b). Benign adipocytic tumors, such as lipomas, are treated with surgical excision, while malignant adipocytic tumors, such as liposarcomas, often need further treatment (Antonescu et al., 2020). Gaining more knowledge by investigating both benign and malignant tumors are therefore important to decide treatment approaches in the respective cases.

2 OBJECTIVES

Acquired genetic alterations are known to drive cellular neoplastic transformation and tumor development (Gomes, 2022; Repana et al., 2019). They are common in neoplasms, such as lipomas and other adipocytic tumors (Mandahl & Mertens, 2015; Mitelman et al., 2024). Many alterations are visible in the microscope when chromosomes condense during mitosis and can be registered as numerical or structural cytogenetic abnormalities (Jackson et al., 2018). Abnormal karyotypes provide valuable insight into the pathogenic events of tumorigenesis and the clonal evolution of neoplastic cell populations (Heim & Mitelman, 2015). At the molecular level, chromosomal aberrations can lead to altered expression of genes involved, or the formation of fusion transcripts (Mitelman & Heim, 2015). The aberrations and their gene products are in many cases characteristic of neoplasms and are used to subclassify tumors (Antonescu et al., 2020). They are thereby valuable biomarkers for diagnostic, prognostic and therapeutic purposes (Annala et al., 2013).

The aim of this project was to increase the knowledge of chromosomal abnormalities and their molecular consequences in lipomas. Although lipomas are benign (Antonescu et al., 2020), the knowledge may contribute to distinguish them more efficiently from malignant adipocytic tumors, and thereby improve the subclassification and diagnostic accuracy of these tumors.

We selected two lipomas showing rearrangements of chromosomes 8 and 12 with similar breakpoint positions in their karyotypes. Karyograms from each of the cases are presented in Figure 3. Since the gene product associated with these rearrangements has not previously been identified, we investigated these tumors on a molecular level.

Furthermore, we wanted to screen these genomes for presence of imbalanced cryptic rearrangements, which are below the chromosomal resolution level at approximately 10 Mb (Bass, 2018), by applying aCGH. This was conducted as an exploratory method to further investigate chromosomal aberrations in the selected lipomas.

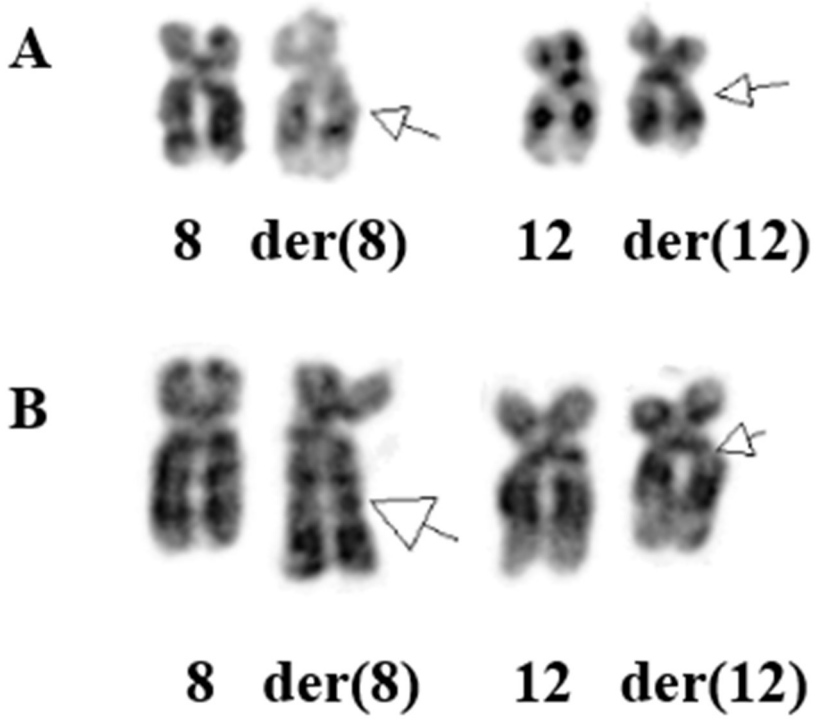

Figure 3. Karyograms of normal and derivative (der) chromosomes 8 and 12 of the two lipoma cases investigated: case 1 (A) and case 2 (B). The karyograms are from the diagnostic routine at the Section of Cancer Cytogenetics at the Norwegian Radium Hospital, Oslo University Hospital.

3 MATERIALS AND METHODS

A summary of all methods performed are briefly illustrated in flowcharts in Figure A1, A2 and A3 in Appendix A.

3.1 TUMOR SAMPLES

In this project, two tumor samples with similar chromosomal aberrations in their karyotypes, identified prior to this project, were investigated. Both samples were lipomas and collected from female patients, aged 71 and 58. The tumors were located at the thigh and flank, and were 13 x 9 x 3 cm and 7 x 4.7 x 9.5 cm in size, respectively. This study has been approved by the Regional Ethics Committee (S-0747a) and the patient information has been de-identified. The samples will hereby be named case 1 and 2, as seen in Table 2. The tumors were surgically removed and sent to cytogenetic analysis as a part of the diagnostic routine at the Norwegian Radium Hospital to detect chromosomal abnormalities by karyotyping. The karyotypic description for each case is shown in Table 2. Analysis performed in the diagnostic routine, found that both karyotypes showed a rearrangement involving chromosomes 8 and 12. In case 1, a balanced 8;12-translocation was seen, whereas in case 2 an 8;12-insertion was described. Since the breakpoints of both rearrangements map on the same chromosomal bands, namely 8q21-q22 and 12q14, it was thought that these aberrations could form the same gene products. Furthermore, case 1 showed the presence of two related clones where the stem line was characterized by the t(8;12), suggesting that this could be the primary aberration present in the neoplastic cells. This master project was created to investigate these tumors on a molecular level, as the gene product of these similar aberrations have not been previously identified.

Two lipomas showing a balanced 9;12-translocation as a sole change in their karyotype were used as controls in some of the methods. They will hereby be named control 1 and 2. Specifications about the controls are shown in Table 2.

Table 2. Clinical and karyotypic data for the two lipoma samples investigated and the two controls.

Lipoma sample	Sex/Age	Tumor location	Size (cm)	Karyotype
Case 1	F/71	Thigh	13 x 9 x 3	46,XX,t(8;12)(q21;q14)[10]/46,idem,t(5;15)(q13;q15),t(9;20)(q12;p11),del(10)(p13)[cp5]
Case 2	F/58	Flank	7 x 4.7 x 9.5	46,XX,?ins(8;12)(q22;q13q14),?t(11;16)(p15;p12)[8]/46,XX[2]
Control 1*	M/47	Thigh	4.5 x 4 x 1.5	46,XY,t(9;12)(q33;q14)[10]/46,XY[5]
Control 2*	F/68	Shoulder	7 x 6 x 4	46,XX,t(9;12)(q33;q14)[8]/46,XX[5]

* Investigated in a previous study (Panagopoulos et al., 2023).

3.2 METHODS PERFORMED IN ADVANCE OF THIS PROJECT

Total RNA extraction and RNA-Seq were performed in advance of this project. RNA concentrations and quality measurements are shown in Table B1 in Appendix B.

3.2.1 TOTAL RNA EXTRACTION

RNA extraction was performed in advance on frozen lipoma tissue of the two cases and the two controls. Total RNA from the four lipomas was extracted using the Protocol from the miRNeasy Mini Kit (QIAGEN, Hilde, Germany), TissueLyser II (QIAGEN) and QIAcube (QIAGEN). RNA concentrations of the samples were measured using a QIAxpert spectrophotometer (QIAGEN) with the A260 RNA pre-programmed method. Additionally, RNA quality of case 1 and 2 was evaluated using Agilent RNA 6000 nano assay kit from Bioanalyzer (Agilent, Santa Clara, CA, USA).

3.2.2 RNA SEQUENCING OF TOTAL RNA

For case 1 and 2, 177 ng and 244 ng of total RNA, respectively, was sent to the Genomics Core Facility at the Norwegian Radium Hospital, Oslo University Hospital (https://oslo.genomics.no/). They performed high-throughput paired-end total RNA-Seq using an Illumina sequencing instrument (Illumina, San Diego, CA, USA).

The deFuse software package (version 0.6.1) was utilized to identify putative fusion transcripts by analyzing the output generated from RNA-Seq. The software examines paired-end reads that align across a genomic fusion boundary and can identify fusions with breakpoints between

known exons, within exons, intronic and intergenic sequences. Both spliced and unspliced sequences are used as reference, as fusion genes may produce variants where intronic sequences are expressed (McPherson et al., 2011). In return, raw data containing detected fusion transcripts was provided for each case. Fusion transcripts involving both chromosomes 8 and 12 were of interest.

3.3 ALGORITHMIC DETECTION OF GENE REGIONS INVOLVED IN FUSION TRANSCRIPTS

In this project, BLAST-Like Alignment Tool (BLAT), where BLAST stands for Basic Local Alignment Search Tool, was used to identify genes and genomic regions involved in the formation of the detected fusion transcripts from deFuse. The transcripts that involved chromosomes 8 and 12, with the q21-22 and q14 chromosomal bands, respectively, were run through BLAT Search Genome from the Human BLAT Search web tool (genome.ucsc.edu/cgi-bin/hgBlat). The assembly from February 2009 (GRCh37/hg19) was used. Query type was set to "BLAT's guess" and the human genome was selected as the reference genome, against which the genes and genomic regions involved in the formation of the fusion transcripts were mapped. Furthermore, it was determined whether the fusion had occurred within introns or exons.

3.4 PCR PRIMER DESIGN

PCR primers were designed to verify if the putative fusion transcripts detected by deFuse were present in the lipoma tumors. They were designed using the Primer-BLAST web tool from the National Center for Biotechnology Information (NCBI) (ncbi.nlm.nih.gov/tools/primer-blast/index.cgi). Selecting a good primer combination is important for the PCR performance (Ye et al., 2012). Primers bind to 5'- and 3'-end of a single-stranded DNA (ssDNA) sequence of interest and are extended by DNA-polymerase activity, creating a complimentary copy of the target-sequence (Krebs et al., 2018). Designing primers that are target-specific is important to only amplify the DNA sequence of interest. In general, the primers should have similar melting temperatures (Tm), a balanced guanine-cytosine content (G-C content), not be self-complementary and be target-specific (Krebs et al., 2018; Ye et al., 2012). The melting temperature (T_m) is used to set the annealing temperature (T_a) for the primers, and is often 3-5°C lower than the T_m (Thermo Fisher Scientific, n.d.-a). The length of the PCR product should be between 100-500 bp (Krebs et al., 2018).

Detected fusion transcript sequences from deFuse were used as templates. These sequences represent cDNA sequences derived from total RNA. In Primer-BLAST, the "Primers for target on one template" folder was selected. Considering the predicted breakpoint of the fusion transcripts, the position range for the forward and reverse primers was chosen. The aim was to design primers that created a PCR product encompassing the breakpoint. PCR product size was then set to default, ranging from 70 to 1000 nucleotides (nt). The criteria for primer melting temperatures were between 58-65 ℃, with a maximum difference of 1 ℃ between the forward and the reverse primers. Primer sizes were selected to be between 15 and 25 nt.

The first eight primers shown in Table 3 were designed and selected to test for fusion transcripts, producing PCR products approximately 200-400 nt in length. Primers were ordered from Thermo Fisher Scientific (Thermo Fisher Scientific, Waltham, MA, USA). Their location in each putative fusion transcript sequence detected with deFuse is displayed in Appendix C. Followingly, primers were diluted to make a primer stock solution of 100 µM using nuclease-free water. They were subsequently incubated overnight at room temperature on a rotating platform to get a homogenous mixture. Primers were then diluted to a concentration of 10 µM and 3.2 µM which were used for PCR and Sanger sequencing.

Table 3. Primers used for PCR and Sanger sequencing. Primers in bold letters will be the focus later in this thesis.

Designation*	**Sequence (5'->3')**	**Length (nt)**	**T_m (ºC)****	**Reference sequence: Position**
HMGA2-853FW (F)	CAGCGCCTCAGAAGAGAGGACG	22	64.85	NM_003483.6: 853-874
8q22-intron-SEQ2-94333331-Rev2 (R)	GCATACACTGTGGCTGGTGGTTG	23	64.30	ENST00000520096.5: 99806-99828
HMGA2-947FW (F)	AGGCAGCAAAAACAAGAGTCCC	22	61.60	NM_003483.6: 947-968
8q22-intron-SEQ3-94333290-Rev3 (R)	GACATTCTGGACCAGGTAGAAGAGA	25	61.33	ENST00000520096.5: 99845-99869
Chr12-66232065-SEQ1-Fw1 (F)	GGGCGAGGGGTTGCATAGATA	21	61.72	ENST00000403681.7: 13853-13873
8q22-intron-94310329-SEQ4-Rev4 (R)	AGAGTGCACATTTTGGGCTGTG	22	61.91	ENST00000520096.5: 122809-122830
HMGA2-985F (F)	TTGCAGAAAGCAGAAGCCAC	20	59.33	ENST00000403681.7: 14081-14100
8q22-intron-seq2-94333341-R (R)	GCTGGTGGTTGATGCCATTATT	22	59.57	ENST00000520096.5: 99818-99839
ABL1-185F (F)	ATGACCCCAACCTTTTCGTTGCA	23	63.40	NM_007313.3: 1516-1538
ABL1-325R (R)	TAGTTGCTTGGGACCCAGCCTTG	23	64.86	NM_007313.3: 1678-1656

* F: forward; R: reverse.

** Tm: melting temperature.

3.5 REVERSE TRANSCRIPTION PCR

PCR is utilized to amplify DNA sequences of interest and is used both in diagnostic and research settings (Ye et al., 2012). The amplification is typically carried out through thermal cycling steps (Fletcher & Hickey, 2013). First, the method includes an incubation step where double-stranded DNA (dsDNA) is denatured into ssDNA. Next, a cycling step where dsDNA is denatured, primers anneal to the ssDNA template and an extension of the primers by DNA-polymerase activity is performed.

In cases where there is a need to investigate RNA rather than DNA, RNA needs to be converted into cDNA by reverse transcriptase (RT) enzymes (Krebs et al., 2018). In Krebs et al. (2018), coupling a reverse transcription reaction with PCR is referred to as a Reverse Transcription PCR (RT-PCR). The cDNA synthesis may be performed during a one-step PCR, or it could be performed in advance of a conventional PCR. A series of thermal cycling steps are performed

to make cDNA, including an incubation step for primer annealing, then a reverse transcription step, where RNA is transcribed to cDNA by RT, and a step for inactivating the enzyme to stop cDNA synthesis (Thermo Fisher Scientific, n.d.-b). The following sections describes how RT-PCR was performed in this project.

3.5.1 cDNA Synthesis

RT-PCR was carried out by first performing cDNA synthesis to transcribe total RNA to cDNA. The cDNA was prepared using the protocol from the iScript™ Advanced cDNA Synthesis Kit for RT-qPCR (Bio-Rad, Hercules, CA, USA), although this was not a quantitative PCR (qPCR).

A cDNA synthesis reaction mixture was prepared for the two investigated cases and the two controls. Reagents in the reaction mixtures are shown in Table 4. Total RNA concentration for case 1 and 2 was 17.7 ng/µl and 24.4 ng/µl, respectively. The total RNA concentration for control 1 and 2 was 16.8 ng/µl and 30.1 ng/µl, respectively.

Table 4. Reagents in one cDNA synthesis reaction mixture prepared for synthesis reaction.

Reagents	**Volume (µl)**
iScript™ Advanced Reaction Mix (5X)	4
iScript™ Advanced Reverse Transcriptase	1
Template (total RNA)	10
Nuclease-free water	5
Total	20

Synthesis reaction was then performed using a C1000 Thermal Cycler (Bio-Rad). Table 5 shows the thermal cycler program used for the cDNA synthesis.

Table 5. Thermal cycler program used to perform cDNA synthesis. Temperature and incubation time for different stages are displayed.

Parameter	Thermal cycler stages			
	Incubation	Reverse transcription	RT* inactivation	Hold
Temperature	25 °C	42 °C	85 °C	12 °C
Time (mm:ss)	05:00	30:00	05:00	∞

* Reverse transcriptase (RT).

3.5.2 CONVENTIONAL PCR

Several conventional PCRs were performed to identify putative fusion transcripts, using five different primer combinations. Primer combinations are shown in Table D1 in Appendix D and specifications for each primer is shown in Table 3. The HMGA2-853FW and 8q22-intron-SEQ2-94333331-Rev2 primer combination will be the focus in this thesis. In addition, a ABL1-185F and ABL1-325R primer combination was used to amplify transcripts from the *ABL1* housekeeping gene, as a control of presence and quality of cDNA in the samples. Both primer combinations are in bold letters in Table 3. The five PCRs and the *ABL1* PCR were performed on case 1 and 2. Some of the PCRs were then performed on control 1 and 2, as shown in Table D2 in Appendix D. The controls cDNA was also quality tested with the *ABL1* PCR.

PCR mixtures containing different primer combinations were prepared in PCR microtubes and the volume of the reagents used are shown in Table 6. General reaction mixture recommendations for the Premix Ex Taq™ HS (Takara Bio Inc., Otsu, Shiga, Japan) DNA-polymerase were followed. As total RNA was converted to cDNA, the cDNA concentrations used for the PCRs were assumed to be 8.85 ng/µl, 12.2 ng/µl, 8.4 ng/µl and 15.05 ng/µl for case 1, case 2, control 1 and control 2, respectively. Nuclease-free water was used as an internal negative control for the reactions. The microtubes were then briefly vortexed and spun down before placed in a C1000 Thermal Cycler (Bio-Rad).

Table 6. Overview of PCR reaction mixture for one PCR reaction.

Reagents	**Volume (µl)**
Nuclease-free water	9.5
Forward primer (10 µM)	1
Reverse primer (10 µM)	1
Premix Ex Taq™ HS (2X)	12.5
Template (cDNA)	1
Total	25

Different PCR programs were used for different primer combinations. The T_a for each primer combination was set 2-5 °C below their T_m. The PCR programs used for the HMGA2-853FW with 8q22-intron-SEQ2-94333331-Rev2 and ABL1-185F with ABL1-325R primer combinations are shown in Table 7 and Table 8, respectively. PCR-programs used for the remaining primer combinations are shown in Table E1 and E2 in Appendix E.

Table 7. PCR program used with the HMGA2-853FW and 8q22-intron-SEQ2-94333331-Rev2 primer combination. Temperature and time for different PCR stages are displayed.

Parameter	PCR stages				
	Incubation	35 cycles			Hold
		Denaturing	Annealing	Extension	
Temperature	96 ℃	96 ℃	60 ℃	60 ℃	12 ℃
Time (mm:ss)	01:00	00:20	00:30	04:00	∞

Table 8. PCR program used with the ABL1-195F and ABL1-325R primer combination. Temperature and time for different PCR stages are displayed.

Parameter	PCR stages				
	Incubation	35 cycles			Hold
		Denaturing	Annealing	Extension	
Temperature	96 ℃	96 ℃	55 ℃	60 ℃	12 ℃
Time (mm:ss)	01:00	00:20	00:30	04:00	∞

3.6 GEL ELECTROPHORESIS

Gel electrophoresis was performed to assess if an amplification of the target fusion transcripts and the *ABL1* housekeeping transcript had occurred during the PCR reactions. The method relies on separating negatively charged DNA by size when sequences migrate through an agarose gel, under the influence of an electric field (Rana & Joshi, 2023).

Agarose gels (1%) were made by mixing Faster Better LB 20X Lithium boric acid ultralow-conductive medium buffer (Thermo Fisher Scientific), which was diluted with distilled water to a 2X concentration, with certified molecular biology agarose (Bio-Rad). The mixture was heated in a microwave at 800 Watt until the agarose had dissolved and then poured into a tray containing a 12 well comb. It was left at room temperature to solidify. The tray containing the gel was placed in a buffer chamber using the Owl™ EasyCast™ B1A Mini Gel Electrophoresis Systems (Thermo Fisher Scientific) filled with 250 ml of the same buffer type used to create the gel. The well comb was then removed.

Electrophoresis was performed on PCR products, and a GeneRuler 1 kb Plus DNA Ladder (0.5 µg/µl) (Thermo Fisher Scientific) was used as a DNA marker. Before loading the gel, marker and PCR products were separately mixed with nuclease-free water and DNA Gel Loading Dye (6X) (Thermo Fisher Scientific) containing GelRed (Biotium, Hayward, CA, USA) making the final concentration of the loading dye to be 1X. Reagent volumes used for each sample is shown

in Table F1 in Appendix F. The marker and the samples were loaded onto the gel and ran for approximately 20 minutes at 180 V. Gel imaging was then performed using the UV transilluminator in SMART5 Gel Documentation System (VWR, Radnor, PA, USA).

3.7 SANGER SEQUENCING

Sanger sequencing was performed on PCR products to verify the presence of the putative fusion transcript as well as to characterize the breakpoint position. The principle is described in Krebs et al. (2018), and the method requires amplified DNA, primers, DNA-polymerase, deoxynucleotide triphosphates (dNTPs) and dideoxynucleotide triphosphates (ddNTPs) labeled with distinct fluorophores corresponding to each of the four nucleotide types. While primers bind to corresponding ssDNA templates and DNA polymerases synthesize new DNA, ddNTPs, lacking the 3' hydroxyl group needed to attach the next nucleotide, are randomly incorporated and terminates the synthesis reaction. Labeled DNA fragments are separated by size by electrophoresis through a capillary gel. Simultaneously, a laser beam hits the fragments and a detector records which ddNTPs terminated each fragment. As the fragments are size-separated and each nucleotide type is tagged with unique fluorophores, the DNA sequence can be precisely determined (Krebs et al., 2018). Sequences between 100-1000 bp may be generated (Crossley et al., 2020).

The following method was used to perform Sanger sequencing. All remaining PCR products were purified prior to the sequencing reaction to remove unincorporated dNTPs, unbound primers, salts and enzymes from the PCR reactions (Crossley et al., 2020). The purification was carried out using the Protocol for the MinElute® PCR Purification Kit (QIAGEN). One volume of PCR product was mixed with five volumes of Buffer PB (QIAGEN). The samples were then purified according to the manufacturer's recommendations, but with some modifications. Collection tubes were replaced when flowthroughs were discarded to minimize the contamination of reagents between the steps. To eluate the DNA, 20 µl of nuclease-free water was added to the center of the membranes. The columns were incubated for one minute at room temperature and then centrifuged to collect the eluate.

Purified DNA was used as template for sequencing reactions. These were carried out using the BigDye™ Terminator v1.1 Cycle Sequencing Kit user guide (Thermo Fisher Scientific, 2016). Table 9 shows the reagents used for each sequencing reaction. Table G1 in Appendix G shows the DNA concentration and purity for each PCR product used as template, measured using a

QIAxpert spectrophotometer (QIAGEN). According to the user guide, the optimum absorbance ratio A_{260}/A_{280} is between 1.8 and 2.0. The negative PCR controls were also sequenced to eliminate contamination in the PCR and Sanger sequencing steps. Same volume as for their corresponding samples were used. For the reaction mixture tubes, forward and reverse primers were separately added, using the same primers as for the PCRs. The microtubes were vortexed and spun down before the cycle sequencing, which was performed on a C1000 Thermal Cycler (Bio-Rad). PCR program used for the cycle sequencing is displayed in Table 10.

Table 9. Overview of the sequencing reaction mixture for one sample.

Reagents	**Volume (µl)**
BigDye™ Terminator v1.1 Ready Reaction Mix	8
Forward/Reverse primer (3.2 µM)*	2
Nuclease-free water**	9
Template (PCR product)**	1
Total	20

* Forward and reverse primers were added to separate sequencing reaction mixtures. The same primers as for the PCRs were used. ** For samples with dsDNA concentrations below 10 ng/µl: 2 µl template and 8 µl nuclease-free water was used.

Table 10. PCR program used to perform cycle sequencing prior to the Sanger sequencing. Temperature and time for different PCR stages are displayed.

Parameter	PCR stages				
	Incubation	35 cycles			Hold
		Denaturing	Annealing	Extension	
Temperature	96 °C	96 °C	55 °C	60 °C	12 °C
Time (mm:ss)	01:00	00:20	00:30	04:00	∞

Purification of the PCR products from the sequencing reactions was performed using the BigDye™ Terminator v1.1 Cycle Sequencing Kit user guide (Thermo Fisher Scientific, 2016). This was performed to remove salts and unincorporated nucleotides. In a MicroAmp™ Optical 96-well Reaction Plate with Barcode (Thermo Fisher Scientific), 10 µl of sequencing reaction product, 45 µl of SAM™ Solution (Thermo Fisher Scientific) and 10 µl of BigDye™ Xterminator™ Solution (Thermo Fisher Scientific) were added. The plate was sealed using MicroAmp™ Clear Adhesive Film (Thermo Fisher Scientific), vortexed for 20 minutes at 2000 rpm on a MixMate™ (Eppendorf, Hamburg, Germany) and spun down for two minutes at 1000

x g. Finally, the adhesive film was removed and replaced by a Septa for 96-Well Plates for 3500/SeqStudio™ Flex (Thermo Fisher Scientific). The plate was then placed in the Applied Biosystems™ SeqStudio™ Genetic Analyzer (Thermo Fisher Scientific), where the MediumSeq BDX run module was selected. Output from Sanger sequencing was analyzed using the chromatogram viewer Chromas (version 2.6.6) and BLAST (NCBI).

3.8 FLUORESCENCE IN SITU HYBRIDIZATION

FISH was performed on tumor cells from case 1 and 2 to validate structural aberrations found by karyotyping that may play a role in the formation of the detected fusion transcripts. The method relies on hybridization of a locus specific probe tagged with fluorophores to complementary target ssDNA, which can be visualized through fluorescence microscopy (Wan, 2017).

An *HMGA2* Break-apart Probe MPP16360 (CytoCell, Milton, Cambridge, UK) with the target chromosomal region of 12q14.3, encompassing the *HMGA2* gene, was used. The probe mix contained two different probes labeled with red fluorophores, placed upstream of the gene, and three labeled with green fluorophores placed downstream of the gene. The red and green fluorochromes are located approximately 200 kb apart from one another. The *HMGA2* gene is located between these, occupying around 140 kb (Mansoori et al., 2021). The regions of the 12q14.3 chromosomal band covered by red and green fluorescence probes are shown in Figure 4. Since the probes are labeled with both red and green fluorophores, a yellow fluorescence signal is expected in intact *HMGA2* genes, and a separate red and green fluorescence signal is expected if a rearrangement has occurred.

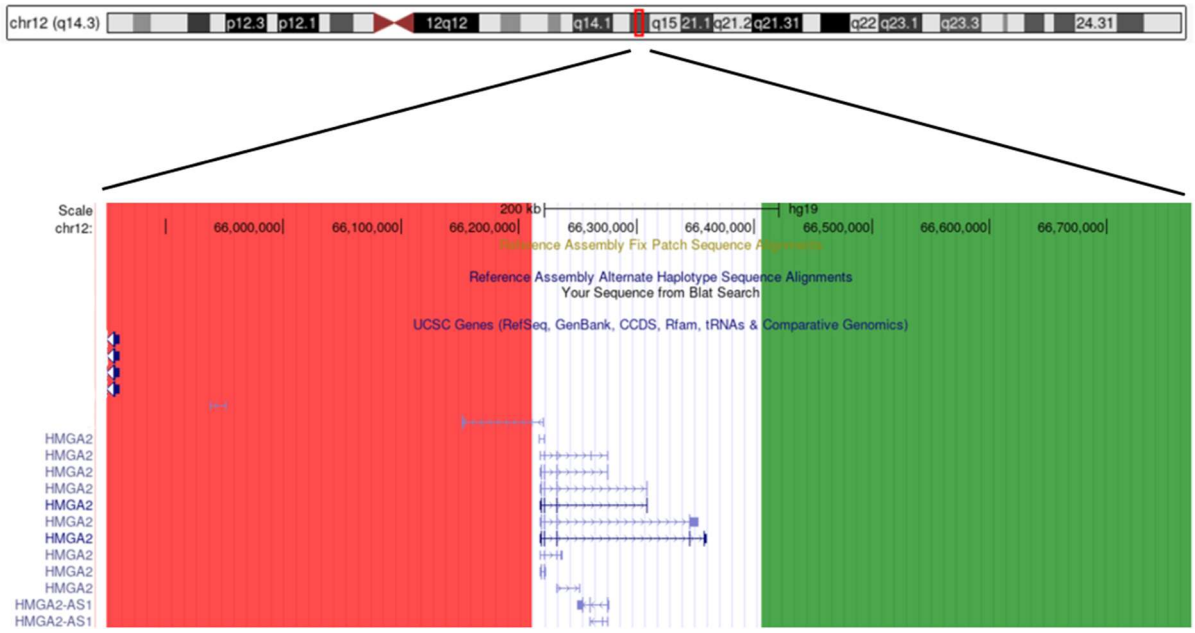

Figure 4. Illustration of genomic regions covered by the red fluorescence probes (red) and the green fluorescence probes (green) in the 12q14.3 chromosomal band, using the *HMGA2* break-apart probe. *HMGA2* is located between these areas. Created with screenshots from UCSC Genome Browser on Human (GRCh37/hg19).

The following method was performed to prepare tumor cells from case 1 and 2 for FISH analysis. A cell suspension from fixed cells used for karyotyping was used. The suspension included both interphase and metaphase cells. For each case, two drops of cell suspension were applied on a SuperFrost Microscope Slide (Epredia, Kalamazoo, MI, USA) using a glass Pasteur pipette. When dried, the cell density was assessed under a light microscope. Applying cell suspension to the slide was repeated until the cell density was high enough, without overlapping cells.

The slide preparations were then pretreated with a short, preprogrammed procedure using ThermoBrite Elite (Leica, Wetzlar, Germany) to remove excess cytoplasm and cell components. All the steps were performed under a temperature condition of 23°C. They were first incubated with 2X Saline-Sodium Citrate (SSC) for two minutes. The solution was prepared in advance of this project from a 20X SSC solution made by mixing 175.3 g Sodium chloride (Merck, Darmstadt, Germany) and 88.2 g Trisodium citrate dihydrate (Merck) with 800 ml of MilliQ water, adjusting the pH to 7. Subsequentially, a two-minute series of ethanol incubation steps with 70%, 85% and 100% ethanol (Leica) was performed to remove excess liquid from the slides. The slides were then left to dry to remove residual ethanol.

Followingly, 10 µl of *HMGA2* Break-apart Probe MPP16360 (CytoCell) was applied to the slides. They were then covered with 22 x 22 mm Menzel-Gläser cover slips (Epredia) and edges were sealed with Rubbercement (Royal Talens, Apeldoorn, Netherlands) to prevent access of air. Slides were placed in ThermoBrite Elite (Leica) and a 16 h long incubation was performed to denature dsDNA and probes, and for hybridization of the probes to ssDNA targets. The denaturing step was carried out in a demineralized water bath at 75℃ for two minutes followed by a hybridization step at 37℃ overnight.

The cover slips were then removed, and the slides underwent a post washing procedure to remove unbound and non-specifically bound probes. The slides were incubated in Wash Buffer V (LK-141C) (10X SSC, 0.5% Igepal) (Leica) at 62℃ for six seconds and 72℃ for two minutes. Followingly, the slides were cooled down to 25℃ and incubated with Wash Buffer V (LK-141C) (10X SSC, 0.5% Igepal) (Leica) at 25℃ for two minutes, followed by a one-minute series of ethanol washing steps with 70%, 85% and 100% ethanol (Leica) to remove excess liquid.

Furthermore, staining cell nuclei for interphase and metaphase cells to become visible in fluorescence microscope was done by applying 10 µl of DAPI Antifade ES (CytoCell) to the slides. They were then covered with a 22 x 22 mm Menzel-Gläser cover slip (Epredia). Slides were put in a light proof box and incubated at -20℃ for 30 minutes. Cells were examined using a Nikon ECLIPSE Ni (Nikon, Minato City, Tokyo, Japan) microscope with a light source and excitation and emission filters to filter out the wavelengths required to excite the fluorochromes. Excitation wavelengths used to excite the green and red fluorophores were set to 495 nm and 596 nm, respectively. The emission wavelength used for the green fluorophores was 521 nm and it was 615 nm for the red fluorophores.

3.9 ARRAY COMPARATIVE GENOMIC HYBRIDIZATION

Case 1 and 2 were examined for imbalanced cryptic aberrations, such as deletions and duplications, using aCGH. Since the method has a higher resolution than karyotyping, which has a resolution of about 10 Mb, it can be used to detect cryptic chromosomal aberrations (Bass, 2018; Krebs et al., 2018). As described in Krebs et al. (2018), array chips covered with sequences, representing the entire human genome, is typically used. They further describe that DNA from an investigated sample and a reference is labeled with two different fluorophores, then mixed to an equal total DNA ratio and hybridized to the chip. Image analysis is then carried

out computationally to quantitate color changes on each array spot to determine if the investigated sample has a loss or gain of a specific sequence (Krebs et al., 2018).

The method was performed on tissue from the original sample used in RNA-Seq, RT-PCR and Sanger sequencing and tissue from a new location of the tumor. It was necessary due to the limited amount of available tumor tissue used in the other methods. Additional tissue from the two tumors was obtained from the Pathology Department at the Norwegian Radium Hospital, Oslo University Hospital (https://www.oslo-universitetssykehus.no/avdelinger/klinikk-for-laboratoriemedisin/avdeling-for-patologi/seksjon-radiumhospitalet/).

3.9.1 DNA EXTRACTION

DNA extraction of the two cases was performed following the Protocol for preparation of tissue samples and automated DNA purification from Maxwell RSC Tissue DNA Kit (Promega, Madison, WI, USA) and by using a Maxwell RSC Instrument (Promega). The extraction steps were carried out according to the manufacturer's recommendations, with some modifications made to increase the DNA yield. The modifications were inspired by the DNA isolation method used by Tap et al. (2011) when they successfully performed aCGH on lipoma tumors. Briefly, 0.5 cm^3 of lipoma tissue was mixed with 60 µl of TE buffer and 20 µl of proteinase K from the kit and disrupted using TissueLyser II (QIAGEN) and a 7 mm stainless steel bead (QIAGEN). The mixture was then incubated on an Eppendorf Thermomixer C (Eppendorf) at 56°C and 750 rpm overnight to dissolve the tissue. The Maxwell Automated DNA Purification Protocol (Promega) was then carried out according to the manufacturer's instructions, except that the DNA was eluted in 50 µl of nuclease-free water rather than in 100 µl elution buffer. Reduced volume was utilized to increase the concentration of eluted DNA and nuclease-free water was added in case the DNA concentration was too low, necessitating a DNA up-concentration procedure. The method was performed several times to reach a recommended DNA concentration of 1000 ng, according to the CytoSure™ Array Handbook (4x44k and 4x180k formats) (Oxford Gene Technology, 2015). The handbook also suggests maintaining DNA purity with A_{260}/A_{280} and A_{260}/A_{230} ratios of 1.8 and >1.5, respectively. DNA concentrations and purities were measured using a QIAxpert spectrophotometer (QIAGEN) with the A260 dsDNA pre-programmed method.

3.9.2 Array Comparative Genomic Hybridization

The aCGH method was performed according to the manufacturer's recommendations in the CytoSure™ Array Handbook (4x44k and 4x180k formats) (Oxford Gene Technology, 2015), using the CytoSure™ Genomic DNA Labelling Kit (Oxford Gene Technology, Oxfordshire, UK). Reagents not provided in the kit were: SureSeq™ COT Human DNA (Oxford Gene Technology), Agilent Oligo aCGH/ChIP-on-chip Hybridization Kit (Agilent) and Agilent Oligo aCGH/ChIP-on-Chip Wash Buffer 1 and 2 (Agilent).

Of extracted DNA, 18 µl from case 1 (26.2 ng/ µl) and 2 (27.9 ng/ µl) was used. Human Genomic DNA: Female (170 ng/µl) (Promega) was used as a reference to compare the two cases with a normal genome. Based on the concentration of DNA used for case 1 and 2, calculations were done to dilute the references to concentrations that matched these. The final DNA concentrations for the references were 26.4 ng/µl and 28.3 ng/µl used for case 1 and 2, respectively.

Samples and their respective references were mixed and hybridized to an array chip covered with 180 000 sequences, representing the entire human genome. The CytoSure™ Array Handbook (Oxford Gene Technology, 2015) suggests that the incubation time of the chip should be increased from 22 h to 40 h if it contains Single Nucleotide Polymorphism (SNP) probes. In this case, custom designed CytoSure™ Cancer + SNP Arrays (Oxford Gene Technology) were used, and the slides were incubated at 67°C for 40 h in a hybridization oven.

Array slides were then scanned using Agilent SureScan Dx Microarray Scanner (Agilent) with the Agilent Microarray Scan Control (version 9.1.15.0) program. Output was converted from image to text file using Agilent Feature Extraction (version 12.2.0.7) program and then analyzed with CytoSure™ Interpret Software (version 4.11.39) to identify CNVs. The program compares the probe signal intensity from the sample with the reference signals and normalizes it with $\log_2$. A ratio of 0 indicates no CNV in the genome of the investigated case. Positive numbers indicate gain of chromosomal material in the genome, while negative numbers indicate loss of chromosomal material. The cut-off value for the mean log ratio was set to a minimum of absolute 0.5, and only genomic imbalances larger than 0,5 Mb were investigated as CNVs. Additionally, a minimum of 5 probes should cover the genomic area of a detected loss or gain, increasing the reliability of the results (Brunetti et al., 2022; Schoumans et al., 2016). CNVs detected close to the centromeres and telomers were carefully evaluated. The resolution of the

exon targeted genes, whole genes and genomic backbone were 1 probe every 1.2 kb, 1 probe every 23 kb and 1 probe every 36 kb, respectively.

4 RESULTS

4.1 IDENTIFIED PUTATIVE FUSION TRANSCRIPTS

The output from deFuse showed 688 detected fusion transcripts for case 1 and 1050 for case 2. Based on the karyotypic information, the fusions of interest should include sequences mapping on chromosome 8, band q21-22, and chromosome 12, band q14. The deFuse output showed presence of three such transcripts for case 1 and two for case 2. Table 11 shows a shorter part of the fusion transcript sequences of interest and the possible breakpoint within each of these. Additionally, it shows that the fusions are between the *HMGA2* gene and the *Long Intergenic Non-protein Coding RNA 535* (*LINC00535*), also known as *CIBAR1 Divergent Transcript* (*CIBAR1-DT*). Through BLAT it was found that exon 3 of *HMGA2*, which is located on 12q14.3, and intron 4 of *LINC00535*, which is located on 8q22.1, were involved in the breakpoint of almost all the five putative fusion transcripts. Figure H1 and H2 in Appendix H shows some of the output from deFuse for case 1 and 2, respectively, that was analyzed to identify the fusion transcripts of interest. The putative fusions from deFuse were validated with PCRs and the PCR products were assessed through agarose gels.

Table 11. Overview of the fusion transcripts identified with RNA-Seq and deFuse for case 1 and 2, where both chromosome regions 8q21-22 and 12q14 are involved. Bolder letters mark sequences from chromosome 12.

Lipoma sample	Transcript nr.	Transcript sequence (5'->3')	Chromosomes involved	Genes: exon/intron	Accession number
Case 1	1	**CAGACCTAGGAAATG**|GGCCCCTGATGTCAC	12	*HMGA2*: exon 3	NM_003483.6
			8	*LINC00535*: intron 4	ENST00000520096.5
Case 1	2	**CAGACCTAGGAAATG**|GGCCCCTGATGTCAC	12	*HMGA2*: exon 3	NM_003483.6
			8	*LINC00535*: intron 4	ENST00000520096.5
Case 1	3	**GACCTAGGAAATGGG**|CCCCTGATGTCACCA	12	*HMGA2*: intron 2/ exon 3	ENST00000403681.7
			8	*LINC00535*: intron 4	ENST00000520096.5
Case 2	1	**CAGACCTAGGAAATG**|GGCCCCTGATGTCAC	12	*HMGA2*: exon 3	NM_003483.6
			8	*LINC00535*: intron 4	ENST00000520096.5
Case 2	2	**GACCTAGGAAATGGC**|AGATCGTATAAGACT	12	*HMGA2*: exon 3	NM_003483.6
			8	*LINC00535*: intron 5	ENST00000520096.5

4.2 PCR PRODUCTS VERIFIED THROUGH AGAROSE GEL IMAGING

Agarose gel images of PCR products that were relevant for Sanger sequencing results are shown in this section. Gel images for the additional primer combinations are shown in Figure I1 and I2 in Appendix I.

4.2.1 PCR PRODUCTS USING THE HMGA2-853FW AND 8Q22-INTRON-SEQ2-94333331-REV2 PRIMER COMBINATION

Agarose gel image from electrophoresis of PCR products generated with the HMGA2-853FW and 8q22-intron-SEQ2-94333331-Rev2 primer combination for the two cases and controls are shown in Figure 5. The expected PCR product size was 331 bp, as seen in Table D1 in Appendix D, and should include sequences from *HMGA2* exon 3 and *LINC00535* intron 4. Figure 5 shows that the PCR control (nuclease-free water in lane 3) was negative, indicating no contamination of the samples. Additionally, case 1 (lane 4) and 2 (lane 5) had DNA bands between 300-400 bp, which indicates a successful amplification of the potential fusion transcript. Case 2 has a faint band. However, the PCR and gel electrophoresis were performed multiple times giving the same results. Figure I2 and I3 in Appendix I shows other gels where case 1 (lane 1) showed a band with higher intensity than case 2 (lane 2) for the same primer combination. Another primer combination targeting the same breakpoint showed the same intensity of the bands, as seen in case 1 (lane 4) and case 2 (lane 5) in Figure I2 and I3 in Appendix I. This indicates that case 2 most likely has a lower concentration of the PCR product than case 1, rather than contamination.

Conversely, no PCR product is shown for control 1 (lane 1) and 2 (lane 2) in Figure 5. The controls had the balanced 9;12-translocation as their sole change in their karyotype, generating a fusion between the *HMGA2* gene and the *Gelsolin* (*GSN)* gene (Panagopoulos et al., 2023). Since the primers here are designed to target the fusion between *HMGA2* and *LINC00535*, amplification of the putative fusion transcript was not expected in the controls. Overall, the gel shows that a potential fusion transcript was present in case 1 and 2, which have the specific chromosomal aberrations investigated.

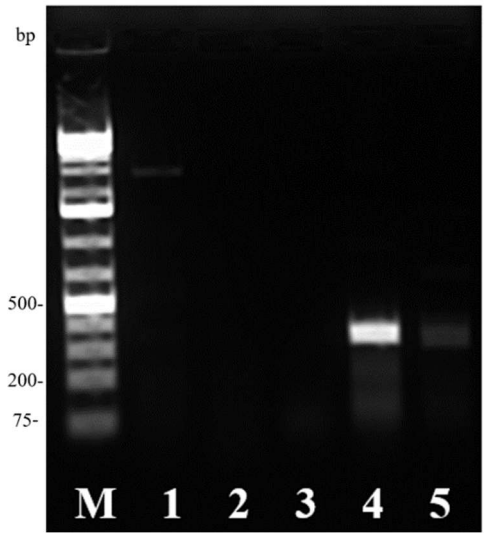

Figure 5. Agarose gel (1%) image of PCR products from case 1, 2 and controls run with the HMGA2-853FW and 8q22-intron-SEQ2-94333331-Rev2 primer combination. M: 1 kb DNA marker, 1: control 1, 2: control 2, 3: negative control for the PCR (nuclease-free water), 4: case 1, 5: case 2.

4.2.2 PCR Products Using the ABL1-195F and ABL1-325R Primer Combination

Amplification of the *ABL1* transcript for the two cases and the two controls using the ABL1-195F and ABL1-325R primer combination is shown in Figure 6. The expected PCR product size was 163 bp. Although some bands are more diffuse than others, both cases and controls displayed PCR products less than 200 bp, which most likely is the amplified *ABL1* transcript. As shown in Figure 6, case 2 (lane 5) has a band with higher intensity than case 1 (lane 4), which is the opposite of what is seen in Figure 5. However, Figure 5 and Figure 6 show products from two separate PCRs with different targets and the amount of amplified transcripts is not necessarily expected to be the same, as they may be expressed at different levels. In summary, the *ABL1*-PCR illustrates the presence of cDNA in case 1, case 2 and the controls used in other PCRs.

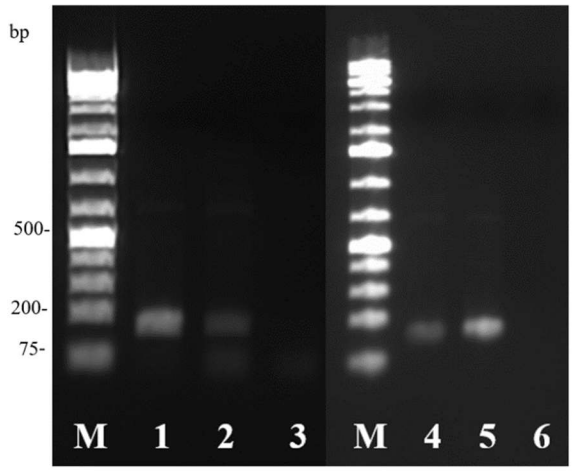

Figure 6. Agarose gel (1%) image of PCR products from case 1, 2 and controls run with the ABL1-195F and ABL1-325R primer combination. M: 1 kb DNA marker, 1: control 1, 2: control 2, 3: negative control for the PCR (nuclease-free water), 4: case 1, 5: case 2, 6: negative control for the PCR (nuclease-free water).

4.3 CHROMATOGRAMS FROM SANGER SEQUENCING

The negative controls (nuclease-free water) showed no chromatograms, indicating no contamination of the samples through the PCR and Sanger sequencing steps. Purified PCR products from Figure 5, using the HMGA2-853FW and 8q22-intron-SEQ2-94333331-Rev2 primer combination, showed the most reliable Sanger sequencing chromatograms. Case 1 had a sequencing product of 277 bp and case 2 had a product of 287 bp, when combining trimmed sequences from forward and reverse primers. The breakpoints of the fusion transcripts were precisely localized, and the genomic regions involved in the transcripts were identified. The chromatograms revealed that case 1 and 2 had the same breakpoints as seen in the putative fusion transcript that included exon 3 from *HMGA2* and intron 4 from *LINC00535.* The fusion was between nucleotide 1036 in reference sequence with accession number NM_003483.6 and nucleotide 99951 in reference sequence with accession number ENST00000520096.5. Figure 7 displays partial chromatograms depicting the breakpoint of the fusion transcript identified in case 1 and 2. The other primer combinations could either not identify the breakpoint of a fusion transcript or not identify the breakpoint with the same quality of the chromatograms as the above-mentioned primer combination. Through BLAST, it was found that the sequences aligned with chromosome 8 and 12, and that there was a match to the *HMGA2* reference transcript (NM_003483.6) with a 100% identity and an E-value of 7×10^{-62} and 3×10^{-80} for case 1 and case 2, respectively. Since the fusion transcript included an intronic region from *LINC00535,* no transcript match was found for this lncRNA using BLAST. However, the

sequence downstream of the breakpoint matched the intron 4 sequence of *LINC00535* in the putative fusion transcript with 146/146 (case 1) and 124/124 bases matching (case 2).

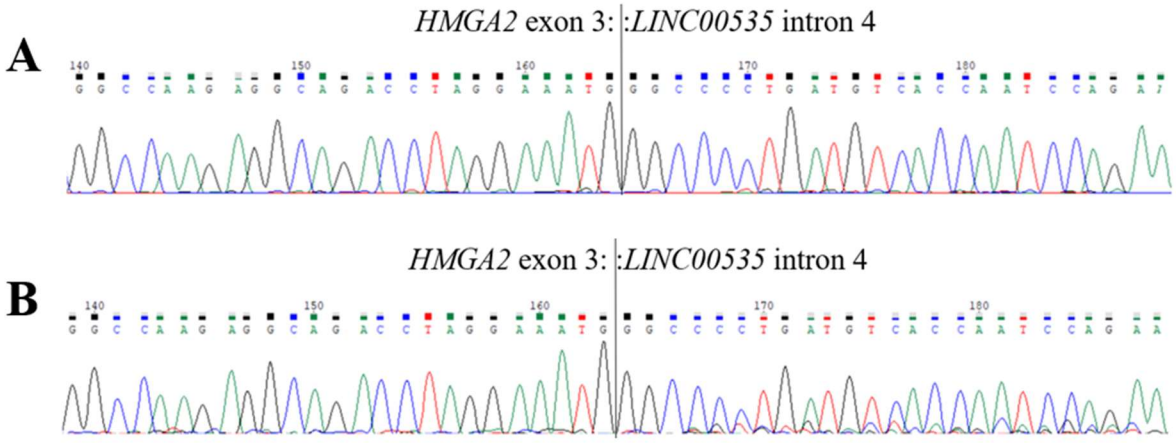

Figure 7. Partial Sanger sequencing chromatograms of case 1 (A) and 2 (B), using the HMGA2-853F primer. Breakpoint (black vertical line) of the fusion transcript is shown between exon 3 in *HMGA2* and intron 4 in *LINC00535*. Illustration made from Chromas (version 2.6.6) output.

4.4 FLUORESCENT IN SITU HYBRIDIZATION

Examination of interphase nuclei, hybridized with the *HMGA2* break-apart probe in case 1 and 2, showed fluorescence signal patterns with one red, one green and one yellow signal. Figure 8 displays an example for each case. The figure also shows an example of a metaphase spread from case 1, where the signals are seen on separate chromosomes. This pattern indicates that one homologous chromosome has the intact *HMGA2* gene (yellow signal), while it is disrupted in the other homologous chromosome (separate red and green signals). For case 1, a translocation involving chromosomes 8 and 12 is described in the karyotype. A red fluorescence signal is therefore expected to be present on the derivative chromosome 12, while the green signal has relocated to a partner chromosome, which could be a derivative chromosome 8. The metaphase spread in Figure 8 shows the translocation in case 1. In contrast, a possible insertion involving chromosomes 8 and 12 is described in the karyotype of case 2. A separate green and red fluorescence signal is therefore expected to be located on the same chromosome 12. However, no metaphase spreads were found for case 2 and it is therefore uncertain if the rearrangement seen in the interphase nuclei was due to an insertion. However, the disruption of the 12q14 chromosomal region coincide with the karyotypes in case 1 and 2.

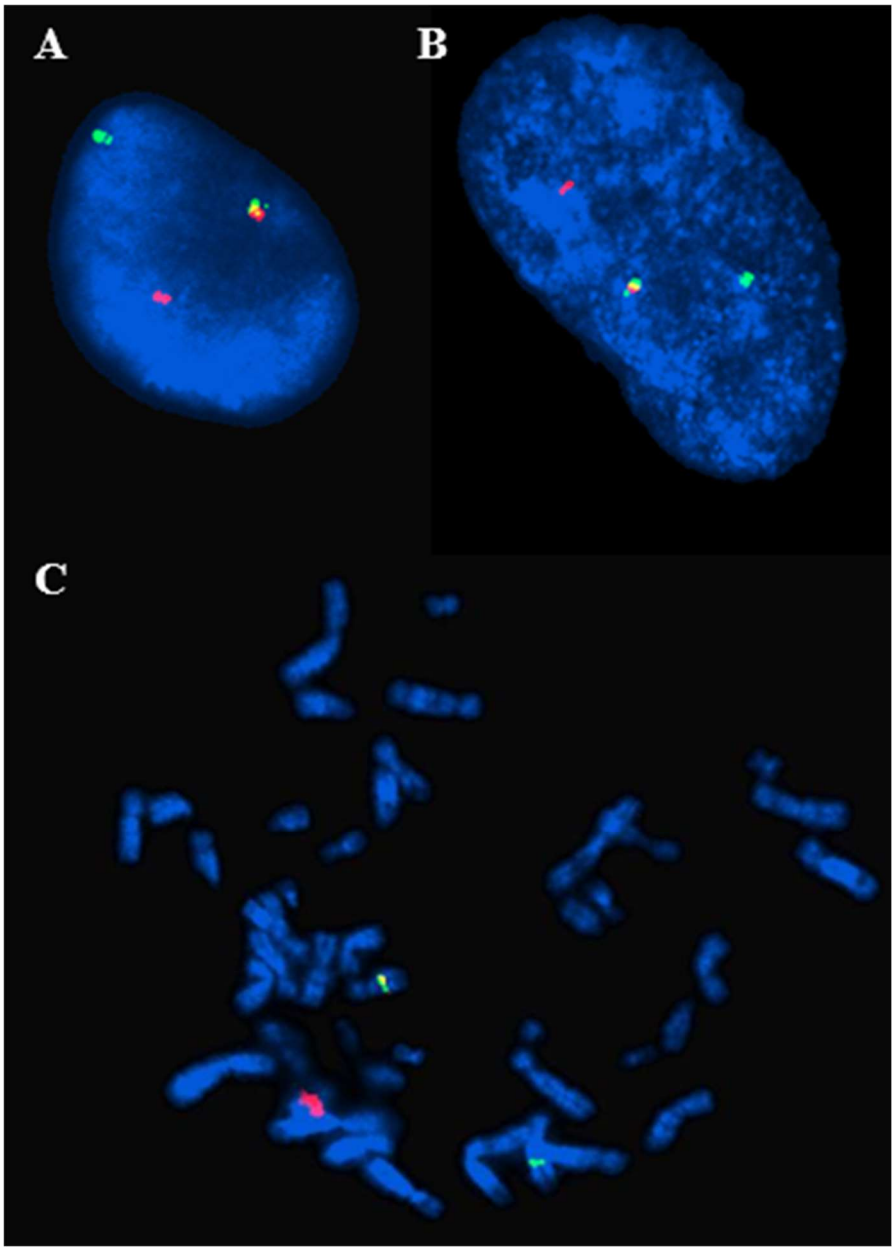

Figure 8. Images (100X magnification) from FISH analysis of interphase nuclei from case 1 (A) and 2 (B) with the *HMGA2* break-apart probe showing the intact *HMGA2* gene (yellow signal) on one of the homologous chromosomes and a disrupted gene in the other (separate red and green signals). FISH analysis on metaphase spread from case 1 (C) shows red and green fluorescence signals on derivative chromosome 12 and a possible derivative chromosome 8, respectively, and a yellow signal in the normal chromosome 12.

4.5 ARRAY COMPARATIVE GENOMIC HYBRIDIZATION

4.5.1 DNA CONCENTRATIONS AND QUALITY

DNA concentrations and quality, measured after DNA extraction of adipose tissue from case 1 and 2, are shown in Table 12. Although the extraction method was performed several times, only a total DNA concentration of 26.2 ng/µl (case 1) and 27.9 ng/µl (case 2) was obtained. Therefore, only 471.6 ng of DNA for case 1 and 502.2 ng of DNA for case 2 was used for aCGH analysis, instead of the recommended 1000 ng. The A_{260}/A_{280} ratios were close to the recommended 1.8, but the A_{260}/A_{230} ratios were lower than the recommended 1.5.

Table 12. Concentrations and quality of dsDNA measured after DNA extraction for case 1 and 2.

Lipoma sample	**dsDNA concentration (ng/µl)**	**A_{260}/A_{280}**	**A_{260}/A_{230}**
Case 1	26.2	1.72	0.83
Case 2	27.9	1.82	0.72

4.5.2 ARRAY COMPARATIVE GENOMIC HYBRIDIZATION OUTPUT

Figure J1 and J2 in Appendix J show that the quality of the aCGH output for case 1 and 2 were either excellent or satisfactory, in accordance with the CytoSure™ Interpret Software user guide (Oxford Gene Technology, 2017). Interpretation of the aCGH output for case 1 and 2 revealed no imbalances in the genome when applying the cut-off values mentioned in the methods section. Figure 9 displays an overview of the probe signals for all chromosomes in case 1 and 2. In general, it shows probe signals that are around 0, indicating no CNVs in the investigated samples, compared to the reference genome. Some vertical blue lines are seen, but these represent detected CNVs that were not within the cut-off values set in this project.

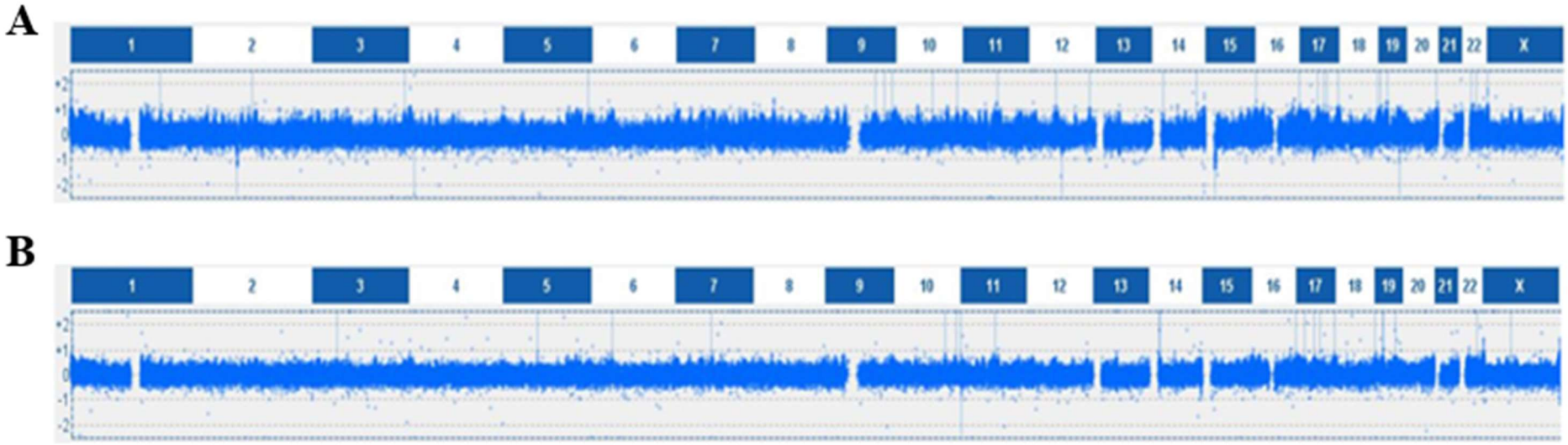

Figure 9. aCGH results showing an overview of probe signals on chromosomes 1-22 and chromosome X in case 1 (A) and 2 (B). Created from screenshots from the CytoSure™ Interpret Software (version 4.11.39).

5 DISCUSSION

The aim of this project was to increase the knowledge of chromosomal aberrations and their molecular consequences in lipomas, as this may contribute to distinguish them from malignant adipocytic tumors, and thereby improve the subclassification and diagnostic accuracy of these tumors. Two lipomas, showing rearrangements of the 8q21-22 and 12q14 chromosomal bands in their karyotypes, were investigated. The tumors were investigated on a molecular level since the gene product associated with these rearrangements had not previously been identified. They were also screened for imbalanced cryptic aberrations in the genome to gain further knowledge about chromosomal aberrations in these lipomas.

5.1 IDENTIFICATION OF THE *HMGA2::LINC00535* FUSION TRANSCRIPT

The findings show that aberrations involving chromosomes 8 and 12, with the chromosomal bands, namely 8q22 and 12q14, are recurrent in lipomas. Based on the karyotypes, a rearrangement event had occurred in each of the cases. It was found that these aberrations led to the formation of *HMGA2::LINC00535* fusion transcripts, where exon 3 from *HMGA2* was fused with intron 4 from *LINC00535*. Since fusion transcripts are investigated, introns are expected to be spliced out. However, somatic mutations may affect splicing (Calabrese et al., 2020), and fusion genes have been found to produce splice variants that include intronic sequences (McPherson et al., 2011). Combining these molecular results with the abnormal karyotype and FISH results, could indicate that the transcripts were produced from a fusion between the *HMGA2* gene and the *LINC00535* region in the genome.

According to the Mitelman database, which lists 33 800 unique fusion genes (last updated 16th of January 2024), an *HMGA2::LINC00535* fusion transcript has not been previously identified (Mitelman et al., 2024). However, *HMGA2* is known to be located in the chromosomal region most frequently involved in rearrangements in lipomas and has been identified in several fusion genes (Mandahl & Mertens, 2015; Mitelman et al., 2024). As an example, the fusion between the *HMGA2* gene and the *LIM Domain Containing Preferred Translocation Partner In Lipoma* (*LPP*) produces an HMGA2::LPP chimeric protein that is thought to have tumorigenic properties (Crombez et al., 2005; Panagopoulos et al., 2023). Conversely, *LINC00535* is a lncRNA that has not been identified in fusions in lipomas. In the past decades, researchers have focused on protein-coding genes rather than non-coding structures and their possible role in

tumorigenesis. However, lncRNAs have shown to serve a role in tumorigenesis in specific cancer types (Taniue & Akimitsu, 2021a). Tolomeo et al. (2021) explain that the lncRNA *Plasmacytoma Variant Translocation 1* (*PVT1*) is involved in 98 different fusion transcripts in solid tumors and hematological malignancies. Although many lncRNAs have been discovered through massive parallel sequencing, most of their functions remain unclear (Taniue & Akimitsu, 2021a).

Since *LINC00535* is not transcribed into a protein, a fusion protein will not be generated with HMGA2. However, the consequence of a fusion between a coding gene as the 5' partner and a lncRNA as the 3' partner is truncation of the coding gene and thereby its loss of function (Tolomeo et al., 2021). Fusion of exon 3 of the *HMGA2* gene and intron 4 of the *LINC00535* lncRNA may produce a truncated *HMGA2* transcript, and thereby a truncated HMGA2 protein. Truncated *HMGA2* transcripts have been previously identified in lipomas, where *HMGA2* is fused with a variety of partner genes, where some have shown to serve a role in tumor development (Antonescu et al., 2020). Other tumor types, such as angiomyxomas, ovarian tumors and myeloid malignancies, have also been shown to express truncated *HMGA2* transcripts (Agostini et al., 2016b; Lee et al., 2019; Odero et al., 2005). Although the partner genes may have other cellular functions than the lncRNA, many of the fusions involve only the first three exons of the *HMGA2* gene (Antonescu et al., 2020; Panagopoulos et al., 2023), as also seen in this project. Consequently, there is an absence of the C-terminal domain and the 3' untranslated region (UTR) containing regulatory parts, and the HMGA2 expression may be affected (Antonescu et al., 2020).

The relevance of a fusion transcript depends on the biological functions of the genes or RNAs involved, and its expression (Dorney et al., 2023). HMGA2 is normally not expressed in adult tissue, due to increasing let-7b miRNA expression (Mansoori et al., 2021). The *Let-7* family consists of miRNAs that directly bind to the 3'UTR part of *HMGA2* transcripts and represses its post-transcriptional expression (Zhang et al., 2019). Truncation of *HMGA2* results in a loss of *let-7* binding site and, consequently, HMGA2Tr proteins are over-expressed (Klemke et al., 2010), possibly explaining why the transcript is detected even when the RNA concentration is low. These truncated proteins, as well as full-length proteins, are sufficient to induce tumorigenesis in transgenic mice (Mansoori et al., 2021). In a study performed by Arlotta et al. (2000), it was found that expression of HMGA2Tr in transgenic mice led to obesity and the development of lipomas. Analysis could be performed to investigate if HMGA2Tr is produced in case 1 and 2, and if so, supporting the idea that the protein is involved in lipoma development.

The biological function of *LINC00535* is not clearly understood, but in a study performed by Zheng et al. (2020), it was found that this lncRNA was up-regulated and that high expressions correlated with poor survival rates in patients with Wilms' Tumor, a type of kidney cancer found in children. This could indicate that *LINC00535* plays a role in the development of pathogenic diseases. However, the biological function of full-length and shortened *LINC00535* in lipomas, the latter seen in the *HMGA2::LINC00535* fusion transcript, is not known.

Since the chromosomal aberrations involving the 8q21-22 and 12q14 chromosomal bands are found in the stem line of the clones from both case 1 and case 2, they could be important for the primary event of the lipoma development. Aberrations involving *HMGA2* has been found in the stem line of clones from other investigated lipomas (Agostini et al., 2016a; Panagopoulos et al., 2023). Hence, indicating that the gene may be relevant for the establishment of lipoma tumors.

As mentioned in Section 1.1, WHO uses molecular information when classifying specific tissue types. Identification of the *HMGA2::LINC00535* fusion transcript may therefore play a role in the subclassification of adipocytic tumors, and thereby more efficiently distinguishing lipomas from malignant adipocytic tumors. As seen here, the *HMGA2::LINC00535* fusion transcript is recurrent in lipomas. This information could be of vital importance when distinguishing benign and malignant adipocytic tumors. In the first instance, only a surgical removal of the tumor is necessary, and the patient does not need further treatment (Antonescu et al., 2020). Combined with clinical information, further subclassification of adipocytic tumors could lead to more precise diagnosis and improve therapeutic decision-making. However, a greater number of lipoma samples should be analyzed to understand the relevance of the *HMGA2::LINC00535* fusion transcript in lipomas. Additionally, other adipocytic tumors should be investigated to find out if it is specific for lipomas.

5.2 SCREENING FOR IMBALANCED CRYPTIC ABERRATIONS

In this project, no imbalanced cryptic aberrations were found in case 1 and 2, when screened with aCGH. However, the composition of the tumor is important when evaluating the results. Although aCGH revealed that all rearrangements were balanced, it could be due to few abnormal cells present in the samples. For reliable detection and accurate interpretation of aCGH results, abnormal cells should account for > 30% of the cells investigated (Lee et al., 2013). In case 1, a deletion on chromosome 10p13 was observed in ≤ 33% of the cells with

karyotyping. No deletions were found at this chromosomal region through aCGH, indicating that it is found in such small numbers of cells that it is not detectable. This could give a false negative result. For case 2, a possible insertion of (8;12)(q22;q13q14) was found in 80% of the cells through karyotyping. Since this was not detected with aCGH, the insertion could be a balanced rearrangement where the inserted chromosome region is lost from its original site. Overall, aCGH revealed no imbalanced structural aberrations. The structural aberrations found in the karyotypes are most likely balanced, which is the case for the majority of lipomas (Mandahl & Mertens, 2015).

5.3 METHODOLOGICAL CONSIDERATIONS

Chromosomal aberrations and their molecular consequences can be studied at different resolution levels and with a variety of methods. Combining cytogenetic and molecular genetic methods, with their resolution differences, is important to understand tumorigenesis and tumor progression (Heim & Mitelman, 2015).

5.3.1 METHODS TO IDENTIFY FUSION TRANSCRIPTS IN LIPOMAS

Through karyotyping, simple chromosomal rearrangements can be identified, and the breakpoints are located to specific bands. In this project, the karyotypes were used to select lipomas with simple and similar chromosomal aberrations. Samples that show a karyotype where the stem line represent a single chromosomal aberration are of interest, because these aberrations are though to represent primary events in tumor development (Heim & Mitelman, 2015). Additionally, they were good indicators of which genomic regions were involved in the chromosomal rearrangements. An advantage of using karyotyping as a screening technique is its capability to investigate rearrangements across the entire genome. This makes it a valuable tool for discriminating among all putative fusion transcripts detected later with sequencing methods, eliminating the need to further examine all transcripts (Panagopoulos, 2015). Due to the low resolution of banding technology, which is less than 10 Mb (Bass, 2018), methods with higher resolutions were used to investigate the molecular consequences of the chromosomal rearrangements.

Panagopoulos (2015) describes that using FISH to identify genes involved in a fusion event is laborious, as different probes need to be tested to find the one that is overlapping the breakpoint

of a fusion. He further explains that NGS-related methods are therefore frequently used to identify fusion genes or fusion transcripts, and that FISH is used later for the validation of the sequencing results. Most fusions are found through high throughput sequencing technologies, such as WGS or RNA-Seq (Panagopoulos, 2015). As explained in Section 1.5, WGS can be used to identify both transcribed and non-transcribed fusion genes, but that the relevance of a fusion gene relies on whether it produces a fusion transcript or not. Therefore, methods such as RT-PCR and Sanger sequencing can be used to confirm that an identified rearrangement generates a fusion transcript (Panagopoulos, 2015). RNA-Seq, on the other hand, is a common method for fusion transcript identification, and has been used in several fusion transcript studies (Panagopoulos, 2015; Panagopoulos et al., 2023). However, since it generates a lot of data, including false positives, it also needs confirmation methods (Carrara et al., 2013; Taniue & Akimitsu, 2021b). RNA-Seq requires less computational analysis and is less time consuming to perform compared to WGS, but cannot detect lowly expressed or non-transcribed genes (Panagopoulos, 2015). Nevertheless, RNA-Seq was proven by Maher et al. (2009) to be an efficient method to detect fusion transcripts. These are presumably some of the reasons why RNA-Seq was used prior to this project.

RT-PCR, followed by Sanger sequencing, was used to validate the putative fusion transcripts identified with RNA-Seq. Sanger sequencing has a higher resolution than banding technologies and can detect sequences between 100-1000 bp in length (Crossley et al., 2020). In this way, we could localize the breakpoint of the fusion more precisely and identify the genomic regions involved in the fusion transcript. When validating fusion transcripts from a transcriptomic level rather than a genomic level, problems related to variable genomic breakpoints are avoided (Panagopoulos, 2015). The identification of a fusion transcript depends on the designed PCR primers. The breakpoint of the *HMGA2::LINC00535* fusion transcript was successfully identified, because the HMGA2-853FW and 8q22-intron-SEQ2-94333331-Rev2 primers were designed such that the product included the fusion breakpoint. Additionally, they were designed to hybridize not too close to the breakpoint, to avoid a potential problem with Sanger sequencing, as the quality tends to be low within the first 15-40 bp due to primer binding (Crossley et al., 2020).

In this project, FISH was used to validate the presence of fusion events at the genomic level. Today, many reported fusion genes are not confirmed at the genome-level and only identified at a transcriptomic level (Panagopoulos & Heim, 2022). By using FISH probes that hybridizes to specific genomic regions, we confirmed that there was a correlation between the

HMGA2::LINC00535 fusion transcript and a structural rearrangement at the DNA level. Although the rearrangements were already observed in the karyotypes, the resolution of FISH, which was approximately 200 kb, made it possible to investigate the gene level consequences of the aberrations observed. However, when using break-apart probes, only one of the genes in a fusion is identified, in this case the *HMGA2* gene. Identifying the partner gene has shown to be important for diagnosis and therapeutic approaches of different diseases (Krystel-Whittemore et al., 2019). With the *HMGA2* break-apart probe, it could not precisely be determined if the partner region was *LINC00535* on chromosome 8 in case 1 and 2. To better determine this, a dual-fusion probe could be used instead of a break-apart probe. Following the principle described in Ventura et al. (2006), one fluorophore would hybridize to the *HMGA2* gene and another one to *LINC00535*, giving a mixed fluorescence signal when fused. Since a positive result requires two fusion signals, it is thought that the chance for false-positives is lower than for break-apart probes and that the sensitivity is therefore higher (Ventura et al., 2006). Although more probes have become available in the later years as a consequence of the Human Genome Project, which was conducted between 1990-2013, they often only represent genes that are commonly involved in rearrangements (Green et al., 2015; Heyer & Blackburn, 2020; Panagopoulos, 2015). A probe complementary to the *LINC00535* sequence therefore needs to be designed. Overall, the FISH method has high specificity and sensitivity, and has therefore been commonly used in fusion gene detection (Heyer & Blackburn, 2020).

5.3.2 Methods to Investigate Imbalanced Cryptic Aberrations in Lipomas

With aCGH, the genome of the two lipoma cases could be screened for imbalanced cryptic rearrangements. This is not possible with karyotyping, because cryptic rearrangements are below the chromosomal resolution level at approximately 10 Mb (Bass, 2018). FISH could be used to identify cryptic rearrangements, however, its multiplexing capacity is low, which makes it less useful for screening (Heyer & Blackburn, 2020). Screening for cryptic rearrangements in lipomas may be important for the subclassification of adipocytic tumors. A study performed by Svobodova and her colleagues showed that cryptic aberrations identified in bone-marrow from patients with myelodysplastic syndromes may be of prognostic significance (Svobodova et al., 2020). This also indicates that cryptic aberrations could be relevant for the classification of diseases. Although the majority of structural aberrations in lipomas are balanced and cannot be identified with aCGH, identification of imbalanced cryptic rearrangements could potentially improve the accuracy of the subclassification of adipocytic tumors (Mandahl & Mertens, 2015).

5.3.3 RNA Quality and Its Influence on the Detection of Fusion Transcripts

Reliable PCR and RNA-Seq results relies on total RNA quality. RNA Integrity Number (RIN) values range from 1 to 10, where 1 represents totally degraded RNA and 10 represents intact RNA (Schroeder et al., 2006). According to Puchta et al. (2020), high quality RNA should at least have values around 7-8. However, obtaining high quality RNA from some tissue types is not possible and using low quality RNA could be the only option when investigating such tissues (Sigurgeirsson et al., 2014). According to Stroh et al. (2021), RIN values of lipomas could reach up to 8.8. However, RNA integrity depends on the sample storage before the tissue enters the lab, which in some cases are taken care of by other personnel outside the lab. The quality of the RNA in lipomas could therefore vary and are not always under the lab-personnels control.

As seen in Table B1 in Appendix B, the RIN for case 1 and 2 measured in advance of this project, is low. However, there seems to be no correlation between the concentration of PCR product, shown in Figure 5, and the RIN values. Case 1 had a higher intensity band on the gel compared to case 2, but poorer RIN value. An explanation of the non-correlation could be that RIN measurements depends on 18S and 28S ribosomal RNA, which have shown to have low correlation with RNA integrity of other RNA molecules (Puchta et al., 2020). It has been reported that degraded RNA may affect quantitative RT-PCR analysis, giving inaccurate expression results (Fleige & Pfaffl, 2006). If RNA is highly degraded, both forward and reverse PCR primers, which are needed for amplification of a template, cannot bind, leading to no amplification and a false negative result.

False positive fusion transcripts are not expected as a consequence of sequencing samples with low RIN values, but there could be false negative results. Although Illumina states that low-quality samples may be used for their total RNA-Seq analysis (Illumina, n.d.), it is expected that degraded RNA will reduce the read coverage. Davila et al. (2016) showed that degraded RNA can result in the loss of reads covering the breakpoint of a fusion transcript, consequently affecting the fusion detection sensitivity. Additionally, Sigurgeirsson et al. (2014) have not only shown that sequencing degraded RNA can lead to the underrepresentation of full-length transcripts, but also that it can lead to the overrepresentation of short transcripts, which may have an impact on differential expression analysis. All of this suggests that false negative fusion transcripts may occur and that the quantitative amount of an identified fusion transcript may be false when sequencing low integrity RNA.

However, a more reliable determinant of RNA quality for RNA-Seq is the percentage of RNA fragments larger than 200 nt (Illumina, 2016). As seen in Table B1 in Appendix B, both cases had values above 50%, which is recommended as the lower limit for RNA-Seq by The Genomics Core Facility at the Norwegian Radium Hospital, Oslo University Hospital.

Although putative fusions are selected through filtering steps, both negative and false positive results are common for algorithms detecting fusion transcripts, including deFuse (Carrara et al., 2013; Taniue & Akimitsu, 2021b). The *HMGA2::LINC00535* fusion transcript was verified through PCR analysis, which strengthens our impression that the RIN value for case 1 and 2 did not affect the identification of this fusion. If an expression analysis was to be performed, higher RIN values would be of importance.

5.3.4 Extraction of DNA From Lipomas

The DNA content in adipose tissue in humans are relatively understudied and different methods have been used to extract DNA from lipomas (Stroh et al., 2021; Tap et al., 2011). Obtaining high DNA yield is method dependent. Although methods, such as the DNeasy Blood & Tissue kit (QIAGEN), have shown to yield high quantities of DNA (Stroh et al., 2021), no column-based method was used in this project, as this was not available. By using the DNeasy Blood & Tissue kit (QIAGEN), Stroh et al. (2021) found that the concentration of DNA in adipose tissue was 52 ± 14 ng/mg tissue. As a reference, this only corresponds to 28% of the DNA found in muscle tissue (Stroh et al., 2021), indicating that greater amount of tissue is necessary to obtain sufficient DNA concentrations from lipomas.

In this project, the first attempt of DNA extraction was performed according to the protocol for the Maxwell RSC Tissue DNA Kit (Promega), with low DNA yield. To obtain sufficient DNA for aCGH, the extraction method was improved by incubating the tissue on a heating block to melt the adipose tissue, as described by Tap et al. (2011) and Stroh et al. (2021). This made it easier to extract more tissue compared to tissue that had only been homogenized with a pestle, consequently leading to a higher DNA concentration of the eluate. In addition, lipids are known to cause some challenges when extracting DNA (Stroh et al., 2021), and by heating the tissue, a lipid layer was formed, making it easier to prevent it from contaminating the sample. According to Macherey-Nagel (2016), lipids may affect tissue disruption or influence the chemistry of the extraction buffers. This could potentially lead to lower DNA yield.

Using less eluate volume also increased the concentration, but less than 50 µl of eluate was obtained after extraction. Evaporation was observed, and a higher eluate volume should therefore be used to ensure enough material for sequential analysis, such as aCGH. However, using Maxwell RSC Tissue DNA Kit (Promega) instead of a column-based method was more labor efficient since most of the process was automated.

Adding proteinase K to the tissue was done to improve the quality of the DNA by degrading nucleases (Sweeney & Walker, 1993). It is unsure whether proteinase K improved the DNA quality, as we did not perform aCGH on DNA incubated without proteinase K. As shown in Table 12, the A_{260}/A_{230} absorbance ratios measured in the eluates were lower than the recommended 1.5, which may indicate a carryover of reagents with absorbance around 230 nm from the extraction procedure. An explanation could be that the Maxwell RSC Tissue DNA Kit (Promega) contains reagents with guanidine, which has an absorbance around 230 nm (Thermo Scientific, n.d.). However, it did not seem to influence the quality of aCGH output.

Overall, although we did not obtain the recommended DNA yield for aCGH with the extraction method used in this project, we showed that aCGH can still be performed on ~500 ng of DNA from lipomas. To summarize, improvements to extraction methods could be done to increase the DNA yield from lipomas. As a suggestion, the DNeasy Blood & Tissue kit (QIAGEN) could be used to extract DNA, with modifications mentioned in this section.

6 CONCLUSION

In conclusion, aberrations involving the 8q22 and 12q14 chromosomal bands were found to be recurrent in lipomas. The aberrations led to the formation of an *HMGA2::LINC00535* fusion transcript, which has not been previously identified. If translated into a protein, a truncated HMGA2 protein may be formed, which is already thought to play a role in the development of lipomas and other neoplasms. Furthermore, no imbalanced cryptic aberrations were identified in the two lipomas. Although lipomas are benign, the identification of the fusion transcript could be valuable to distinguish them from other, including malignant, adipocytic tumors. Combining these results with clinical information may improve the subclassification of adipocytic tumors and thereby enhancing the decision-making regarding diagnosis, prognosis and treatment of these tumors. However, further studies, involving other lipomas and adipocytic tumors, are needed to determine if the *HMGA2::LINC00535* fusion transcript is specific for lipomas.

7 FUTURE PERSPECTIVES

Ideally, a higher number of lipomas should be investigated, and it would be preferable to also investigate matched normal adipose tissue. However, obtaining normal adipose tissue for research purposes introduces ethical issues, which makes it less feasible, especially considering the amount needed to obtain a sufficient amount of DNA or RNA.

However, it would be interesting to investigate the consequences of the *HMGA2::LINC00535* fusion transcript. With gene expression analysis, it may be determined if the fusion is differentially expressed in lipomas compared to normal adipose tissue. A differential expressed fusion gene could indicate its role in lipoma development. If the *HMGA2* is missing the *let-7* binding site, it is expected to be overexpressed (Klemke et al., 2010). Quantification of a truncated *HMGA2* transcript can be performed through quantitative Reverse Transcription PCR (RT-qPCR) by using two primer targets, one for exon 1-2 and one for exon 4-5, excluding exon 3 where the rearrangements were identified (Bartuma et al., 2009). A higher expression of either of them compared to a normal control would indicate differential expressed fusion genes (Bartuma et al., 2009; Panagopoulos et al., 2015).

Identification of the truncated HMGA2 protein could give an indication of the biological consequence of the *HMGA2::LINC00535* fusion transcript. As previously mentioned, the expression of HMGA2Tr is sufficient to induce lipoma development in transgenic mice (Arlotta et al., 2000). If the *HMGA2::LINC00535* fusion transcript produces an HMGA2Tr protein, it could be related to lipoma development. Its presence in benign lipomas could also indicate that the protein alone is not sufficient for malignant transformation. Truncated HMGA2 proteins can suggestively be detected through Western blot analysis, using a full-length HMGA2 protein as a control (Arlotta et al., 2000).

REFERENCES

Agostini, A., Gorunova, L., Bjerkehagen, B., Lobmaier, I., Heim, S. & Panagopoulos, I. (2016a). Molecular characterization of the t(4;12)(q27~28;q14~15) chromosomal rearrangement in lipoma. *Oncology Letters*, 12 (3): 1701-1704. doi: 10.3892/ol.2016.4834.

Agostini, A., Panagopoulos, I., Davidson, B., Trope, C. G., Heim, S. & Micci, F. (2016b). A novel truncated form of HMGA2 in tumors of the ovaries. *Oncology Letters*, 12 (2): 1559-1563. doi: 10.3892/ol.2016.4805.

Annala, M. J., Parker, B. C., Zhang, W. & Nykter, M. (2013). Fusion genes and their discovery using high throughput sequencing. *Cancer Letters*, 340 (2): 192-200. doi: 10.1016/j.canlet.2013.01.011.

Antonescu, C. R., Bridge, J. A., Cunha, I. W., Dei Tos, A. P., Fletcher, C. D. M., Folpe, A. L., Goldblum, J. R., Hornick, J. L., Miettinen, M. & Oda, Y. (2020). Soft tissue tumours. In WHO Classification of Tumours Editorial Board (ed.) *Soft Tissue and Bone Tumours*, pp. 6-318. Lyon, France: International Agency for Research on Cancer.

Arlotta, P., Tai, A. K., Manfioletti, G., Clifford, C., Jay, G. & Ono, S. J. (2000). Transgenic mice expressing a truncated form of the high mobility group I-C protein develop adiposity and an abnormally high prevalence of lipomas. *The Journal of Biological Chemistry*, 275 (19): 14394-400. doi: 10.1074/jbc.m000564200.

Bartuma, H., Panagopoulos, I., Collin, A., Trombetta, D., Domanski, H. A., Mandahl, N. & Mertens, F. (2009). Expression levels of HMGA2 in adipocytic tumors correlate with morphologic and cytogenetic subgroups. *Molecular Cancer*, 8 (1): 36. doi: 10.1186/1476-4598-8-36.

Bass, H. W. (2018). Chromosomes. In Krebs, J. E., Goldstein, E. S. & Kilpatrick, S. T. (eds) *Lewin's genes XII*, pp. 161-185. Burlington, MA, USA: Jones & Barlett Learning.

Brien, G. L., Stegmaier, K. & Armstrong, S. A. (2019). Targeting chromatin complexes in fusion protein-driven malignancies. *Nature Reviews Cancer*, 19 (5): 255-269. doi: 10.1038/s41568-019-0132-x.

Broberg, K., Toksvig-Larsen, S., Lindstrand, A. & Mertens, F. (2001). Trisomy 7 accumulates with age in solid tumors and non-neoplastic synovia. *Genes, Chromosomes & Cancer*, 30 (3): 310-5 https://pubmed.ncbi.nlm.nih.gov/11170291/#:~:text=Trisomy%207%20is%20a%20common,than%20the%20disease%20process%20itself. (accessed: 13 May 2024).

Brunetti, M., Panagopoulos, I., Vitelli, V., Andersen, K., Hveem, T. S., Davidson, B., Eriksson, A. G. Z., Trent, P. K. B., Heim, S. & Micci, F. (2022). Endometrial Carcinoma: Molecular Cytogenetics and Transcriptomic Profile. *Cancers (Basel)*, 14 (14). doi: 10.3390/cancers14143536.

Bäsecke, J., Griesinger, F., Trümper, L. & Brittinger, G. (2002). Leukemia- and lymphoma-associated genetic aberrations in healthy individuals. *Annals of Hematology*, 81 (2): 64-75. doi: 10.1007/s00277-002-0427-x.

Calabrese, C., Davidson, N. R., Demircioğlu, D., Fonseca, N. A., He, Y., Kahles, A., Lehmann, K. V., Liu, F., Shiraishi, Y., Soulette, C. M., et al. (2020). Genomic basis for RNA alterations in cancer. *Nature*, 578 (7793): 129-136. doi: 10.1038/s41586-020-1970-0.

Carrara, M., Beccuti, M., Lazzarato, F., Cavallo, F., Cordero, F., Donatelli, S. & Calogero, R. A. (2013). State-of-the-Art Fusion-Finder Algorithms Sensitivity and Specificity. *BioMed Research International*, 2013: 340620. doi: 10.1155/2013/340620.

Cohen, P., Cross, D. & Jänne, P. A. (2021). Kinase drug discovery 20 years after imatinib: progress and future directions. *Nature Reviews Drug Discovery*, 20 (7): 551-569. doi: 10.1038/s41573-021-00195-4.

Crombez, K. R., Vanoirbeek, E. M., Van de Ven, W. J. & Petit, M. M. (2005). Transactivation functions of the tumor-specific HMGA2/LPP fusion protein are augmented by wild-type HMGA2. *Molecular Cancer Research*, 3 (2): 63-70. doi: 10.1158/1541-7786.Mcr-04-0181.

Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., Clement, T. & Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic Investigation*, 32 (6): 767-775. doi: 10.1177/1040638720905833.

Davila, J. I., Fadra, N. M., Wang, X., McDonald, A. M., Nair, A. A., Crusan, B. R., Wu, X., Blommel, J. H., Jen, J., Rumilla, K. M., et al. (2016). Impact of RNA degradation on fusion detection by RNA-seq. *BMC Genomics*, 17 (1): 814. doi: 10.1186/s12864-016-3161-9.

Dorney, R., Dhungel, B. P., Rasko, J. E. J., Hebbard, L. & Schmitz, U. (2023). Recent advances in cancer fusion transcript detection. *Briefings in Bioinformatics*, 24 (1). doi: 10.1093/bib/bbac519.

Fleige, S. & Pfaffl, M. W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular Aspects of Medicine*, 27 (2-3): 126-39. doi: 10.1016/j.mam.2005.12.003.

Fletcher, H. & Hickey, I. (2013). PCR and related technology. In *Genetics*, pp. 266-271. London: Garland Science.

Gomes, C. C. (2022). Recurrent driver mutations in benign tumors. *Mutation Research - Reviews in Mutation Research*, 789: 108412. doi: 10.1016/j.mrrev.2022.108412.

Goodwin, G. H., Sanders, C. & Johns, E. W. (1973). A new group of chromatin-associated proteins with a high content of acidic and basic amino acids. *European Journal of Biochemistry*, 38 (1): 14-9. doi: 10.1111/j.1432-1033.1973.tb03026.x.

Green, E. D., Watson, J. D. & Collins, F. S. (2015). Human Genome Project: Twenty-five years of big biology. *Nature*, 526 (7571): 29-31. doi: 10.1038/526029a.

Heim, S. & Mitelman, F. (2015). Nonrandom chromosome abnormalities in cancer. In Heim, S. & Mitelman, F. (eds) *Cancer Cytogenetics*, pp. 26-41. Chichester, UK: Wiley-Blackwell. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118795569.ch4 (accessed: 20 February 2024).

Heyer, E. E. & Blackburn, J. (2020). Sequencing Strategies for Fusion Gene Detection. *Bioessays*, 42 (7): e2000016. doi: 10.1002/bies.202000016.

Hochhaus, A., Larson, R. A., Guilhot, F., Radich, J. P., Branford, S., Hughes, T. P., Baccarani, M., Deininger, M. W., Cervantes, F., Fujihara, S., et al. (2017). Long-Term Outcomes of Imatinib Treatment for Chronic Myeloid Leukemia. *The New England Journal of Medicine*, 376 (10): 917-927. doi: 10.1056/NEJMoa1609324.

Illumina. (2016). *Evaluating RNA Quality from FFPE Samples*. Available at: https://www.illumina.com/content/dam/illumina-marketing/documents/products/technotes/evaluating-rna-quality-from-ffpe-samples-technical-note-470-2014-001.pdf (accessed: 3 May 2024).

Illumina. (n.d.). *A comprehensive picture of the transcriptome*. Available at: https://emea.illumina.com/techniques/sequencing/rna-sequencing/total-rna-seq.html (accessed: 3 April 2024).

Jackson, M., Marks, L., May, G. H. W. & Wilson, J. B. (2018). The genetic basis of disease. *Essays in Biochemistry*, 62 (5): 643-723. doi: 10.1042/ebc20170053.

Kang, Z. J., Liu, Y. F., Xu, L. Z., Long, Z. J., Huang, D., Yang, Y., Liu, B., Feng, J. X., Pan, Y. J., Yan, J. S. & Liu, Q. (2016). The Philadelphia chromosome in leukemogenesis. *Chinese Journal of Cancer*, 35: 48. doi: 10.1186/s40880-016-0108-0.

Kannan, T. P. & Zilfalil, B. A. (2009). Cytogenetics: past, present and future. *Malaysian Journal of Medical Science*, 16 (2): 4-9. Available at: https://pubmed.ncbi.nlm.nih.gov/22589651/ (accessed: 1 March 2024).

Klemke, M., Meyer, A., Hashemi Nezhad, M., Belge, G., Bartnitzke, S. & Bullerdiek, J. (2010). Loss of let-7 binding sites resulting from truncations of the 3' untranslated region of HMGA2 mRNA in uterine leiomyomas. *Cancer Genetics and Cytogenetics*, 196 (2): 119-23. doi: 10.1016/j.cancergencyto.2009.09.021.

Kosztyuova, T. & Shim, T. N. (2017). Rapidly enlarging lipoma. *BMJ Case Reports*, 2017. doi: 10.1136/bcr-2017-221272.

Krebs, J. E., Goldstein, E. S. & Kilpatrick, S. T. (2018). Methods in Molecular Biology and Genetic Engineering. In *Lewin's genes XII*, pp. 35-69. Burlington, MA, USA: Jones & Barlett Learning.

Krystel-Whittemore, M., Taylor, M. S., Rivera, M., Lennerz, J. K., Le, L. P., Dias-Santagata, D., Iafrate, A. J., Deshpande, V., Chebib, I., Nielsen, G. P., et al. (2019). Novel and established EWSR1 gene fusions and associations identified by next-generation sequencing and fluorescence in-situ hybridization. *Human Pathology*, 93: 65-73. doi: 10.1016/j.humpath.2019.08.006.

Lee, C. G., Yun, J.-N., Park, S.-J. & Sohn, Y. B. (2013). Low-frequency Mosaicism of Trisomy 14, Missed by Array CGH. *Journal of Genetic Medicine*, 10 (1): 52-56. doi: 10.5734/JGM.2013.10.1.52.

Lee, M. Y., da Silva, B., Ramirez, D. C. & Maki, R. G. (2019). Novel HMGA2-YAP1 fusion gene in aggressive angiomyxoma. *BMJ Case Reports*, 12 (5). doi: 10.1136/bcr-2018-227475.

Macherey-Nagel. (2016). *Genomic DNA from lipid-rich tissue*. Available at: http://bioke.be/blobs/Plaatjes/MN/UM_gDNALipidTissue.pdf (accessed: 28 April 2024).

Maher, C. A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N. & Chinnaiyan, A. M. (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature*, 458 (7234): 97-101. doi: 10.1038/nature07638.

Mandahl, N. & Mertens, F. (2015). Soft tissue tumors. In Heim, S. & Mitelman, F. (eds) *Cancer Cytogenetics*, pp. 583-614. Chichester, UK: Wiley-Blackwell. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118795569.ch24 (accessed: 16 February 2024).

Mansoori, B., Mohammadi, A., Ditzel, H. J., Duijf, P. H. G., Khaze, V., Gjerstorff, M. F. & Baradaran, B. (2021). HMGA2 as a Critical Regulator in Cancer Development. *Genes (Basel)*, 12 (2). doi: 10.3390/genes12020269.

McPherson, A., Hormozdiari, F., Zayed, A., Giuliany, R., Ha, G., Sun, M. G., Griffith, M., Heravi Moussavi, A., Senz, J., Melnyk, N., et al. (2011). deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Computational Biology*, 7 (5): e1001138. doi: 10.1371/journal.pcbi.1001138.

Mitelman, F., Johansson, B. & Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nature Reviews Cancer*, 7 (4): 233-245. doi: 10.1038/nrc2091.

Mitelman, F. & Heim, S. (2015). How it all began. In Heim, S. & Mitelman, F. (eds) *Cancer Cytogenetics*, pp. 1-10. Chichester, UK: Wiley-Blackwell. Available at:

https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118795569.ch1 (accessed: 9 February 2024).
Mitelman, F., Johansson, B. & Mertens, F. (2024). *Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer*. Available at: https://mitelmandatabase.isb-cgc.org/ (accessed: 27 February 2024).
Odero, M. D., Grand, F. H., Iqbal, S., Ross, F., Roman, J. P., Vizmanos, J. L., Andrieux, J., Laï, J. L., Calasanz, M. J. & Cross, N. C. (2005). Disruption and aberrant expression of HMGA2 as a consequence of diverse chromosomal translocations in myeloid malignancies. *Leukemia*, 19 (2): 245-52. doi: 10.1038/sj.leu.2403605.
Oxford Gene Technology. (2015). *CytoSure Array Handbook (4x44k and 4x180k formats)*. Available at: https://www.ogt.com/media/tqxfzfay/cytosure__array_handbook__4x44k_and_4x180k_formats_.pdf (accessed: 4 March 2024).
Oxford Gene Technology. (2017). *CytoSure Interpret Software User Guide*. Available at: https://www.ogt.com/media/bpalughv/cytosure-interpret-software-user-guide.pdf (accessed: 13 May 2024).
Panagopoulos, I. (2015). From chromosomes to genes. In Heim, S. & Mitelman, F. (eds) *Cancer Cytogenetics*, pp. 42-61. Chichester, UK: Wiley-Blackwell. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118795569.ch5 (accessed: 9 February 2024).
Panagopoulos, I., Gorunova, L., Bjerkehagen, B., Lobmaier, I. & Heim, S. (2015). The recurrent chromosomal translocation t(12;18)(q14~15;q12~21) causes the fusion gene HMGA2-SETBP1 and HMGA2 expression in lipoma and osteochondrolipoma. *International Journal of Oncology*, 47 (3): 884-90. doi: 10.3892/ijo.2015.3099.
Panagopoulos, I. & Heim, S. (2021). Interstitial Deletions Generating Fusion Genes. *Cancer Genomics & Proteomics*, 18 (3): 167-196. doi: 10.21873/cgp.20251.
Panagopoulos, I. & Heim, S. (2022). Neoplasia-associated Chromosome Translocations Resulting in Gene Truncation. *Cancer Genomics & Proteomics*, 19 (6): 647-672. doi: 10.21873/cgp.20349.
Panagopoulos, I., Andersen, K., Brunetti, M., Gorunova, L., Lund-Iversen, M., Micci, F. & Heim, S. (2023). Fusion of the High-mobility Group AT-Hook 2 (HMGA2) and the Gelsolin (GSN) Genes in Lipomas With t(9;12)(q33;q14) Chromosomal Translocation. *In Vivo*, 37 (2): 524-530. doi: 10.21873/invivo.13110.
Puchta, M., Boczkowska, M. & Groszyk, J. (2020). Low RIN Value for RNA-Seq Library Construction from Long-Term Stored Seeds: A Case Study of Barley Seeds. *Genes (Basel)*, 11 (10). doi: 10.3390/genes11101190.
Pös, O., Radvanszky, J., Buglyó, G., Pös, Z., Rusnakova, D., Nagy, B. & Szemes, T. (2021). DNA copy number variation: Main characteristics, evolutionary significance, and pathological aspects. *Biomedical Journal*, 44 (5): 548-559. doi: 10.1016/j.bj.2021.02.003.
Rana, B. & Joshi, G. K. (2023). Chapter 11 - Electrophoresis: Basic principle, types, and applications. In Bhatt, A. K., Bhatia, R. K. & Bhalla, T. C. (eds) *Basic Biotechniques for Bioprocess and Bioentrepreneurship*, pp. 183-193: Academic Press. Available at: https://www.sciencedirect.com/science/article/pii/B9780128161098000118 (accessed: 9 March 2024).
Repana, D., Nulsen, J., Dressler, L., Bortolomeazzi, M., Venkata, S. K., Tourna, A., Yakovleva, A., Palmieri, T. & Ciccarelli, F. D. (2019). The Network of Cancer Genes (NCG): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biology*, 20 (1): 1. doi: 10.1186/s13059-018-1612-0.

Sampaio, M. M., Santos, M. L. C., Marques, H. S., Gonçalves, V. L. S., Araújo, G. R. L., Lopes, L. W., Apolonio, J. S., Silva, C. S., Santos, L. K. S., Cuzzuol, B. R., et al. (2021). Chronic myeloid leukemia-from the Philadelphia chromosome to specific target drugs: A literature review. *World Journal of Clinical Oncology*, 12 (2): 69-94. doi: 10.5306/wjco.v12.i2.69.

Schoumans, J., Suela, J., Hastings, R., Muehlematter, D., Rack, K., van den Berg, E., Berna Beverloo, H. & Stevens-Kroef, M. (2016). Guidelines for genomic array analysis in acquired haematological neoplastic disorders. *Genes, Chromosomes & Cancer*, 55 (5): 480-91. doi: 10.1002/gcc.22350.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. & Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, 7: 3. doi: 10.1186/1471-2199-7-3.

Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Månér, S., Massa, H., Walker, M., Chi, M., et al. (2004). Large-scale copy number polymorphism in the human genome. *Science*, 305 (5683): 525-8. doi: 10.1126/science.1098918.

Sigurgeirsson, B., Emanuelsson, O. & Lundeberg, J. (2014). Sequencing degraded RNA addressed by 3' tag counting. *PLoS One*, 9 (3): e91851. doi: 10.1371/journal.pone.0091851.

Stroh, A. M., Lynch, C. E., Lester, B. E., Minchev, K., Chambers, T. L., Montenegro, C. F., Chavez Martinez, C., Fountain, W. A., Trappe, T. A. & Trappe, S. W. (2021). Human adipose and skeletal muscle tissue DNA, RNA, and protein content. *Journal of Applied Physiology*, 131 (4): 1370-1379. doi: 10.1152/japplphysiol.00343.2021.

Svobodova, K., Lhotska, H., Hodanova, L., Pavlistova, L., Vesela, D., Belickova, M., Vesela, J., Brezinova, J., Sarova, I., Izakova, S., et al. (2020). Cryptic aberrations may allow more accurate prognostic classification of patients with myelodysplastic syndromes and clonal evolution. *Genes, Chromosomes & Cancer*, 59 (7): 396-405. doi: 10.1002/gcc.22841.

Sweeney, P. J. & Walker, J. M. (1993). Proteinase K (EC 3.4.21.14). In Burrell, M. M. (ed.) *Enzymes of Molecular Biology*, pp. 305-311. Totowa, NJ, USA: Humana Press. Available at: https://doi.org/10.1385/0-89603-234-5:305 (accessed: 28 April 2024).

Taniue, K. & Akimitsu, N. (2021a). The Functions and Unique Features of LncRNAs in Cancer Development and Tumorigenesis. *International Journal of Molecular Sciences*, 22 (2). doi: 10.3390/ijms22020632.

Taniue, K. & Akimitsu, N. (2021b). Fusion Genes and RNAs in Cancer Development. *Noncoding RNA*, 7 (1). doi: 10.3390/ncrna7010010.

Tap, W. D., Eilber, F. C., Ginther, C., Dry, S. M., Reese, N., Barzan-Smith, K., Chen, H. W., Wu, H., Eilber, F. R., Slamon, D. J. & Anderson, L. (2011). Evaluation of well-differentiated/de-differentiated liposarcomas by high-resolution oligonucleotide array-based comparative genomic hybridization. *Genes, Chromosomes & Cancer*, 50 (2): 95-112. doi: 10.1002/gcc.20835.

Thermo Fisher Scientific. (2016). *BigDye Terminator v1.1 Cycle Sequencing Kit User Guide*. Available at: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_041330.pdf (accessed: 8 February 2024).

Thermo Fisher Scientific. (n.d.-a). *PCR Cycling Parameters - Six Key Considerations for Success*. Available at: https://www.thermofisher.com/no/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/pcr-education/pcr-reagents-enzymes/pcr-cycling-considerations.html#:~:text=The%20annealing%20temperature%20is%20determined,Tm%20of%20the%20primers. (accessed: 29 February 2024).

Thermo Fisher Scientific. (n.d.-b). *Reverse Transcription Reaction Setup - Seven Important Considerations*. Available at: https://www.thermofisher.com/no/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/rt-education/reverse-transcription-setup.html#rt_a6 (accessed: 12 March 2024).

Thermo Scientific. (n.d.). *260/280 and 260/230 Ratios*. Available at: https://dna.uga.edu/wp-content/uploads/sites/51/2019/02/Note-on-the-260_280-and-260_230-Ratios.pdf (accessed: 28 May 2024).

Tolomeo, D., Agostini, A., Visci, G., Traversa, D. & Storlazzi, C. T. (2021). PVT1: A long non-coding RNA recurrently involved in neoplasia-associated fusion transcripts. *Gene*, 779: 145497. doi: 10.1016/j.gene.2021.145497.

Ventura, R. A., Martin-Subero, J. I., Jones, M., McParland, J., Gesk, S., Mason, D. Y. & Siebert, R. (2006). FISH analysis for the detection of lymphoma-associated chromosomal abnormalities in routine paraffin-embedded tissue. *The Journal of Molecular Diagnostics*, 8 (2): 141-51. doi: 10.2353/jmoldx.2006.050083.

Wan, T. S. K. (2017). Cancer Cytogenetics: An Introduction. In Wan, T. S. K. (ed.) *Cancer Cytogenetics, Methods and Protocols* pp. 1-10. New York, NY, USA: Springer Science + Business Media LLC. Available at: https://link.springer.com/book/10.1007/978-1-4939-6703-2 (accessed: 1 March 2024).

Wang, X., Wang, J. & Wu, J. (2021). Emerging roles for HMGA2 in colorectal cancer. *Translational Oncology*, 14 (1): 100894. doi: 10.1016/j.tranon.2020.100894.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13: 134. doi: 10.1186/1471-2105-13-134.

Zhang, F., Gu, W., Hurles, M. E. & Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annual Review of Genomics and Human Genetics*, 10: 451-81. doi: 10.1146/annurev.genom.9.081307.164217.

Zhang, S., Mo, Q. & Wang, X. (2019). Oncological role of HMGA2 (Review). *International Journal of Oncology*, 55 (4): 775-788. doi: 10.3892/ijo.2019.4856.

Zheng, H., Li, B. H., Liu, C., Jia, L. & Liu, F. T. (2020). Comprehensive Analysis of lncRNA-Mediated ceRNA Crosstalk and Identification of Prognostic Biomarkers in Wilms' Tumor. *BioMed Research International*, 2020: 4951692. doi: 10.1155/2020/4951692.

APPENDICES

A: Flowcharts of Methods

B: RNA Quality

C: Primer Locations in the Putative Fusion Transcript Sequences from deFuse

D: Primer Combinations

E: PCR Programs

F: Reagents Used in Methods Section

G: DNA Concentration and Quality of Purified PCR Products

H: Output from deFuse

I: Agarose Gel Images of PCR Products

J: Quality of aCGH Output

APPENDIX A: FLOWCHARTS OF METHODS

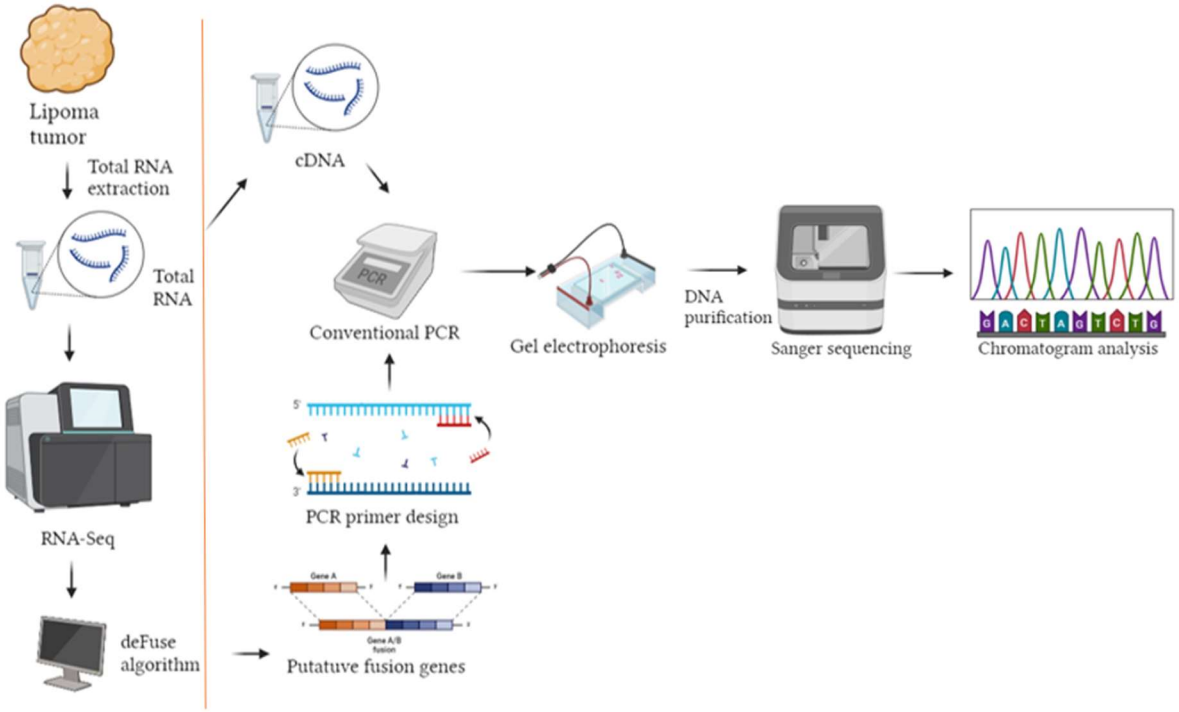

Figure A1. Brief overview of methods used to identify fusion transcripts. The methods to the right of the orange vertical line demonstrates the methods performed in this project. Created with BioRender.com.

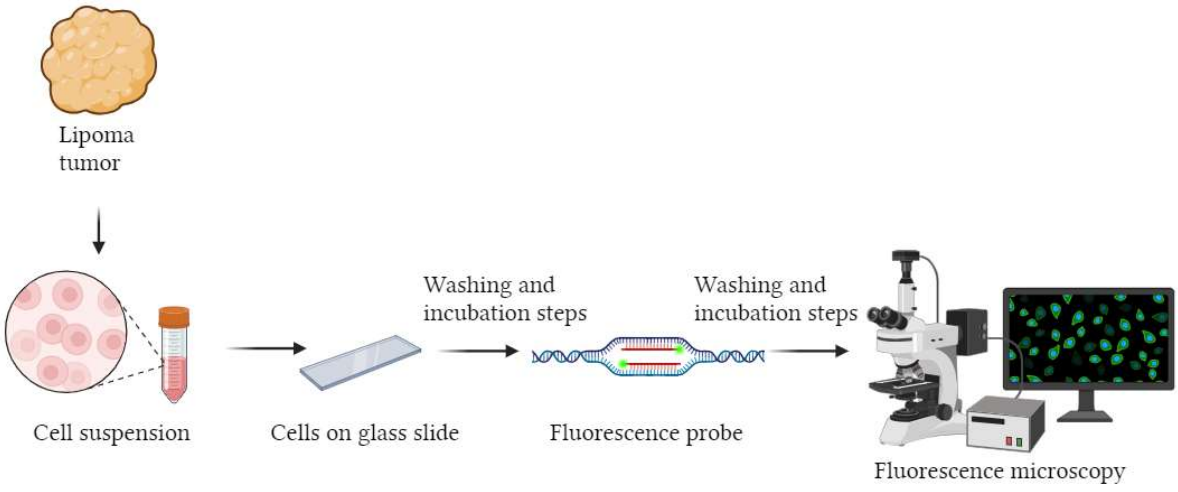

Figure A2. Brief overview of FISH analysis. Created with BioRender.com.

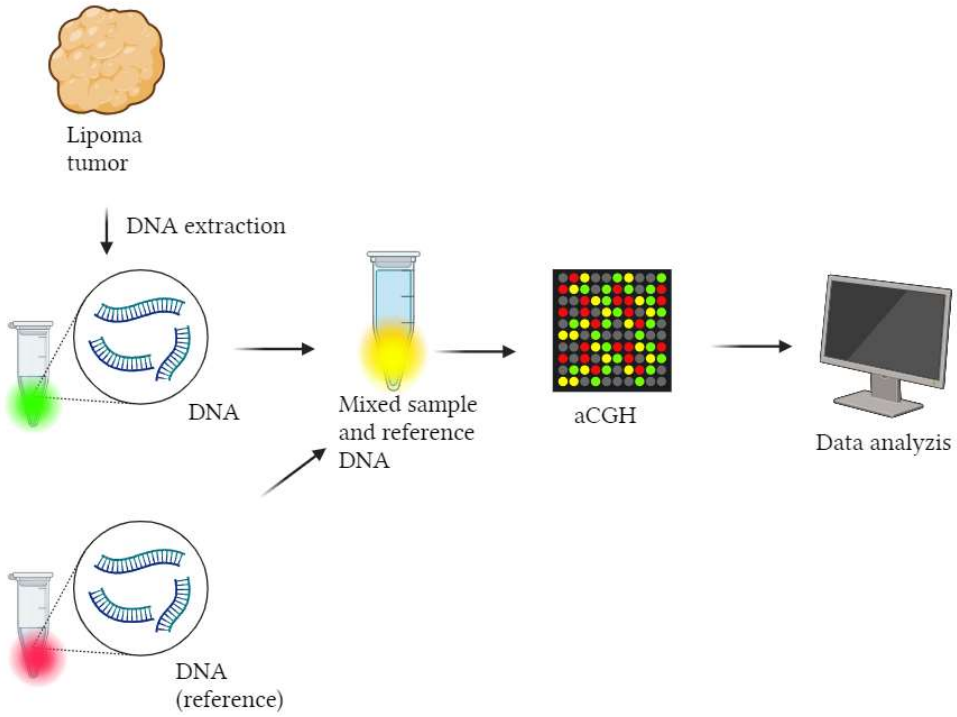

Figure A3. Brief overview of aCGH analysis. Created with BioRender.com.

APPENDIX B: RNA QUALITY

The measurements of RNA concentration and RNA quality for case 1 and 2, performed in advance of this project, as well as concentrations for control 1 and 2, are shown in Table B1.

Table B1. Total RNA concentration, RNA Integrity Number (RIN) and percentage of RNA fragments larger than 200 nucleotides for case 1 and 2, as well as total RNA concentrations for control 1 and 2.

Lipoma sample	**Total RNA concentration (ng/µl)**	**RIN**	**RNA fragments > 200 nt (%)**
Case 1	17.7	2.2	59
Case 2	24.4	4.3	64
Control 1	16.8	N/A*	N/A
Control 2	30.1	N/A	N/A

*N/A: not available.

APPENDIX C: PRIMER LOCATIONS IN THE PUTATIVE FUSION TRANSCRIPT SEQUENCES FROM DEFUSE

Case 1 – Fusion transcript nr. 1: ***HMGA2*** **exon 3** and *LINC00535* intron 4:

GTGGCAGCGGCGGCAGCCTAAGCAACAGCAGCCCTCGCAGCCCGCCAGCTCGCGCTCGCCC
CGCCGGCGTCCCCAGCCCTATCACCTCATCTCCCGAAAGGTGCTGGGCAGCTCCGGGGCGGT
CGAGGCGAAGCGGCTGCAGCGGCGGTAGCGGCGGCGGGAGGCAGGATGAGCGCACGCGGT
GAGGGCGCGGGGCAGCCGTCCACTTCAGCCCAGGGACAACCTGCCGCCCCAGCGCCTCAGA
AGAGAGGACGCGGCCGCCCCAGGAAGCAGCAGCAAGAACCAACCGGTGAGCCCTCTCCTAA
GAGACCCAGGGGAAGACCCAAAGGCAGCAAAAACAAGAGTCCCTCTAAAGCAGCTCAAAAG
AAAGCAGAAGCCACTGGAGAAAAACGGCCAAGAGGCAGACCTAGGAAATG*GGCCCCTGATG
TCACCAATCCAGAAATAAGGGCACATAGGCCCTAGGATGCCTTCAACAAGACCAAATTAAGGAGC
AAATGATCTCTTCTACCTGGTCCAGAATGTCTACCTAATAATGGCATCAACCACCAGCCACAGTGT
ATGCCAGTACTCAAATGCTGGTATCAAGCTAGCTGCGTATAAATTACCGGGGGAGTTGTGGGGCAG
GGAACTTACATCTTTAATGGATATTTCAGAAAATTCTAATTTACTCAGTCCTAATGGAAGCCCAGA
AATCCATATTTTTAAAAATCTCCCAGGTGATTCTGATGAATGGCTAAGTTTGCAAAATCCACCAATA
CATTTCCTTTATATAAAAATATAAGTGAAATTTCCATGAATTTATT

Case 1 – Fusion transcript nr. 2: ***HMGA2*** **exon 3** and *LINC00535* intron 4:

TGAGCGCACGCGGTGAGGGCGCGGGGCAGCCGTCCACTTCAGCCCAGGGACAACCTGCCGC
CCCAGCGCCTCAGAAGAGAGGACGCGGCCGCCCCAGGAAGCAGCAGCAAGAACCAACCGGT
GAGCCCTCTCCTAAGAGACCCAGGGGAAGACCCAAAGGCAGCAAAAACAAGAGTCCCTCTAA
AGCAGCTCAAAAGAAAGCAGAAGCCACTGGAGAAAAACGGCCAAGAGGCAGACCTAGGAAA
TG*GGCCCCTGATGTCACCAATCCAGAAATAAGGGCACATAGGCCCTAGGATGCCTTCAACAAGAC
CAAATTAAGGAGCAAATGATCTCTTCTACCTGGTCCAGAATGTCTACCTAATAATGGCATCAACCA
CCAGCCACAGTGTATGCCAGTACTCAAATGCTGGTATCAAGCTAGCTGCGTATAAATTACCGGGGG
AGTTGTGGGGCAGGGAACTTACATCTTTAATGGATATTTCAGAAAATTCTAATTTACTCAGTCCTAA
TGGAAGCCCAGAAATCCATATTTTTAAAAATCTCCCAGGTGATTC

Case 1 – Fusion transcript nr. 3: ***HMGA2*** **intron 2/exon 3** and *LINC00535* intron 4:

TCATGTATTTCAGAAGAGATCACTACAGGGGGTGTCTTTTAGGGCGAGGGGTTGCATAGATA
CCTCTTTACATCATGCAATGAAGAAGAATCTACTCAGAAATGTGGAAAAAGATTAACCTTAGA
GGAGACGAAGTTTGTTAAAAAAAACAAAAAAAAAAAAACCGATACGTCATCTGCAAAGCAGA
ACATTCTTACTTCAAATGTGTCCCTTGGTGCAGTACTTATAAACAATGTCAGGTAGAAAACTA
TAATGACTTCCTTTTTCATTTGCAGAAAGCAGAAGCCACTGGAGAAAAACGGCCAAGAGGCA
GACCTAGGAAATGGG*CCCCTGATGTCACCAATCCAGAAATAAGGGCACATAGGCCCTAGGATGC
CTTCAACAAGACCAAATTAAGGAGCAAATGATCTCTTCTACCTGGTCCAGAATGTCTACCTAATAA
TGGCATCAACCACCAGCCACAGTGTATGCCAGTACTCAAATGCTGGTATCAAGCTAGCTGCGTATA
AATTACCGGGGGAGTTGTGGGGCAGGGAACTTACATCTTTAATGGATATTTCAGAAAATTCTAATT

TACTCAGTCCTAATGGAAGCCCAGAAATCCATATTTTTAAAAATCTCCCAGGTGATTCTGATGAAT
GGCTAAG

Case 2 – Fusion transcript nr. 1: ***HMGA2*** **exon 3** and *LINC00535* intron 4:

CCCAGCCCTATCACCTCATCTCCCGAAAGGTGCTGGGCAGCTCCGGGGCGGTCGAGGCGAA
GCGGCTGCAGCGGCGGTAGCGGCGGCGGGAGGCAGGATGAGCGCACGCGGTGAGGGCGCG
GGGCAGCCGTCCACTTCAGCCCAGGGACAACCTGCCGCCCCAGCGCCTCAGAAGAGAGGAC
GCGGCCGCCCCAGGAAGCAGCAGCAAGAACCAACCGGTGAGCCCTCTCCTAAGAGACCCAG
GGGAAGACCCAAAGGCAGCAAAAACAAGAGTCCCTCTAAAGCAGCTCAAAAGAAAGCAGAA
GCCACTGGAGAAAAACGGCCAAGAGGCAGACCTAGGAAATG*GGCCCCTGATGTCACCAATCC
AGAAATAAGGGCACATAGGCCCTAGGATGCCTTCAACAAGACCAAATTAAGGAGCAAATGATCTC
TTCTACCTGGTCCAGAATGTCTACCTAATAATGGCATCAACCACCAGCCACAGTGTATGCCAGTAC
TCAAATGCTGGTATCAAGCTAGCTGCGTATAAATTACCGGGGGAGTTGTGGGGCAGGGAACTTACA
TCTTTAATGGATATTTCAGAAAATTCTAATTTACTCAGTCCTAATGGAAGCCCAGAAATCCATATTT
TTAAAAATCTCCCAGGTGA

Case 2 – Fusion transcript nr. 2: ***HMGA2*** **exon 3** and *LINC00535* intron 5:

GTGGCAGCGGCGGCAGCCTAAGCAACAGCAGCCCTCGCAGCCCGCCAGCTCGCGCTCGCCC
CGCCGGCGTCCCCAGCCCTATCACCTCATCTCCCGAAAGGTGCTGGGCAGCTCCGGGGCGGT
CGAGGCGAAGCGGCTGCAGCGGCGGTAGCGGCGGCGGGAGGCAGGATGAGCGCACGCGGT
GAGGGCGCGGGGCAGCCGTCCACTTCAGCCCAGGGACAACCTGCCGCCCCAGCGCCTCAGA
AGAGAGGACGCGGCCGCCCCAGGAAGCAGCAGCAAGAACCAACCGGTGAGCCCTCTCCTAA
GAGACCCAGGGGAAGACCCAAAGGCAGCAAAAACAAGAGTCCCTCTAAAGCAGCTCAAAAG
AAAGCAGAAGCCACTGGAGAAAAACGGCCAAGAGGCAGACCTAGGAAATGGC*AGATCGTAT
AAGACTAAGTCAAAGCTGATCCTCCTTCTAAACCCGCCTCTCCCCACAGCCCAAAATGTGCACTCT
GTGAACATAATTGCCCAAGGAAAACGGAAGCTTGAGCTCTCCGAGCATCCGTATTCATGGGAGCA
GCACTTTTGTACTTTGCAAGACCAGAATGGGCTCCCCACC

HMGA2-853FW

8q22-intron-SEQ2-94333331-Rev2

HMGA2-947FW

8q22-intron-SEQ3-94333290-Rev3

Chr12-66232065-SEQ1-Fw1

8q22-intron-94310329-SEQ4-Rev4

HMGA2-985F

8q22-intron-seq2-94333341-R

Overlap between 8q22-intron-seq2-94333341-R and 8q22-intron-SEQ2-94333331-Rev2

APPENDIX D: PRIMER COMBINATIONS

Table D1. Primer combinations for PCRs used to test for fusion genes in case 1 and 2.

Forward primer	Reverse primer	PCR product size (bp)
HMGA2-853FW	8q22-intron-SEQ2-94333331-Rev2	331
HMGA2-947FW	8q22-intron-SEQ3-94333290-Rev3	198
Chr12-66232065-SEQ1-Fw1	8q22-intron-SEQ3-94333290-Rev3	385
HMGA2-947FW	8q22-intron-94310329-SEQ4-Rev4	167
HMGA2-985F	8q22-intron-seq2-94333341-R	190

Table D2. Primer combinations for PCRs used to test for fusion genes in control 1 and 2.

Forward primer	Reverse primer	PCR product size (bp)
HMGA2-853FW	8q22-intron-SEQ2-94333331-Rev2	331
HMGA2-947FW	8q22-intron-SEQ3-94333290-Rev3	198

APPENDIX E: PCR PROGRAMS

Table E1. PCR program used to perform RT-PCR with HMGA2-985F and 8q22-intron-seq2-94333341-R primer combination. Temperature and time for different PCR stages are displayed.

Parameter	PCR stages				
	Incubation	35 cycles			Hold
		Denaturing	Annealing	Extension	
Temperature	96 ℃	96 ℃	57 ℃	60 ℃	12 ℃
Time (mm:ss)	01:00	00:20	00:30	04:00	∞

Table E2. PCR program used to perform PCR with the Chr12-66232065-SEQ1-Fw1 and 8q22-intron-SEQ3-94333290-Rev3, the HMGA2-947FW and 8q22-intron-SEQ3-94333290-Rev3, and the HMGA2-947FW and 8q22-intron-94310329-SEQ4-Rev4 primer combinations. Temperature and time for different PCR stages are displayed.

Parameter	PCR stages				
	Incubation	35 cycles			Hold
		Denaturing	Annealing	Extension	
Temperature	96 ℃	96 ℃	58 ℃	60 ℃	12 ℃
Time (mm:ss)	01:00	00:20	00:30	04:00	∞

APPENDIX F: REAGENTS USED IN METHODS SECTION

Table F1. Reagents and volumes mixed to visualize PCR products with gel electrophoresis for one sample.

Reagents	**Volume (µl)**
Nuclease-free water	9/7*
DNA Gel Loading Dye (6X) (containing GelRed)	2
GeneRuler 1 kb Plus DNA Ladder (0.5 µg/µl)/Sample (PCR product)	1/3*
Total	12

* For the DNA marker, 9 µl of nuclease-free water and 1 µl of marker were mixed. For each sample, 7 µl of nuclease-free water and 3 µl of PCR product were mixed.

APPENDIX G: DNA CONCENTRATION AND QUALITY OF PURIFIED PCR PRODUCTS

Table G1. DNA concentration and purity of PCR products after purification of the DNA and before sequencing reactions.

Case	Primer combination used for PCR	dsDNA [ng/µl]	A_{260}/A_{280}
1	HMGA2-853FW and 8q22-intron-SEQ2-94333331-Rev2	19.5	1.82
2	HMGA2-853FW and 8q22-intron-SEQ2-94333331-Rev2	12.3	1.75
1	HMGA2-947FW and 8q22-intron-SEQ3-94333290-Rev3	22.4	1.86
2	HMGA2-947FW and 8q22-intron-SEQ3-94333290-Rev3	11.8	1.80
1	Chr12-66232065-SEQ1-Fw1 and 8q22-intron-SEQ3-94333290-Rev3	7.9	1.62
2	Chr12-66232065-SEQ1-Fw1 and 8q22-intron-SEQ3-94333290-Rev3	6.8	1.99
1	HMGA2-947FW and 8q22-intron-94310329-SEQ4-Rev4	14.4	1.68
2	HMGA2-947FW and 8q22-intron-94310329-SEQ4-Rev4	19.0	1.78
1	ABL1-195F and ABL1-325R	18.1	1.67
2	ABL1-195F and ABL1-325R	24.5	1.80
1	HMGA2-985F and 8q22-intron-seq2-94333341-R	13.3	1.56
2	HMGA2-985F and 8q22-intron-seq2-94333341-R	12.1	1.96

APPENDIX H: OUTPUT FROM DEFUSE

	A	B	U	V	Y	Z	AC	AD	AE	AF
1	cluster_id	splitr_sequence	gene1	gene2	gene_chromosome1	gene_chromosome2	gene_location1	gene_location2	gene_name1	gene_name2
120	171113	GAAATGTGGCGTAC	ENSG00000141161	ENSG00000162669	17	1	intron	intron	UNC45B	HFM1
121	189091	CTTCATTCCCACGGT	ENSG00000170345	ENSG00000198626	14	1	coding	intron	FOS	RYR2
122	179645	CTGACTTTGATTTGC	ENSG00000124942	ENSG00000162669	11	1	coding	intron	AHNAK	HFM1
123	91850	CAGACAGAGTGCAA	ENSG00000227345	ENSG00000225784	10	10	coding	intron	PARG	RP11-592B15.4
124	51810	GTGGCAGCGGCGGC	ENSG00000149948	ENSG00000246662	12	8	coding	intron	HMGA2	LINC00535
125	166686	GTCCAGGTCAACGT	ENSG00000170323	ENSG00000162669	8	1	utr5p	intron	FABP4	HFM1
126	253569	CACGCCATACCCTTG	ENSG00000132356	ENSG00000113638	5	5	coding	utr5p	PRKAA1	TTC33
127	6237	CTCCGGAATCGAACC	ENSG00000264063	ENSG00000124942	21	11	downstream	coding	MIR3687	AHNAK
128	33808	CCTGTTACAGCTTAG	ENSG00000150687	ENSG00000174804	11	11	intron	coding	PRSS23	FZD4
129	282546	CCCAGCACCTGGAG	ENSG00000148180	ENSG00000264063	9	21	coding	downstream	GSN	MIR3687
130	150270	GCGAGCTCAGGGAG	ENSG00000162669	ENSG00000233654	1	2	intron	intron	HFM1	AC093388.3
131	28801	GCCGGGATTTGGGT	ENSG00000150991	ENSG00000150672	12	11	coding	intron	UBC	DLG2
132	182954	GGCTGAGGAGACCA	ENSG00000135404	ENSG00000162669	12	1	coding	intron	CD63	HFM1
133	151862	AACGAGCTGCTCTGC	ENSG00000241781	ENSG00000162669	9	1	downstream	intron	AL161626.1	HFM1
134	114956	TGTTTGTTTGTTTGTT	ENSG00000157306	ENSG00000105675	14	19	intron	upstream	RP11-66N24.4	ATP4A
135	69258	GGCGGCCCCGCCCC	ENSG00000260902	ENSG00000154783	3	3	intron	upstream	RP11-95M5.1	FGD5
136	190609	GTCTTATGCTCTACC	ENSG00000105675	ENSG00000162669	19	1	upstream	intron	ATP4A	HFM1
137	269986	AGGACCGCCTTTCCG	ENSG00000224905	ENSG00000256642	21	16	intron	upstream	AP001347.6	LINC00273
138	172767	CCCAGGACGAGGGC	ENSG00000129657	ENSG00000162669	17	1	intron	intron	SEC14L1	HFM1
139	61786	GGGAGTTGAATGTT	ENSG00000175899	ENSG00000141161	12	17	coding	intron	A2M	UNC45B
140	220883	TTGGGTGGGGGTGG	ENSG00000233991	ENSG00000143429	2	2	intron	intron	AC116050.1	AC027612.6
141	3186	ACCTGGCTAATTTTT	ENSG00000141161	ENSG00000150672	17	11	intron	intron	UNC45B	DLG2
142	287322	TACTTCTCCAGTTCG	ENSG00000164692	ENSG00000264063	7	21	coding	downstream	COL1A2	MIR3687

Figure H1. Some of the detected fusion genes for case 1 when using deFuse on RNA-Seq data. Fusion gene of interest is highlighted in green.

	A	B	U	V	Y	Z	AC	AD	AE	AF
1	cluster_id	splitr_sequence	gene1	gene2	gene_chromosome1	gene_chromosome2	gene_location1	gene_location2	gene_name1	gene_name2
528	78989	TCAAAGGGCTTCTTCT	ENSG00000073578	ENSG00000260774	5	5	coding	intron	SDHA	CTD-2083E4.4
529	307158	CAAGACCTTCGACTCT	ENSG00000113140	ENSG00000256642	5	16	coding	upstream	SPARC	LINC00273
530	157578	CCTGGAGGTGGAGGGC	ENSG00000133112	ENSG00000172572	13	12	coding	intron	TPT1	PDE3A
531	313126	GAGCATCGAGGGGGCC	ENSG00000256642	ENSG00000156194	16	4	upstream	intron	LINC00273	PPEF2
532	188429	CCCAGCCCTATCACCT	ENSG00000149948	ENSG00000246662	12	8	coding	intron	HMGA2	LINC00535
533	231510	ATACAGGAAGTGACGA	ENSG00000248333	ENSG00000008128	1	1	coding	coding	CDK11B	CDK11A
534	48604	GAAATGTGGCGTACGG	ENSG00000141161	ENSG00000226958	17	X	intron	intron	UNC45B	RNA28S5
535	9318	CCAGCGGGAGTCGCGC	ENSG00000150093	ENSG00000150672	10	11	utr5p	intron	ITGB1	DLG2
536	322081	GTTGTAGAGTTCAATG	ENSG00000170323	ENSG00000256642	8	16	coding	upstream	FABP4	LINC00273
537	242589	GTCCCTCTTAATCATGC	ENSG00000256642	ENSG00000162669	16	1	upstream	intron	LINC00273	HFM1
538	115795	CTTCCAGAAAATCCAC	ENSG00000184258	ENSG00000150672	X	11	coding	intron	CDR1	DLG2
539	153332	CTAGCTGTTGCTCCAG	ENSG00000196616	ENSG00000230876	4	2	coding	intron	ADH1B	LINC00486
540	169637	GCAAGGCGGACTAGAA	ENSG00000255652	ENSG00000153823	12	2	upstream	intron	RP11-313F23.4	PID1
541	270574	GTGGGTAATTTGCGCG	ENSG00000264063	ENSG00000233476	21	7	downstream	intron	MIR3687	EEF1A1P6
542	177633	CCATCAACTTCCATTGT	ENSG00000134532	ENSG00000255864	12	12	coding	intron	SOX5	RP11-444D3.1
543	190326	TTTGTTATAAAGATCCA	ENSG00000009694	ENSG00000232599	X	X	upstream	downstream	ODZ1	RP1-161N10.1
544	170002	GCGGACTAGAATGGAT	ENSG00000255652	ENSG00000144283	12	2	upstream	intron	RP11-313F23.4	PKP4
545	61638	CCTCCCCCTGGTTGCC	ENSG00000042753	ENSG00000160007	19	19	upstream	upstream	AP2S1	ARHGAP35
546	220762	TGAAGCCTTTTTTTAAA	ENSG00000232858	ENSG00000232527	13	1	upstream	intron	RPL34P27	RP11-14N7.2
547	115808	AACATCTCCCTTAAGCT	ENSG00000124942	ENSG00000226958	11	X	coding	intron	AHNAK	RNA28S5
548	311795	CCCAAAGACTTTGGTTT	ENSG00000256642	ENSG00000226958	16	X	upstream	intron	LINC00273	RNA28S5
549	86002	TGGAACCTAGACATCC	ENSG00000113140	ENSG00000150672	5	11	coding	intron	SPARC	DLG2
550	182272	GGCCGTCCGGCCGCGT	ENSG00000264063	ENSG00000172572	21	12	downstream	intron	MIR3687	PDE3A
551	138937	CCCGCCGCTCCATTAC	ENSG00000136826	ENSG00000163046	9	2	coding	intron	KLF4	ANKRD30BL

Figure H210. Some of the detected fusion genes for case 2 when using deFuse on RNA-Seq data. Fusion gene of interest is highlighted in green.

APPENDIX I: AGAROSE GEL IMAGES OF PCR PRODUCTS

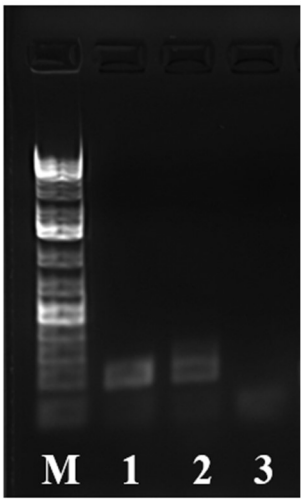

Figure I1. Agarose gel (1,5%) image of PCR products from case 1 and 2 run with the HMGA2-985F and 8q22-intron-seq2-94333341-R primer combination. M: 1 kb DNA marker, 1: case 1, 2: case 2, 3: negative control for the PCR (nuclease-free water).

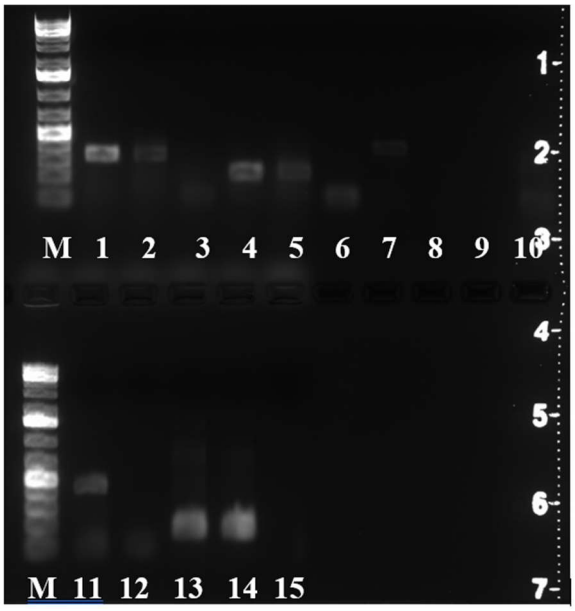

Figure I2. Agarose gel (1%) image of PCR products from case 1 and 2 run with different PCR primer combinations. HMGA2-853FW & 8q22-intron-SEQ2-94333331-Rev2 (well 1-3), HMGA2-947FW & 8q22-intron-SEQ3-94333290-Rev3 (well 4-6), chr12-66232065-SEQ1-Fw1 & 8q22-intron-SEQ3-94333290-Rev3 (well 7-9), HMGA2-947FW & 8q22-intron-94310329-SEQ4-Rev4 (well 10-12) and ABL1 185F & ABL1 325R (well 13-15). M: 1 kb DNA marker. Well 1, 4, 7, 10, 13: case 1. Well 2, 5, 8, 11, 14: case 2. Well 3, 6, 9, 12, 15: negative control (nuclease-free water).

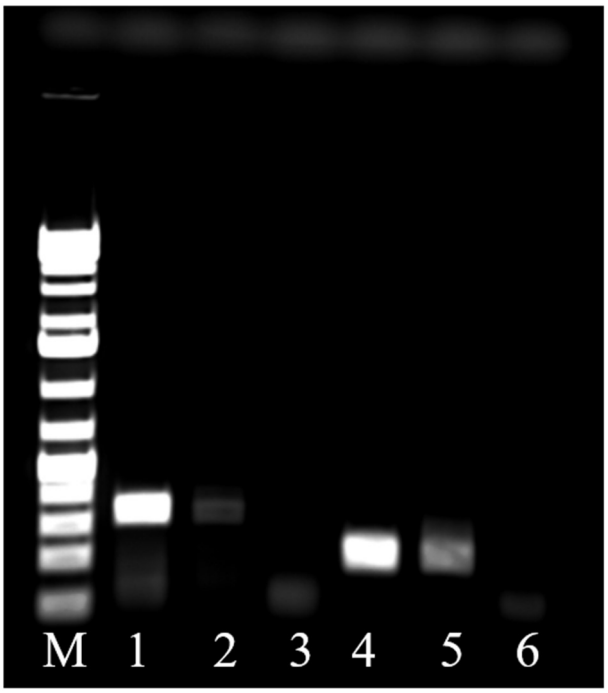

Figure I3. Agarose gel (1%) image of PCR products from case 1 and 2 run with different PCR primer combinations. HMGA2-853FW & 8q22-intron-SEQ2-94333331-Rev2 (well 1-3). HMGA2-947FW & 8q22-intron-SEQ3-94333290-Rev3 (well 4-6). M: 1 kb DNA marker. Well 1 & 4: case 1. Well 2 & 5: case 2. Well 3 & 6: negative control (nuclease-free water).

APPENDIX J: QUALITY OF aCGH OUTPUT

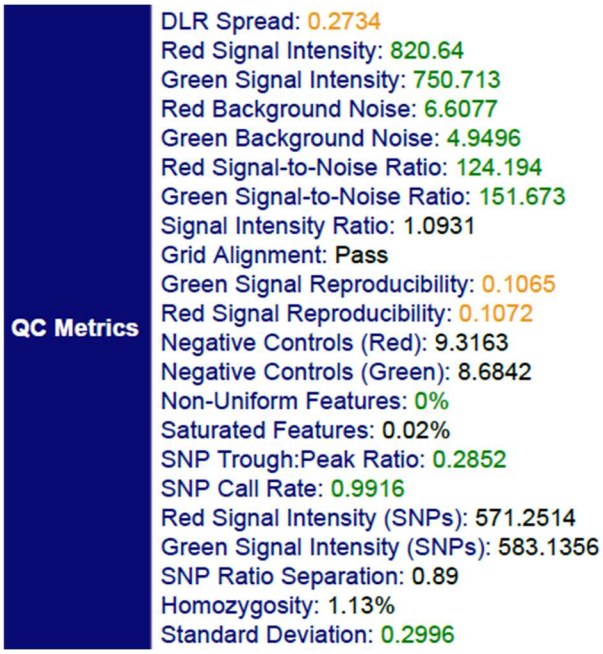

Figure J1. Reported quality control (QC) metrics for aCGH output for case 1. Green values are within the excellent threshold, while yellow values are within the satisfactory threshold reported in the CytoSure™ Interpret Software user guide (Oxford Gene Technology, 2017). Created from screenshot from the CytoSure™ Interpret Software (version 4.11.39).

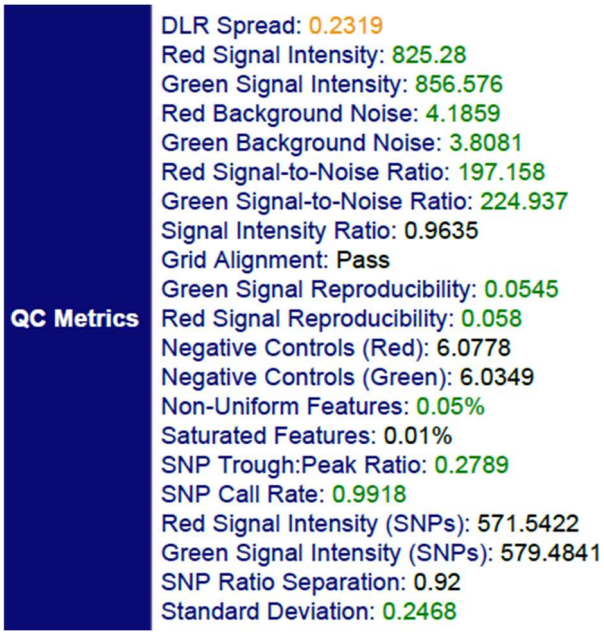

Figure J2. Reported quality control (QC) metrics for aCGH output for case 2. Green values are within the excellent threshold, while yellow values are within the satisfactory threshold reported in the CytoSureTM Interpret Software user guide (Oxford Gene Technology, 2017). Created from screenshot from the CytoSure™ Interpret Software (version 4.11.39).

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway