Vafadar-Isfahani, Natasha (2010) ChIP analysis of the histone modifications at the D4Z4 repeats in facioscapulohumeral muscular dystrophy (FSHD). MRes thesis, University of Nottingham.

ChIP analysis of the histone modifications at the D4Z4 repeats in Facioscapulohumeral muscular dystrophy (FSHD)

Natasha Vafadar-Isfahani

School of Biology

Thesis Submitted to the University of Nottingham for the degree of Masters of Research, September 2010

TABLE OF CONTENTS

LIST OF FIGURES

1. INTRODUCTION

3. MATERIALS AND METHODS

4. RESULTS

5. DISCUSSION

LIST OF TABLES

1. INTRODUCTION

3. MATERIALS AND METHODS

3. RESULTS

ABBREVIATIONS

˚C	Degrees Celsius
3C	Chromatin conformation capture
ANT1	Adenine nucleotide translocator 1
ATR-X	Alpha-thalassemia mental retardation, X-linked
bp	Base pair(s)
ChIP	Chromatin immunoprecipitation
CLB	Cell lysis buffer
Ct	Threshold cycle
DBE	D4Z4 binding element
DMD	Duchenne muscular dystrophy
DMSO	Dimethyl sulfoxide
DNMT	DNA methyltransferase
DRC	D4Z4 repressor complex
DUX4	Double homeobox 4
DUX4C	Double homeobox 4 centromeric
EDMD	Emery-Dreifuss muscular dystrophy
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
ES cells	Embryonic stem cells
FBS	Fetal bovine serum
FISH	Fluorescence in situ hybridisation
FRG1	FSHD region gene 1
FRG2	FSHD region gene 2
FSHD	Facioscapulohumeral muscular dystrophy
g	Grams
g	Centrifugal force
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
h	Hours
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HMGB2	High mobility group B2 protein
HP1	Heterochromatin binding protein
ICF	Immunodeficiency, centromeres instability and facial anomalies
l	Liter
LGMD	Limb-girdle muscular dystrophy
M	molar
Mb	megabase pair(s)
MBD	Methyl-CpG-binding domain
MeCP2	Methyl-CpG-binding protein 2
mg	milligram(s)
min	Minutes
ml	milliliter(s)
mM	millimolar
ng	nanogram(s)
NLB	Nuclear lysis buffer
nM	Nanomolar

ORF	Open reading frame
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PITX1	Paired-like homeodomain transcription factor 1
pmol	picomole(s)
PNK	Polynucleotide kinase
qPCR	Quantitative real-time PCR
rDNA	Ribosomal DNA
RTT	Rett syndrome
SDEV	Standard deviation
SDS	Sodium dodecyl sulfate
SE	Standard error
sec	Seconds
SNP	Single nucleotide polymorphism
SSLP	Single sequence length polymorphism
TE	Tris-EDTA buffer
TRD	Transcriptional repressor domain
TSS	Transcription start site
TUBB4Q	Tubulin, beta polypeptide 4, member Q
V/cm	Volts per centimeter
VNTR	Variable number tandem repeats
YY1	Ying Yang 1
µg	micrograms
µl	microliter
µM	micromolar

ACKNOWLEDGEMENTS

First of all I would like to thank Jane Hewitt and Cristina Tuffarelli for giving me the opportunity to work on this project and for all their help and advice. I would also like to thank all members of the JEH lab: Amanda for answering all my FSHD-related questions, Andy, the superman, for his help and input in IT problems and ChIP calculations, and Jo for brightening up the day to day lab experience and all her sequencing data. I would like to thank my boyfriend, Pablo and my sister, Baharak for their support throughout the year and last but not least, my parents for their encouragement and financial support.

ABSTRACT

Genomic DNA must exist in a particular chromatin configuration and modification of this structure is essential for the correct control of gene expression. There are several human genetic disorders that are caused by misregulation of epigenetic gene control. Facioscapulohumeral muscular dystrophy (FSHD) is a disease that may be caused by alterations in chromatin structure. FSHD is the third most common form of muscular dystrophy. The majority of FSHD cases show contraction of the D4Z4 repeats on the 4q35 chromosome (FSHD1). However, a small number of FSHD cases show no contraction at this region (FSHD2), but share epigenetic changes at the D4Z4 region with the FSHD1 patients. In 2009, Zeng et al. reported a specific loss of H3K9me3 histone modification at the D4Z4 repeats in FSHD patients.

The main focus of this study was to verify the published data by Zeng et al (2009) and further investigate the histone modification changes at the D4Z4 array. Chromatin immunoprecipitation (ChIP) coupled with real-time quantitative PCR (qPCR) was employed to investigate the histone modifications within the D4Z4 array.

The results obtained were in agreement with the previously published data on the reduction of H3K9me3 histone modification at the D4Z4 repeats in FSHD patients. However, contradictory to the previous data, the reduction of this histone modification was also observed on other genomic regions. A global reduction of H3K27me3 was also observed in FSHD patients.

1. Introduction

1.1 Concept of Epigenetics

The term epigenetics has been defined as heritable changes in gene expression that occur without a change in DNA sequence (Wolffe et al. 1999).

This is made possible due to epigenetic modification of the genome, which enables different gene regulation and expression in different cell types. Somatic cells in a multicellular organism all have essentially identical genomes, but each of these cells has a distinct structure and function. Gene regulation has two components, labile regulation and epigenetic regulation. Labile regulation in controlled by fluctuation of nuclear proteins concentration, covalent modifications and subunit associations of transcriptional activators and repressors. The chromatin structure, with its covalent modifications and stable protein associations, provides epigenetic regulation and constitutes the epigenotype. These two components of gene regulation act very closely and in harmony with each other and hence there is not always a clear boundary between their actions (Jiang et al. 2004).

1.1.1 Chromatin structure

An epigenotype is established primarily through the folding of DNA into chromatin and the architecture of the chromatin within the nucleus. In eukaryotic organisms, DNA is packaged with histones within the nucleus of the cell into either euchromatin or heterochromatin regions. Euchromatin is an expanded open structure that consists of mostly actively transcribed genes,

while heterochromatin is a contracted and closed structure, which represents transcriptionally inactive conditions. Identification of these two types of chromatin led to the hypothesis that differences in chromatin structure might be associated with tissue-specific or temporal differences in gene expression (reviewed by Jiang et al. 2004).

A first level of compaction of DNA (5-10x) is achieved by organization into nucleosomes. Approximately 146bp of DNA is wound twice around an octamer of core histones consisting of two of each of the four histones H2A, H2B, H3, and H4 (Figure 1.1). Adjacent nucleosomes are connected by a segment of linker DNA to form a 30-nm fiber, and the addition of histone H1 to each nucleosome lends stability to and facilitates the formation of further higher-order structures (Jiang et al. 2004).

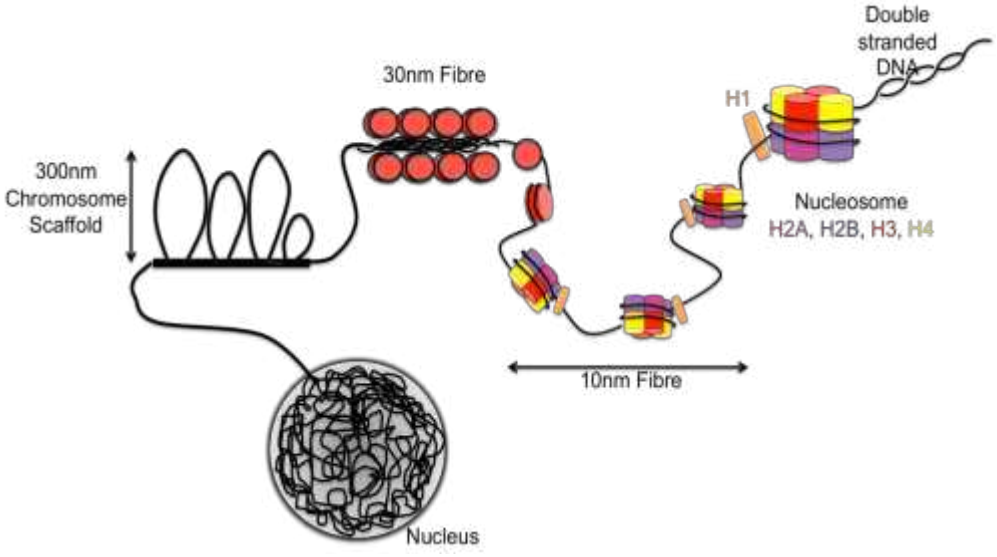


Figure 1.1 Structural organisation of nucleosome and chromatin packaging

1.1.2 Histone modifications

Amino acid residues within the histones, particularly the amino-terminal ends of the histone proteins, are subject to a wide range of different post-translational covalent modifications. There are at least eight distinct types of modifications found on histones which are shown in Table 1 (Kouzarides 2007). Histone modifications affect the histone-histone and histone-DNA interactions, thus affecting the overall structure of the nucleosome and how it interacts with the DNA. As a consequence histone modifications can effect the chromosome structure and gene expression. The entirety of these chemical modifications is called the "histone code" which provides the mammalian cell with another layer of gene regulation. The timing of the appearance of a modification depends on the signaling conditions within the cell (Kouzarides 2007).

The use of modification-specific antibodies in chromatin immunoprecipitation (ChIP) (section 1.4) has enabled study of histone modifications at particular sites such as promoters and enhancers (Barski et al. 2007; He et al. 2010). Monitoring of the global incidence of histone modifications is also possible by coupling of ChIP with DNA-microarray analysis (ChIP-CHIP) (De Gobbi et al. 2007; Pauler et al. 2009) or deep sequencing (ChIP-seq) (Pauler et al. 2009). Certain gene regulatory features have come to light regarding the composition and enrichment of certain histone modifications on various regions of the genome, which are reviewed in Kouzarides (2007).

Chromatin Modifications	Residues Modified	Functions Regulated
Acetylation	**K**-ac	Transcription, Repair, Replication, Condensation
Methylation (lysines)	**K**-me1 **K**-me2 **K**-me3	Transcription, Repair
Methylation (arginines)	**R**-me1 **R**-me2a **R**-me2s	Transcription
Phosphorylation	**S**-ph **T**-ph	Transcription, Repair, Condensation
Ubiquitylation	**K**-ub	Transcription, Repair
Sumoylation	**K**-su	Transcription
ADP ribosylation	**E**-ar	Transcription
Deimination	**R > Cit**	Transcription
Proline Isomerization	**P-cis > P-trans**	Transcription

Table 1.1 Overview of different histone modifications and the functions that have been associated with them.
Table from Kouzarides (2007)

Amongst all the known modifications, acetylation has the most potential to unfold chromatin since it neutralises the basic charge of the lysine. In agreement with this, in general, euchromatin regions of the genome are marked by high levels of acetylation (Kouzarides 2007). High levels of acetylation are detected in promoter regions of active genes (Bernstein et al. 2005; Kim et al. 2005; Roh et al. 2005). Histone acetylation has also been observed in intergenic regions, where it has been correlated with functional enhancers in various cell types (Heintzman et al. 2007; Roh et al. 2007). Histone acetylation is catalysed by Histone acetyltransferases (HATs). HATs are divided into three main families, GNAT, MYST and CBP/p300 (Sterner et al. 2000). In general these enzymes can acetylate more than one lysine but some limited specificity has been observed. For example, HAT1 can only acetylate lysine 5 and lysine 12 of histone H4 (Kouzarides 2007). Acetylation is reversed by histone deacetylase (HDAC) enzymes. Deacetylation correlates with chromatin condensation and gene transcription repression. HDACs are involved in multiple signaling pathways and they are present in a number of repressive

chromatin complexes. Generally, these enzymes do not show any specificity towards a particular acetyl group (Kouzarides 2007).

In comparison with all the various histone modifications, methylations at lysine and arginine residues are relatively stable and are therefore considered potential markers for carrying epigenetic information through the cell cycle. ChIP-Seq analysis (ChIP data analysed by direct sequencing) on human HeLa cells and T-cells has revealed that the three methylation states of lysine 4 of H3 (H3K4) were elevated surrounding the transcription start sites (TSSs). These signals move from mono- to di- to trimethylated H3K4 while approaching the TSSs (Barski et al. 2007; Heintzman et al. 2007). Methylation of H3K27 had previously been correlated with gene repression (Lee et al. 2006; Roh et al. 2006). The ChIP-seq analysis by Barski et. al. (2007) confirmed the correlation between di- and trimethylation of H3K27 with silence promoters. In contrast, H3K27me1 signals were higher at active promoters, particularly downstream of the TSSs (Barski et al. 2007). Methylation of H3K9 is often associated with heterochromatin formation and gene silencing and the data from Barski et.al. (2007) confirmed the presence of H3K9me2 and H3K9me3 in silenced promoters. However, some previous studies have reported the enrichment of H3K9me3 in some active promoters and actively transcribed regions (Vakoc et al. 2005; Squazzo et al. 2006).

The enzymes involved in histone methylation are the most specific and have been implicated in playing critical roles in development and pathological processes (Barski et al. 2007; Kouzarides 2007). These enzymes often work alongside other factors such as DNA methyltransferases or DNA polymerase II (Cedar et al. 2009). For instance, SUV39H1 and SUV39H2 are two of the

enzymes responsible for trimethylation of H3K9. By recruiting DNMT3A and DNMT3B to methylated CpG sites, SUV39H1 and SUV39H2 are responsible for heterochromatinisation of satellite repeats (Fuks et al. 2003; Lehnertz et al. 2003).

1.1.3 DNA methylation

In mammals, DNA methylation is found predominantly at the carbon-5 position of about 80% of all cytosines that are part of symmetrical CpG dinucleotides. Most 5-methylcytosines lie within retrotransposons, endogenous retroviruses, or repetitive sequences (Bestor et al. 1996; Yoder et al. 1996). Consequently it is thought that methylation may have evolved as a host defense mechanism to prevent the mobilization of these elements and to reduce the occurrence of chromosomal rearrangements. It is now clear that DNA methylation has essential roles in mammalian development and plays a crucial role in variety of different biological processes such as genomic imprinting (Ferguson-Smith et al. 2001) and X-inactivation (Lee et al. 1997). Unmethylated CpG dinucleotides are found mainly in short CpG-rich sequence domains known as CpG islands that are in the vicinity of gene promoters (Ng et al. 1999). Unmethylated CpG islands are normally found in transcriptionally active genes, whereas developmental and tissue-specific genes mostly appear to be methylated and silenced in different tissues (Kundu et al. 1999; Bird 2002).

Mouse embryological studies have shown that the methylation pattern is erased in early embryo and is re-established at approximately the time of implantation (Monk et al. 1987; Kafri et al. 1992). Methylation during development is

established by two counteracting mechanisms: a wave of indiscriminate de novo methylation (Okano et al. 1999) and a mechanism that ensures the CpG island remain unmethylated. The precise details of how CpG islands are protected from methylation are not completely understood. Early studies suggested that the protection might be directed by recognition of common cis-acting sequences located in CpG islands (Macleod et al. 1994; Siegfried et al. 1999) and mediated by active demethylation (Frank et al. 1991). Recent studies in mice suggest that histone modification might be accountable for the establishment of the basic DNA methylation profile during early mammalian development (Ooi et al. 2007). Under this model, the pattern of methylation of H3K4 across the genome might be formed in the embryo before de novo DNA methylation. DNA regions that are packaged with nucleosomes containing unmethylated H3K4 are subject to de novo DNA methylation while regions packaged with nucleosomes containing methylated H3K4 are protected from de novo DNA methylation (Ooi et al. 2007). H3K4 methylation might be directed by sequence-directed binding of RNA polymerase II, which was found to recruit specific H3K4 methyltransferases (Guenther et al. 2007). As RNA polymerase II is bound mostly to CpG islands in the early embryo, these regions are marked by methylated H3K4 and are protected from de novo DNA methylation during development.

DNA methyltransferase enzymes DNMT3A and DNMT3B, complexed with DNMT3L are responsible for de novo DNA methylation (Jia et al. 2007; Ooi et al. 2007). DNMT3A and DNMT3B show methyltransferase activity only on unmethylated DNA (CpG sites) (Okano et al. 1999). Some distinction between the functions of DNMT3A and DNMT3B has been achieved. DNMT3B -/- ES

cells lack methylation in minor satellite repeats located in centromeres, however DNMT3A -/- ES cells do not (Okano et al. 1999). DNMT3L is a closely related homologue that co-localises with DNMT3A and DNMT3B. DNMT3L lacks methyltransferase activity itself, but it recruits the methyltransferases to DNA by binding to H3 in the nucleosome. Consistent with the finding of anti-correlation between DNA methylation and the presence of methylated H3K4, contact between DNMT3L and the nucleosome is shown to be inhibited by all forms of methylation on H3K4 (Ooi et al. 2007).

Following differentiation, cells generally lose both their de novo DNA methylation activity and their ability to recognise and protect CpG islands. However, the basic DNA methylation pattern that is generated at the time of implantation is maintained throughout development due to actions of DNMT1 (Chuang et al. 1997). DNMT1 is associated with the replication complex and acts preferentially on hemimethylated DNA during replication to maintain methylation pattern through cell division (Gruenbaum et al. 1982; Leonhardt et al. 1992). It seems that the DNA methylation profile can act as a template to reconstructing epigenetic sate of the genome following each cell division. Methylated regions of DNA are reassembled in a closed configuration (heterochromatin) whereas unmethylated regions tend to get packaged into more open configurations (euchromatin) (Howard 2009). Consistent with this statement it was reported that the heterochromatin binding protein 1 (HP1) has the ability to recruit DNMT1 protein to mediate silencing of euchromatic genes (Smallwood et al. 2007).

DNMT2 has a weak DNA methyltransferase activity (Hermann et al. 2003) and is not essential for either global de novo or maintenance of the DNA

methylation in embryonic stem (ES) cells (Okano et al. 1998), but it is required for methylation of aspartic acid transfer RNA (tRNA(Asp)) (Goll et al. 2006).

1.1.4 Relationship between DNA methylation and histone modification

Although DNA methylation and histone modifications are carried out by different chemical reactions and different enzymes, there seems to be a close biological relationship between the two. It seems that the relationship can work in both directions: DNA methylation can serve as a template for some histone modifications after DNA replication, and histone methylation can direct DNA methylation (reviwed in Cedar et al. 2009). An example of how histone modification can effect DNA methylation was mentioned above. All forms of H3K4 methylation inhibit recruitment of DNA methyltransferases to that site by DNMT3L (Ooi et al. 2007). On the other hand, requirement of the DNA methyltransferase enzymes DNMT3A and DNMT3B for activity of the histone methyltransferase enzymes SUV39H1 and SUV39H2 (Lehnertz 2003, Fuks 2003), demonstrates how DNA methylation is involved in histone modification.

There is also evidence that DNA methylation is important for maintenance of patterns of histone modifications through cell division. It is likely that the replication fork disrupts the chromatin structure as it progresses along the DNA. Although not clearly understood, it seems that the DNA methylation pattern established during early development acts as a template to maintain the histone modification pattern throughout the cell division. Methylation of DNA at specific sites can modulate histone covalent modification through the

recruitment of methyl-binding proteins such as methyl-CpG- binding protein 2 (MeCP2). MeCP2 is an abundant nuclear protein encoded on the X chromosome; it is part of a complex that also consists of histone deacetylase (HDAC) and the co-repressor Sin3a that controls transcription through the deacetylation of core histones at the methylated regions (Jones et al. 1998; Nan et al. 1998). MeCP2 has an N-terminal, methyl-CpG binding domain (MBD) and a transcriptional repressor domain (TRD) (Nan et al. 1993). Presence of DNA methylation at a specific site also inhibits H3K9 dimethylation, perhaps through interactions of DNMT1 and G9a, a histone methyltransferase, with the replication complex (Esteve et al. 2006). There is also evidence that the DNA methylation inhibits H3K4 methylation (Lande-Diner et al. 2007), which is a marker associated with transcription activation.

There are still many mechanical details of how DNA methylation and histone modifications are coordinated at the molecular level that needs to be clarified. For example it is not clear how DNA methylation pattern is actually translated to produce the correct histone modification profile. There are also mysteries about how the formation of histone methylation patterns may affect de novo DNA methylation, because the presence of methylated H3K9 or H3K27 does not always lead to recruitment of DNMT3 proteins and de novo DNA methylation (Cedar et al. 2009).

1.2 Chromatin structure and disease

In 1995, Strohman suggested that we “redirect our attention to epigenetic regulation as a second informational system in parallel with the genome” to understand complex disease traits (Strohman 1995). Genomic DNA must exist

in a particular chromatin configuration and hence the genotype can only give rise to a phenotype through the epigenotype. The epigenotype shows far greater plasticity than the genotype in the normal development of an individual, and therefore it is reasonable to speculate that epigenetic errors could be a major contributor to human diseases (Jiang et al. 2004). It is now widely appreciated that modification of chromatin structure is integral to the correct control of gene expression in mammals. Several human genetic disorders result from mutation in components of chromatin or in enzymes involved in regulation of chromatin structure, these include Immunodeficiency Centromeres instability and Facial anomalies (ICF) syndrome, Rett syndrome (RTT) and Alpha-Thalassemia mental Retardation, X-linked (ATR-X) syndrome.

ICF syndrome is a rare recessive disease caused by mutation in the gene encoding the de novo DNA methyltransferase, DNMT3B (Hansen et al. 1999; Xu et al. 1999). DNMT3b mutations in ICF patients most likely impair, rather than completely abolish, the methyltransferase enzyme activity. There is no genome-wide hypomethylation in the DNA of ICF patients, but the hypomethylation is targeted to specific sequences such as satellite DNAs 2 (Jeanpierre et al. 1993) and the non-satellite repeats D4Z4 and NBL2 (Kondo et al. 2000).

RTT syndrome is a form of mental retardation that is caused by mutation in the gene at Xq28 that encodes the methylated-DNA binding protein MECP2 (Amir et al. 1999). MeCP2 is part of a transcription repressor complex and is thought to be responsible for silencing of certain neural genes (Stancheva et al. 2003). Moreover, it has been proposed that the extensive DNA replication-

independent replacement of H3 in neurons may rely on MeCP2 (Ahmed and Henikoff 2002). As discussed in section 1.1.4, there is a close interplay between DNA methylation and chromatin modifications. Therefore it is likely that silencing of certain genes involved in neurological development relies on DNA methylation and recruitment of MeCP2 (Stancheva et al. 2003).

ATR-X syndrome is associated with mutation in ATRX, an ATP-dependant chromatin remodeling protein. Abnormal levels of DNA methylation at repetitive elements have been reported in ATR-X patients (Gibbons et al. 2000). ATRX and MeCP2 were found to interact in vitro and co-localise at pericentromeric heterochromatin (Nan et al. 2007). Recently, it was reported that ATRX is required for normal recruitment of MeCP2, cohesin and the insulator protein CTCF. Hence loss of the protein in ATRX probably alters the expression of a connected network of imprinted genes in the postnatal brain (Kernohan et al. 2010). MeCP2 and cohesin are known to be involved in chromosomal looping (Horike et al. 2005; Hadjur et al. 2009), which suggests that ATRX may also be involved in chromatin loop formation and long-range chromatin interactions (Cunningham et al. 2010). Involvement of cohesin in the complex sheds light on the importance of cohesin for correct gene regulation.

Facioscapulohumeral Dystrophy (FSHD) is another disease that may be caused by alterations in chromatin structure. Although FSHD was linked to deletions within the macrosatellite, D4Z4, in 1992 (Wijmenga et al. 1992) the underlying mechanism of whereby these deletions cause FSHD is still unclear. Several disease mechanisms have been proposed to explain how FSHD is caused, most of which predict a chromatin structure change followed by loss of

control over expression of genes in or proximal to the D4Z4 repeats (reviewed by de Greef et al. 2008) . This study will further investigate the association of previously identified chromatin structure changes involved in FSHD pathogenesis.

1.3 FSHD

Facioscapulohumeral muscular dystrophy (FSHD) is the third most common form of muscular dystrophy with an incidence of 1 in 20,000 in the Caucasian population (Wijmenga et al. 1990; Lunt et al. 1991). FSHD is an autosomal dominant disorder that in most cases presents in the second decade of life and is predominantly charactarised by progressive weakness and atrophy of a highly selective muscle group (van der Maarel et al. 2005; Tawil et al. 2006). The onset of FSHD involves the weakening of certain facial muscles, such that patients may not be able to smile or close their eyes. As the disease progresses the muscular weakness spreads to the shoulders and upper arm muscles. In more severe cases the weakness can also affect the abdominal and foot extensor muscles affecting the posture and the individual's ability to walk. A notable characteristic feature of the disease is that the muscle weakness displays an asymmetric distribution (Lunt et al. 1991). Additional non-muscular symptoms such as deafness, retinovasculopathy, mental retardation and epilepsy are associated mostly with more severely affected patients (Funakoshi et al. 1998). There is a wide variability in the disease penetrance amongst the FSHD patients, ranging from individuals with very mild muscle weakness, who are almost unaware of being effected, to patients who are wheelchair dependant.

Almost all FSHD cases (~95%) show linkage to a polymorphic EcoRI fragment on the subtelomeric region of chromosome 4q35 (Wijmenga et al. 1991). Cloning of this EcoRI fragment revealed a highly polymorphic array of 3.3kb KpnI repeats (Figure 1.2) (Vandeutekom et al. 1993; Hewitt et al. 1994) that has been named D4Z4. In unaffected individuals the number of the 3.3kb repeats varies between 11 and 100, equivalent to EcoRI fragment of 40-300kb. However in the majority of FSHD patients the D4Z4 repeat array is contracted to 10-40kb containing 1 to 10 repeat units (Figure 1.3) (Wijmenga et al. 1992). Alleles with 1-3 repeats are usually correlated with the earlier age of onset and more severe phenotypes, although a linear relationship between the two has not been observed (Lunt et al. 1995; Tawil et al. 1996). However, loss of a complete D4Z4 array on one allele does not result in FSHD, suggesting that the repeats themselves play a direct role in this disease (Tupler et al. 1996).

Mosaicism for the D4Z4 contraction is a frequent observation in FSHD (Tawil et al. 2006). Mitotic D4Z4 contractions can be encountered in 40% of de novo families, equally divided between affected mosaic patients and non-affected mosaic carriers parents of a non-mosaic affected child (Tawil et al. 2006). The male karyotype has a prominent effect in mosaic cases, where mosaic males are most affected despite having an equal complement of affected cells to a mosaic female. Mosaicism can directly affect the disease presentation and transmission risk, if present in the germ line (Tawil et al. 2006).

What makes the picture even more complex is the fact that around 5% of FSHD patients show no contraction of D4Z4 on chromosome 4q (termed phenotypic FSHD or FSHD2) (Gilbert et al. 1993). These patients share epigenetic changes at the D4Z4 locus with patients with the contraction,

suggesting that this disease mechanism is also mediated through D4Z4 (van Overveld et al. 2003; de Greef et al. 2007). Although FSHD was linked to the 4q35 D4Z4 repeats in 1992 (Wijmenga et al. 1992) the underlying mechanism of whereby these deletions cause FSHD is still unclear.

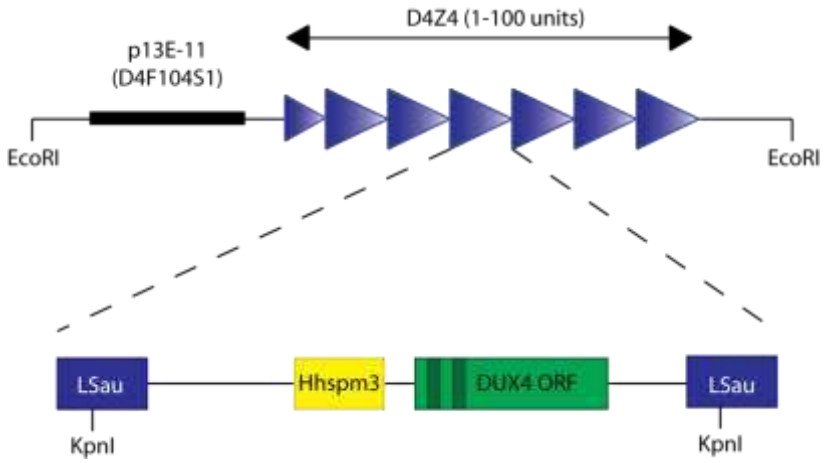


Figure 1.2 The EcoRI fragment identified by the p13E-11 probe.
P13-E11 (D4F10451) region is used as a DNA probe to identify the EcoRI fragment. Each arrowhead represents one 3.3Kb KpnI repeat and the sequence motifs within each individual repeat are shown. Each repeat consists of two low copy CG-rich repeat elements, LSau and Hhspm3 and an open reading frame (ORF) encoding a putative transcription factor with two homeodomains known as DUX4.

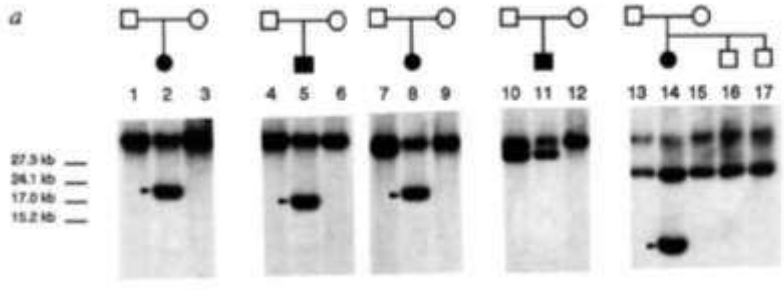


Figure 1.3 Southern analyses with p13E-11 of EcoRI-digested DNA from new cases of FSHD.

In four of the five different cases of FSHD a new EcoRI fragment (arrowhead) is present which is absent in either unaffected parent. The length of the EcoRI fragment is different in all affected individuals and varies between 14-28 kb. Image taken from Wijmenga et al. (1992).

1.3.1 D4Z4 array

D4Z4-like repeats are not restricted to chromosome 4qter. Sequences homologous to D4Z4 have been identified on the short arm of the acrocentric chromosomes (Wijmenga et al. 1992; Hewitt et al. 1994; Winokur et al. 1994). As a result of an ancient duplication a similar tandem array is located on chromosome 10q26 that has 98% nucleotide identity with the D4Z4 repeats on 4q35 (Deidda et al. 1995; van Geel et al. 2002). This region of homology starts from a region 40kb proximal to the D4Z4 repeats and extends into the telomeric repeat itself (van Geel et al. 2002). Similar to the 4q35 locus, 10q26 locus is also polymorphic for the number of D4Z4 repeats but short arrays at this locus are not associated with FSHD (Lemmers et al. 2001; Zhang et al. 2001). Initially the presence of this 10q array complicated the genetic diagnosis of FSHD, since the short 10q EcoRI fragments could not be

distinguished from those arising from 4q35. In 1996 a BlnI recognition site was found to be specific to 10q array (Deidda et al. 1996). Hence a double digestion with EcoRI and BlnI is now used for the genetic diagnosis of FSHD (Deidda et al. 1996).

There are two variants of the 4q subtelomere, termed 4qA and 4qB (van Geel et al. 2002). The most prominent difference between these two variants is the presence of a region of ß-satellite DNA distal to D4Z4 on 4qA type alleles, which is absent from the 4qB chromosomes (Figure 1.4E) (van Geel et al. 2002). Despite the equal frequencies of A and B alleles in the population, FSHD is exclusively associated with contracted D4Z4 repeats on the 4qA allele variant (Lemmers et al. 2002; Lemmers et al. 2004).

There are multiple haplotypes of chromosome 4q identified based on simple sequence length polymorphisms (SSLP) situated 3.5kb proximal to the repeats (Figure 1.4B) and a G/C single nucleotide polymorphism (SNP) on the first repeat (Figure 1.4D). In addition, there are fifteen different SNPs in the p13-E11 (D4F104S1) region (Figure 1.4C) (Lemmers et al. 2007). So far, contraction of D4Z4 repeats only on a certain haplotype of 4qA (4qA161) is associated with FSHD. Contraction of D4Z4 on other common haplotypes is nonpathogenic (Lemmers et al. 2007), suggesting that the contraction of D4Z4 repeats is necessary but not sufficient to cause FSHD (Lemmers et al. 2004). It is thought that certain SNPs might alter the chromatin structure of the 4q region resulting in the disease phenotype (Lemmers et al. 2007).

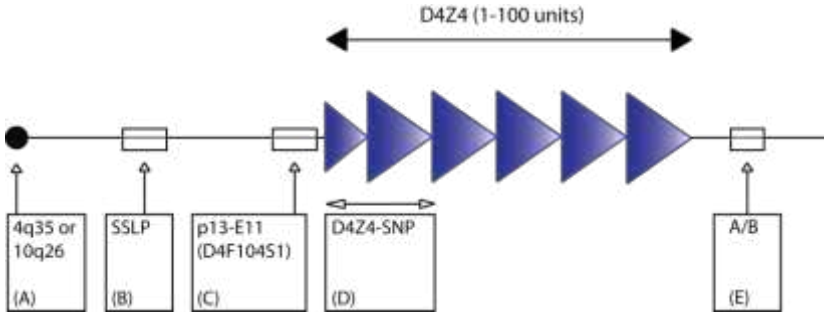


Figure 1.4 Schematic overview of the 4q and 10q subtelomeric region and the key sequence variations.

(A) Chromosomal localisation of the D4Z4 repeats, chromosome 4q or 10q. (B) This simple sequence length polymorphism (SSLP) is a combination of five variable number tandem repeats (VNTRs): a 8bp insertion/deletion, two SNPs localised 3.5kb proximal to D4Z4 and varies in length between 157-182bp. (C) the p13-E11 (D4F104S1) is localised immediately proximal to D4Z4, is 475bp in length and contains 15 SNPs. (D) The most proximal D4Z4 unit contains several SNPs. (E) The two variants of the subtelomeric region distal to D4Z4. Image modified from Lemmers et al. 2010.

1.3.2 Repressor complex associated with D4Z4

To investigate whether the D4Z4 region has any involvement in transcriptional control of the 4q35 genes, the interaction between D4Z4 and nuclear proteins was analysed using an electrophoretic mobility shift assay (EMSA) (Gabellini et al. 2002). This study revealed that a 27bp sequence, called the D4Z4 binding element (DBE) within the D4Z4 region, supported formation of a repressor protein complex (Gabellini et al. 2002). This repressor protein complex comprises of YY1 (Ying Yang 1), HMGB2 and nucleolin, termed the D4Z4 repressing complex (DRC) (Gabellini et al. 2002). YY1 is a complex protein involved in repression and activation of a number of genes. YY1 is known to

interact with numerous key regulatory proteins, suggesting that these interactions are important for determining whether YY1 is involved in gene repression or activation (Thomas et al. 1999). Since DBE contains a putative YY1 recognition sequence, CCATN (Yant et al. 1995), YY1 is the most likely component of DRC to interact directly with the DBE within D4Z4 (Gabellini et al. 2002). HMGB2 is a member of one of the three families of high mobility group (HMG) proteins. A yeast two-hybrid screen identified SP100 to interact with HMGB2. SP100 in turn interacts with HP1, the heterochromatin binding protein. This raises the possibility that HMGB2 might be involved in heterochromatin formation/maintenance (Lehming et al. 1998).

It has been proposed that contraction of the D4Z4 repeats may result in decreased repressor complex binding, leading to miss-regulation of distant neighboring genes (Gabellini et al. 2002). Loss of control is proposed to occur through cis-looping or cis-spreading mechanism, but chromatin studies had failed to identify any cis-acting mechanism operating at 4qter in FSHD (Jiang et al. 2003) until more recently (Petrov et al. 2006; Pirozhkova et al. 2008; Bodega et al. 2009). The findings of these studies are discussed in the next section.

1.3.3 Candidate genes for FSHD and transcription from chromosome 4q35

Assuming that contraction of D4Z4 would influence the transcriptional control of genes in cis, the search for a candidate gene for FSHD was focused on the region surrounding the D4Z4 repeats (Figure 1.5). D4Z4 lies adjacent to a subtelomeric sequence, which is within 20-30kb of a telomeric repeat (van

Geel et al. 2002) and no genes have been described or predicted distal to the array. The first gene identified in the 4qter was FSHD region gene 1 (FRG1) situated at 120kb proximal to D4Z4 (vanDeutekom et al. 1996), encoding a nuclear protein involved in RNA biogenesis (van Koningsbruggen et al. 2007). Chromatin conformation capture (3C) experiments have been taken to indicate that D4Z4 physically interacts with the promoter region of the FRG1, which represses FRG1 expression in myoblasts (Bodega et al. 2009). Using chromatin immunoprecipitation (ChIP) data coupled with data obtained from 3D fluorescence in situ hybridisation (3D-FISH), it was revealed that during normal myogenic differentiation this interaction is loosened and FRG1 expression is up-regulated (Bodega et al. 2009). A small reduction in the frequency of loop formation between D4Z4 array and FRG1 promoter region was observed in FSHD myoblast with contracted D4Z4, but the difference was too small to be involved in mis-regulation of FRG1 gene expression (Bodega et al. 2009). Previously, Gabellini et al. had reported altered expression of FRG1 (2002) and later on showed that over-expression of Frg1 in mice induced a muscle phenotype (Gabellini et al. 2006). However, FRG1 expression in FSHD muscle is highly controversial as several other groups have failed to observe FRG1 up-regulation (vanDeutekom et al. 1996; Gabellini et al. 2002; Jiang et al. 2003; Winokur et al. 2003; Masny et al. 2010). Expression studies have yielded different results, possibly as a consequence of different sources of RNA (site of muscle biopsy, severity of pathology) and different methods to detect expression (Tawil et al. 2006). It has also been suggested that since FRG1 is a multiple copy gene, also present on chromosomes 8, 9, 12, 20 and

all of the acrocentric chromosomes, some of the positive expression signals may come from these sites (Masny et al. 2010).

The closest putative gene to D4Z4 is situated 37kb proximal to the repeats and is termed FSHD region gene 2 (FRG2) (Rijkers et al. 2004). FRG2 was found to be expressed in differentiated FSHD primary myoblast cultures, but not in those of control individuals (Rijkers et al. 2004). The FRG2 homologue on chromosome 10q was also found to be upregulated in differentiating FSHD myoblasts. However, FRG2 absence from 4q in some FSHD patients carrying a proximally extended deletion of D4Z4 (Lemmers et al. 1998; Lemmers et al. 2003) makes the involvement of this gene in FSHD pathogenesis less likely.

Approximately 40kb proximal to D4Z4, lies an inverted and truncated copy of a D4Z4 unit (Wright et al. 1993). This inverted unit defines the proximal boundary of the region of homology between chromosome 4q35 and 10q26 and also contains a putative double homeobox gene DUX4C (double homeobox 4 centromeric), which is absent on chromosome 10q (van Geel et al. 2002). Expression of DUX4C had not been documented until recently. In 2008 a study indicated a possible function for DUX4C in differentiation of mouse C2C12 cells (a mouse myoblast cell line) (Bosnakovski et al. 2008). More recently Ansseau et al. (2009), reported the presence of a putative promoter, which was shown to be functional by fusion upstream of a luciferase reporter gene and transfection of C2C12, TE671 and HeLa cells (2009). In the same study, it was shown that the over-expression of DUX4C could activate human myoblast proliferation and inhibit their differentiation in vitro. It was suggested that this process might be caused by DUX4C inducing MYF5 myogenic regulator and its DNA-binding activity (Ansseau et al. 2009). However, there

are a number of FSHD patients with a proximal deletion that extends to the inverted repeat and hence DUX4C cannot be causative of FSHD.

Another extensively studied FSHD candidate gene is TUBB4Q (Tubulin, beta polypeptide 4, member Q). Although only 80kb proximal to D4Z4, and hence, positionally a good candidate, TUBB4Q is most likely to be a pseudogene (van Geel et al. 2000).

Some genes at a larger distance from D4Z4 have also been considered for their involvement in FSHD. At around 3.5Mb proximal lies the adenine nucleotide translocator-1 gene (ANT1) (Li et al. 1989). ANT1 is involved in the export of ATP through the mitochondrial membrane and is highly expressed in heart and skeletal muscles (Stepien et al. 1992). ANT1 is also known to be involved in apoptosis (Doerner et al. 1997). There is evidence for up-regulation of ANT1 at the protein level in affected and unaffected FSHD muscles (Laoudj-Chenivesse et al. 2005); however, further verification from independent studies is required.

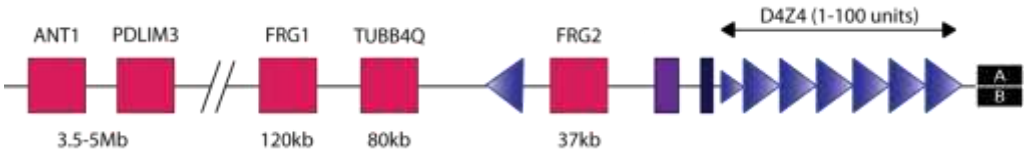


Figure 1.5 Illustration of the positioning of the D4Z4 region on 4q and the known genes proximal to the repeats.
The genes are shown using pink boxes and their distance from the D4Z4 repeats is shown under each gene. The diagram is not drawn to scale.

In summary there is little plausible evidence for the involvement of genes proximal to D4Z4 in FSHD.

The D4Z4 unit itself, as outlined in Figure 1.2, consists of several sequences known to be associated with heterochromatin regions, such as low copy CG-rich repeat elements LSau and hhspm3 (Zhang et al. 1987; Hewitt et al. 1994; Winokur et al. 1994, Meneveri, 1993 #82; Lee et al. 1995). In addition, each of the 3.3kb repeats contains an 1173bp open reading frame (ORF) encoding a putative transcription factor with two homeodomains (Hewitt et al. 1994). This ORF is now known as DUX4 (double homeobox 4) (Gabriels et al. 1999) and is thought to behave as a transcriptional activator (Kawamura-Saito et al. 2006). The DUX4 ORF has been evolutionary conserved for over 100 million years, which supports the idea of an protein coding function for DUX4 (Clapp et al. 2007). These findings, in addition to the positioning of the DUX4 ORF, make this gene a strong candidate for FSHD.

Cells transfected with DUX4 constructs showed characteristic markers of apoptosis suggesting that the protein was toxic to the cells (Gabriels et al. 1999; Kowaljow et al. 2007). However, the toxicity of DUX4 could be derived from its over-expression and therefore may not indicate the real function of DUX4 in vivo. Exogenous expression of DUX4 was shown to activate transient expression of a luciferase reporter gene fused to the Pitx1 (paired-like homeodomain transcription factor 1) promoter as well as the endogenous Pitx gene in transfected C2C12 cells (Dixit et al. 2007). Additionally, genome-wide expression data from muscle biopsies of FSHD patients showed up-regulation of PITX1 transcription in comparison with 11 other neuromuscular disorders (Dixit et al. 2007). However, the data from this study is not supported by other

studies of global gene expression analysis by different groups (Tupler et al. 1999; Winokur et al. 2003; Celegato et al. 2006).

A putative promoter region 149bp upstream of the ORF was inserted upstream of a luciferase reporter gene and was used to transfect TE671 cells. Expression of the reporter gene verified that the region could act as a functional promoter (Gabriels et al. 1999). Nevertheless, confirmation of the expression of full-length RNA and protein from the DUX4 ORF has been somewhat challenging. However, a recent study has reported a low abundance of DUX4 protein and transcripts in some FSHD cells (Dixit et al. 2007). A later study reported the transcription of several overlapping sense and anti-sense RNA transcripts from the 4q D4Z4 region (Snider et al. 2009). Although a full-length DUX4 transcript that can be detected at low levels is initially produced in muscle cells, spliced and cleaved forms are at higher abundance. Cleavage and processing of the DUX4 transcript generates mi/siRNA (small interference RNA)-sized fragments that might have a function in local or distance chromatin silencing or even target RNA from other loci (Snider et al. 2009).

In summary, the most convincing evidence suggests that DUX4 is the best candidate gene for FSHD. As mentioned above, maintenance of at least one D4Z4 repeat is essential for the FSHD pathogenesis (Tupler et al. 1996), which indicates the important role the D4Z4 repeats themselves play in FSHD pathogenesis.

1.3.4 Sub-nuclear localisation of 4q35

Chromosomes occupy distinct territories in the mammalian nucleus. These chromosome territories often reflect the gene density, transcriptional activity, size and replication timing (Sun et al. 2000; Tanabe et al. 2002). Unlike all other chromosomal ends studied, including 10qter, 4qter preferentially localises at the nuclear periphery, independent of cell type and chromosome territory effects (Masny et al. 2004; Tam et al. 2004). It is not the D4Z4 repeats themselves, but sequences proximal to the D4Z4 that are necessary for the perinuclear localisation (Masny et al. 2004). This might explain the difference between 4qter and the 10qter nuclear localisation, since the 98% identity between them only extends 40kb proximal to D4Z4. Although no difference has been observed in localisation of chromosomes between unaffected individuals and FSHD patients, the perinuclear localisation of the 4qter is largely lost in fibroblast lacking lamin A/C (Masny et al. 2004). Mutation in lamin A/C is known to cause Emery-Dreifuss muscular dystrophies (EDMD), thus raising the possibility that FSHD is related to these nuclear envelope muscular dystrophies (Bakay et al. 2006). More recently it was shown that the D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependant transcription insulator and it behaves as a strong perinuclear cis-acting positioning element in FSHD, properties that are lost upon multimerisation of the repeats (Ottaviani et al. 2009). The same group later showed that CTCF and A-type lamin only contribute to and are not essential to the positioning of D4Z4, indicating additional pathways for the anchoring of this region (Ottaviani et al. 2009). The proposed model for the effects of D4Z4

contraction/demethylation in FSHD on 4q35 perinuclear positioning is shown in Figure 1.6.

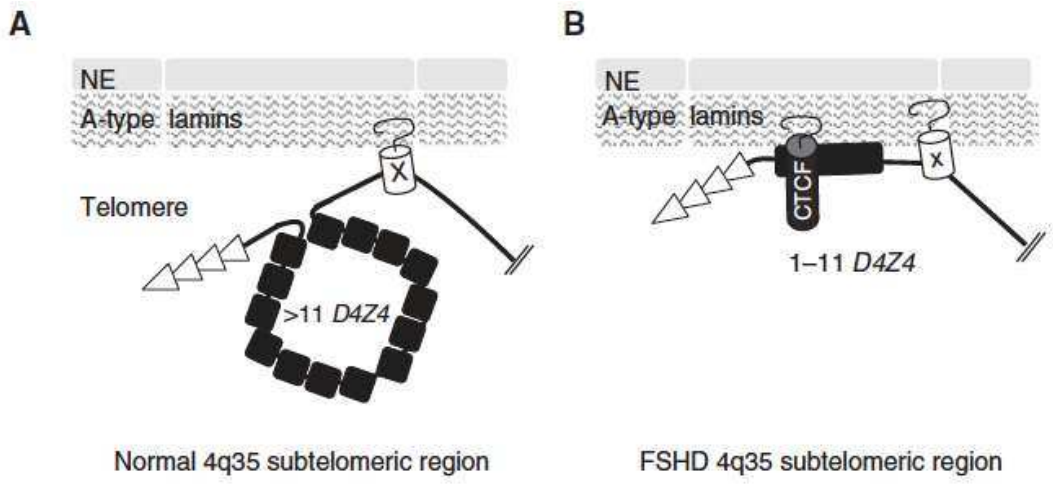


Figure 1.6 A model for tethering of the 4q35 locus at the nuclear periphery.

(A) In normal cells, D4Z4 repeats are methylated and therefore are not bound by CTCF. The 4q35 region is attached to the nuclear periphery through a lamin-A-dependant tethering site located outside the D4Z4 array. (B) In FSHD cells contraction of the D4Z4 repeats allows the binding of CTCF and the D4Z4 repeats become specifically attached to the periphery. This higher order switch is thought to modify the expression of FSHD-associated genes probably by preventing their repression (Ottaviani et al. 2009).

1.3.5 DNA methylation at D4Z4

Human D4Z4 repeats are extremely CG-rich, making them attractive candidates for DNA methylation. Hypomethylation of the proximal D4Z4 repeats in patients with 4q-linked and phenotypic FSHD was observed in both mutated and wild type allele with no tissue-specific methylation differences between DNA isolated from blood lymphocytes and muscles (van Overveld et al. 2003). In 2009 de Greef et al. showed that the D4Z4 contraction below a certain threshold results in significant hypomethylation, regardless of the haplotype on which D4Z4 contraction occurs (2009). In addition, D4Z4 methylation was observed not only on both chromosome 4q, but also on both chromosomes 10q in phenotypic FSHD patients. This argues that an underlying genetic defect, which is responsible for methylation of D4Z4 repeats, exists in FSHD2 patients (de Greef et al. 2009).

More severe hypomethylation at the 4q and 10q repeats is observed in a different group of patients with immunodeficiency-centromeric insatiability-facial anomalies (ICF) (Kondo et al. 2000; van Overveld et al. 2003). ICF syndrome presents with immunodeficiency, facial anomalies, mental retardation and developmental delay and is caused by mutations in DNMT3B (Xu et al. 1999). As these mutations reduce the methyltransferase activity of DNMT3B, hypomethylation of number of repeat arrays, including satellite 2 (Sat2), satellite 3 (Sat3), the NBL2 repeat and the D4Z4 repeat, is observed in ICF patients (Jeanpierre et al. 1993; Xu et al. 1999; Kondo et al. 2000). Since the clinical presence of ICF syndrome shares no similarity with FSHD and seeing as D4Z4 hypomethylation occurs on non-permissive haplotypes in unaffected individuals (de Greef et al. 2009), it can be said that the

hypomethylation of the D4Z4 allele alone is not sufficient to cause FSHD. Other epigenetic factors or haplotype-specific sequence polymorphisms may eventually determine the underlying mechanism to development of the FSHD phenotype (de Greef et al. 2009).

1.3.6 Histone modifications at D4Z4 repeat

As discussed previously, DNA methylation is often associated with increased chromatin condensation and gene inactivation. The D4Z4 repeat array is often considered heterochromatic, mainly because of the presence of the low copy repeats hhspm3 and LSau, which are mainly found in heterochromatic regions of the human genome (Hewitt et al. 1994). However, several studies have reported that D4Z4 repeat array consists of both heterochromatic and euchromatic regions (Jiang et al. 2003; Yang et al. 2004; Zeng et al. 2009). Double-ChIP analysis has revealed that H3K9me3 coincide with H3K27me3, but not H3K4me2, suggesting that the D4Z4 repeats contain distinct heterochromatin regions, marked by H3K9me3 and H3k27me3 as well as euchromatin regions, marked by H3K4me2 (Zeng et al. 2009).

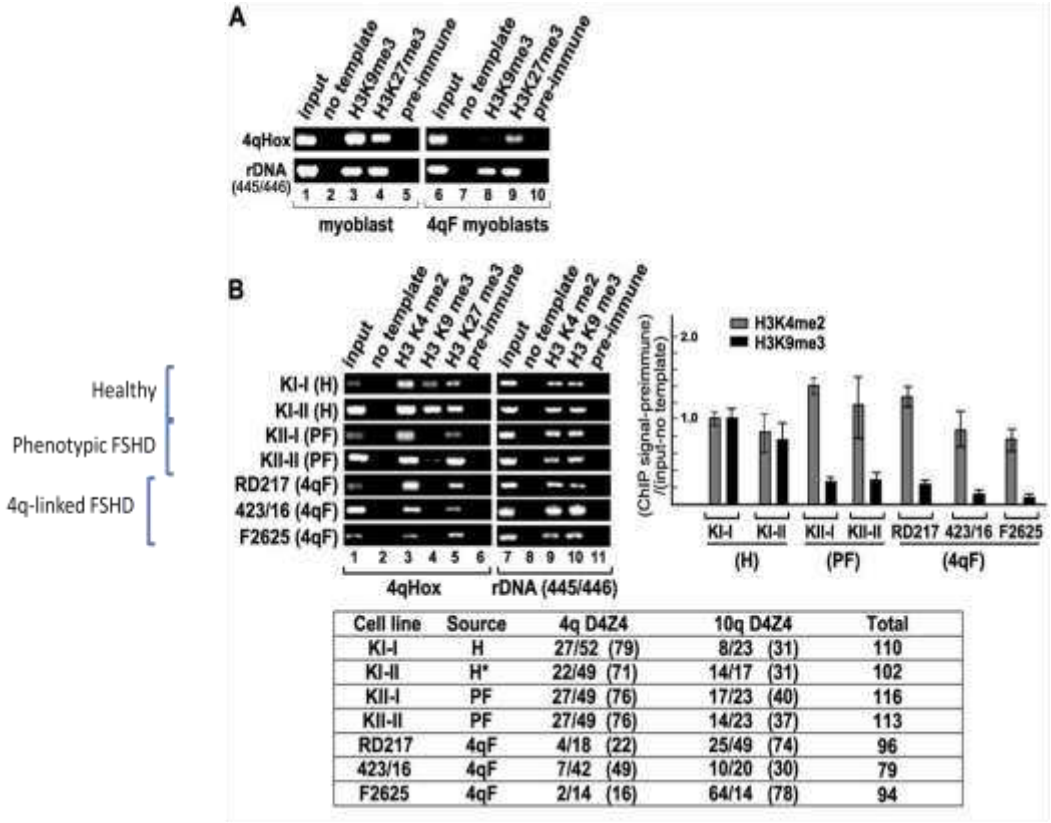
In a recent study, Zeng et al. employed ChIP analysis to examine the D4Z4 chromatin structure at the DUX4 promoter and a region within the DUX4 ORF (2009). ChIP analysis on the non-FSHD lymphoblasts revealed the presence of H3K9me3, H3K27me3, H3K4me2 and H3Ac histone modifications, but the absence of H3K4me3 at these regions. The same study revealed the presence of H3K4me2 and H3K9me3 at the D4Z4 site in human embryonic stem (hES) cell, suggesting that D4Z4 chromatin domains are marked by these histone modifications early on in development and are maintained throughout

differentiation. This was in contrast to H3Ac which was absent in hES and added on at a later stage (Zeng et al. 2009). Additionally, an independent study reported that the levels of H3K27me3 at D4Z4 are significantly reduced after myogenic differentiation (Bodega et al. 2009).

Zeng et al. (2009) reported a significant decrease in H3K9me3 signals at D4Z4 in both FSHD1 (D4Z4 contraction) and FSHD2 (no contraction) patient's myoblasts, fibroblast and lymphoblasts when compared to normal cells from healthy individuals (figure 1.7, Zeng et al. 2009). Using a 4q/10q sequence variant at the D4Z4 region under study, it was verified that D4Z4 on both 4q and 10q undergo the same H3K9me3 modification (Zeng et al. 2009). None of the other forms of muscular dystrophies tested in this study, namely Duchenne muscular dystrophy (DMD) and limb-girdle muscular dystrophy (LGMD), showed any significant loss of H3K9me3 at D4Z4. Hence, it was concluded that the loss of H3K9me3 at 4q and 10q at D4Z4 repeats appear to be a change uniquely associated with FSHD patients (Zeng et al. 2009).

DNA methylation is shown to be the downstream consequence of H3K9 methylation (Lachner et al. 2002), but in some instances DNA methylation was shown to promote trimethylation of H3K9 (Sarraf et al. 2004). Hence the loss of H3K9me3 may have been the downstream effect of the hypomethylation of the D4Z4 repeats in FSHD patients. However, H3K9me3 was largely intact in ICF patient cell lines, which, as previously reported, carry severe hypomethylation at the D4Z4 repeats on 4q and 10q sites (van Overveld et al. 2003; de Greef et al. 2009; Zeng et al. 2009). This suggested that the loss of H3K9me3 at the D4Z4 repeats is a change distinct from the DNA hypomethylation (Zeng et al. 2009). These findings imply that the loss of

H3K9me3 at D4Z4 repeats, rather than DNA hypomethylation, is casually involved in FSHD (Zeng et al. 2009).



Cell line	Source	4q D4Z4	10q D4Z4	Total
KI-I	H	27/52 (79)	8/23 (31)	110
KI-II	H*	22/49 (71)	14/17 (31)	102
KII-I	PF	27/49 (76)	17/23 (40)	116
KII-II	PF	27/49 (76)	14/23 (37)	113
RD217	4qF	4/18 (22)	25/49 (74)	96
423/16	4qF	7/42 (49)	10/20 (30)	79
F2625	4qF	2/14 (16)	64/14 (78)	94

Figure 1.7 H3K9me3 is lost in both 4q-linked and phenotypic FSHD.
(A) ChIP analysis of H3K9me3 and H3K27me3 at D4Z4 in normal and FSHD (4qF) myoblasts. 4qHOX region lies within the DUX4 ORF and rDNA region serves as a positive control. (B) H3K9me3 is specifically lost in fibroblasts. Endpoint PCR analysis with 4qHOX primers is shown on an agarose gel and quantitation of real-time PCR with Q-PCR primers, which lie upstream of DUX4 ORF, are shown as a bar chart. D4Z4 repeat numbers for 4q and 10q and the total repeat number for each cell line is indicated in the Table. (H) Non-FSHD cell line (Pfaffl et al. 2004) phenotypic FSHD cell line (4qF) 4q-linked FSHD cell line. Image adopted from Zeng et al. (2009).

1.4 Chromatin Immunoprecipitation (ChIP) assay

was first developed in 1988 for cultured Drosophila cells (Solomon et al. 1988). ChIP is now the most widely used technique for mapping the in vivo distribution of proteins associated with chromosomal DNA such as histones, transcription factors, regulatory proteins etc (Alexiadis et al. 2007; Bodega et al. 2009; Zeng et al. 2009). ChIP can be used to identify regions that are associated with these proteins, or conversely, it can be used to identify proteins that are associated with a particular region of the genome.

ChIP methodology often involves protein-DNA and protein-protein crosslinking, which consists of treating cells or tissues with formaldehyde. The chromatin is extracted and is fragmented prior to immunoprecipitation with an antibody specific to a target protein associated with the chromosomal DNA. Reversal of the protein-DNA crosslink prepares the DNA fragments for identification. The ChIP procedure is summarised in Figure 1.8. There is variety of identification procedures that can be employed including PCR, DNA microarray and DNA sequencing. Standard or quantitative PCR can be preformed to verify whether a particular region of the genome is associated with the protein of interest (Garrick et al. 2008; Zeng et al. 2009). The combination of ChIP and microarray (ChIP-CHIP) allows genome-wide identification of binding sites for the chromatin-associated proteins (De Gobbi et al. 2007; Pauler et al. 2009). Alternatively, high throughput sequencing of the immunoprecipitated DNA (ChIP-Seq) offers a more precise way of mapping protein-DNA interactions across the genome (Pauler et al. 2009).

The main drawback of ChIP its inherent variability (Mukhopadhyay et al. 2008). ChIP is a multi-step procedure and hence some level of variability is to be expected. The variability in the final results may arise from variability in the crosslinking step, immunoprecipitation or the protein-DNA washing stage. A number of carefully chosen controls are needed to account for the variability in these steps. Other limitations of ChIP analysis include the dependence on the availability of a highly specific antibody as well as broad expression of the DNA-binding protein of interest (Mukhopadhyay et al. 2008).

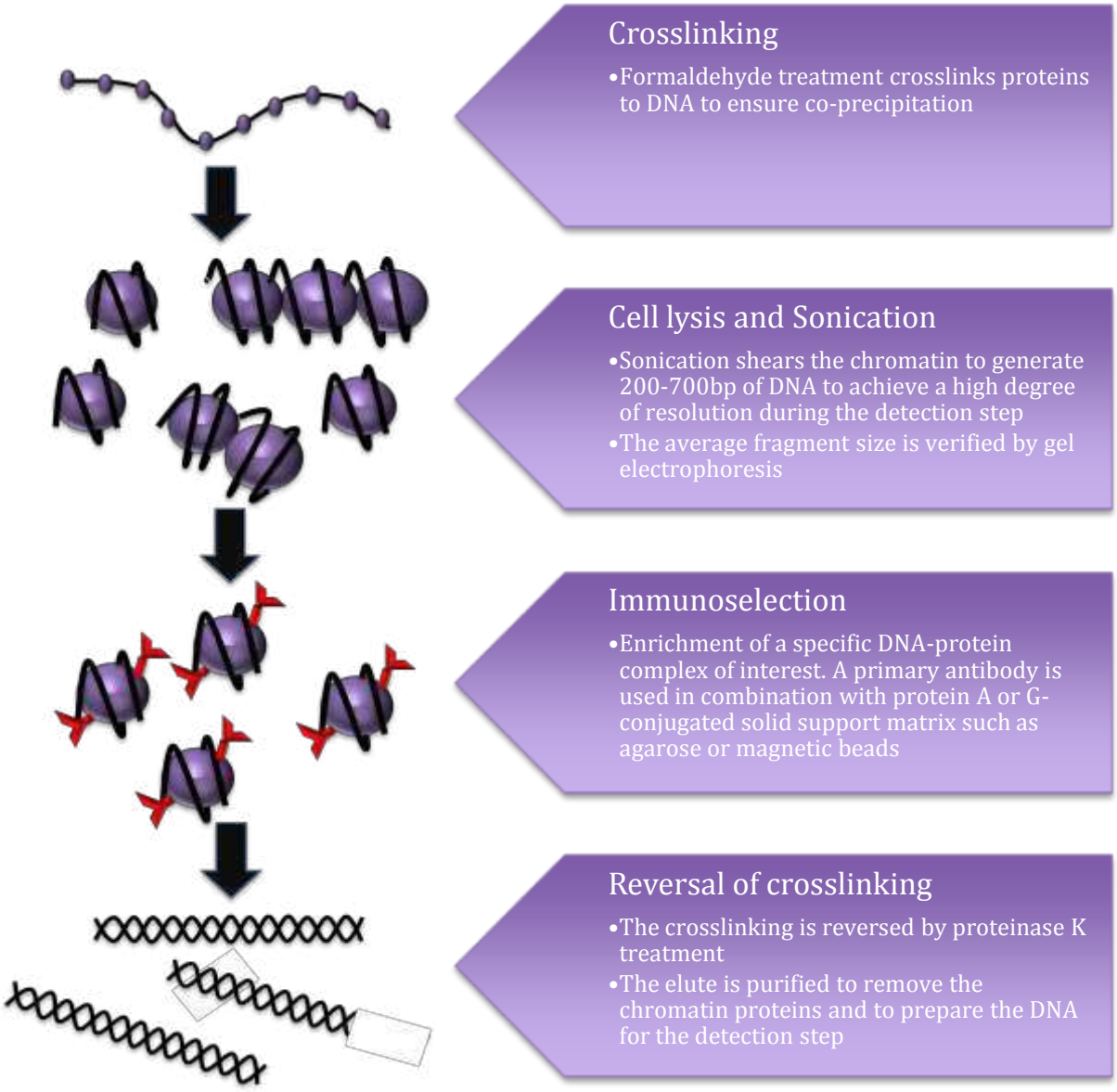


Figure 1.8 a summary of chromatin immunoprecipitation (ChIP) assay

As mentioned above, quantitative real-time PCR (qPCR) can be employed as the method of detection to analyse the ChIP data. qPCR allows close analysis of the histone composition at specific regions of interest within the genome. Section 1.5 explains the basis of qPCR methodology:

1.5 Quantitative real-time PCR (qPCR)

Quantitative real-time PCR (qPCR) is a method for DNA amplification in which fluorescent dyes that interact with the DNA are used to detect the amount of PCR product amplified after each PCR cycle (Higuchi et al 1992).

Threshold cycle (Ct) is the metric used for analysing qPCR results. The Ct value represents the number of cycles needed to reach a set threshold fluorescent signal level. Lower number of cycles needed to cross the threshold, indicates the greater the amount of starting material and vice versa (Figure 1.9). Threshold level should be chosen in a way so that it captures data during the exponential phase and is the same for all the samples analysed in a qPCR run. The threshold value can be chosen either manually or using an instrument-specific algorithm.

In this study, the qPCR assays were carried out on Corbett's Rotor-Gene 6000 and the results were analysed using Rotor-Gene 6000 Software 7.1. The threshold value was calculated using the instrument-specific algorithm, which was carried out by the software after each run.

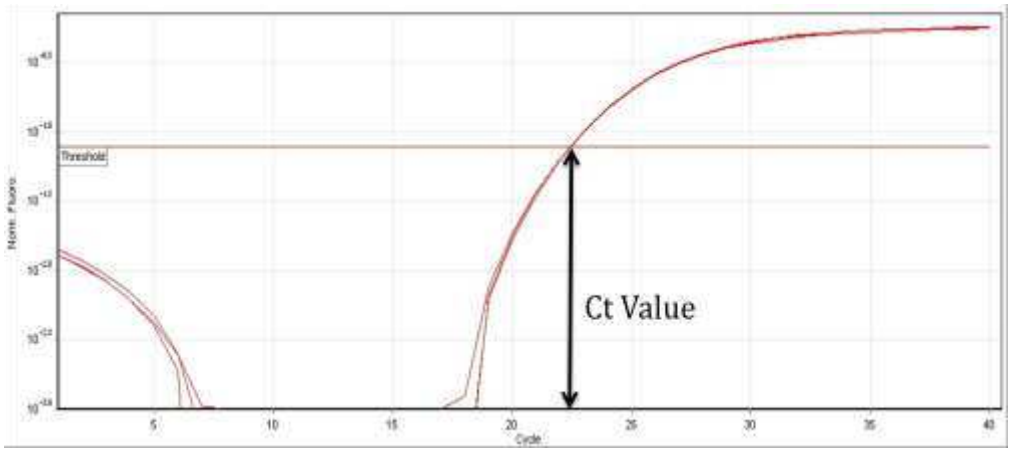


Figure 1.9 An example of a typical qPCR run.

The threshold level is set above background fluorescence and at the exponential level of amplification. The number of cycles required to reach the threshold is determined as Ct value.

1.5.1 Methods of detection:

1.5.1.1 Target-specific detection

This method involves the use of an oligonucleotide probe that has a complementary sequence to the target region. This probe is constructed to contain a reporter fluorescent dye (fluorophore) at the 5’ end and a quencher dye at the 3’end. Fluorophores absorb light energy at one wavelength and almost immediately re-emit light at another, higher wavelength. Quenchers are molecules that can accept energy from fluorophores and then dissipate it without light emission. As long as the probe is intact the proximity of the quencher absorbs the light emitted by the fluorophore by Fluorescence Resonance Energy Transfer (FRET). The specific fluorophore and quencher must be matched to each other and to the optic of the qPCR instrument. The probes can be labeled with two distinguished reporter dyes, which allow

amplification of two different target sequences in one reaction tube. TaqMan probes (also called hydrolysis probes or 5'nuclease oligoprobes) are the most commonly used target-specific detection method (Figure 1.10). Other target-specific detection chemistries include hairpin probes, LightUp probes and hybridization probes.

Although this method offers outstanding specificity and reduction in background and false positive signals for qPCR assays, it requires synthesis of a different probe for each target sequence, which can be costly. In addition prior knowledge of the DNA sequence of the target region is essential for designing these probes.

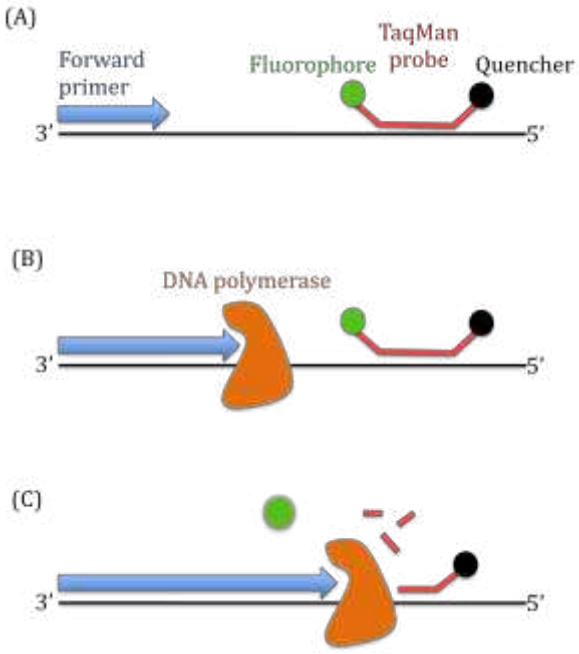


Figure 1.10 Schematic diagram of a qPCR reaction using TaqMan probe. The probe binds to the target sequence downstream of one of the primers (A). The probe is cleaved by the 5' nuclease activity of the Taq polymerase as the primer is extended (c). The cleavage of the probe allows continuous amplification of the target sequence as well as removing the fluorophore from the close proximity of the quencher, which results in increase in reporter signals (C). After each cycle, additional fluorophores are cleaved from their respective probes, resulting in increased fluorescent intensity proportionate to the amount of amplicon produced.

1.5.1.2 Non-specific detection

This approach uses dsDNA-binding fluorophores for non-specific detection of the target DNA sequence. Fluorophore used in non-specific qPCR assays exhibit low fluorescence when unbound in solution, but starts to fluoresce brightly when associated with dsDNA and exposed to a suitable wavelength of light. SYBR green is the most commonly used fluorophore (Figure 1.11). This approach was chosen in this study for its simplicity and lower cost. The advantage of this method is that it requires less optimization and less information about the target sequence. The disadvantage of this approach is that only the primers determine the sensitivity and specificity of the qPCR assay as the dsDNA-binding fluorophore binds to any dsDNA, even unwanted primer dimmer products. For this reason qPCR assays using SYBR green as the method of detection, generally include a post-PCR melting curve analysis to detect the possible presence of primer dimers or any other unwanted products. During the melting curve analysis the qPCR instrument is set to gradually increase the temperature and monitor the fluorescence. A sharp drop in fluorescence is expected when the temperature is high enough to denature dsDNA due to the release of the fluorophore (Figure 1.12). The melting temperature (Tm) is calculated by the instrument's software from the melting curve data by plotting negative first derivative (change in fluorescent signals) versus temperature (-dF/dT). The Tm of a DNA fragment is dependent on its length, G+C content, sequence, strand complementarity, concentration and on buffer component. The difference between the melting temperatures allows the detection of primer dimers or other nonspecific amplification products, which could potentially prevent accurate quantification (Figure 1.12B).

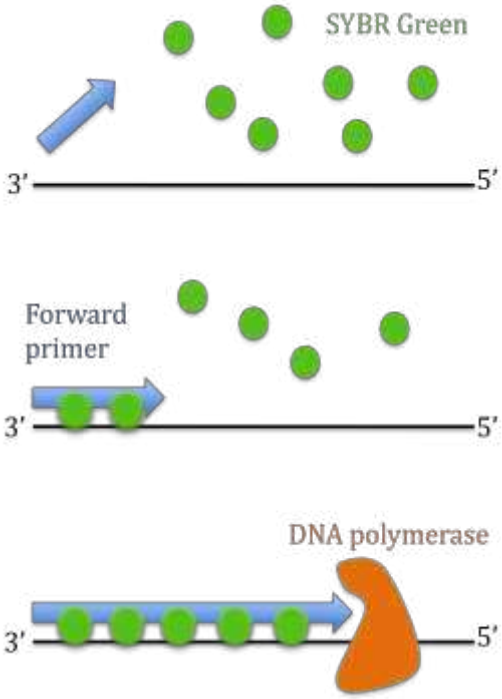


Figure 1.11 Schematic diagram of a qPCR reaction using SYBR green.
SYBR green exhibits low fluorescence when unbound in solution. Upon association with dsDNA, SYBR green starts to fluoresce brightly, which indicates product synthesis.

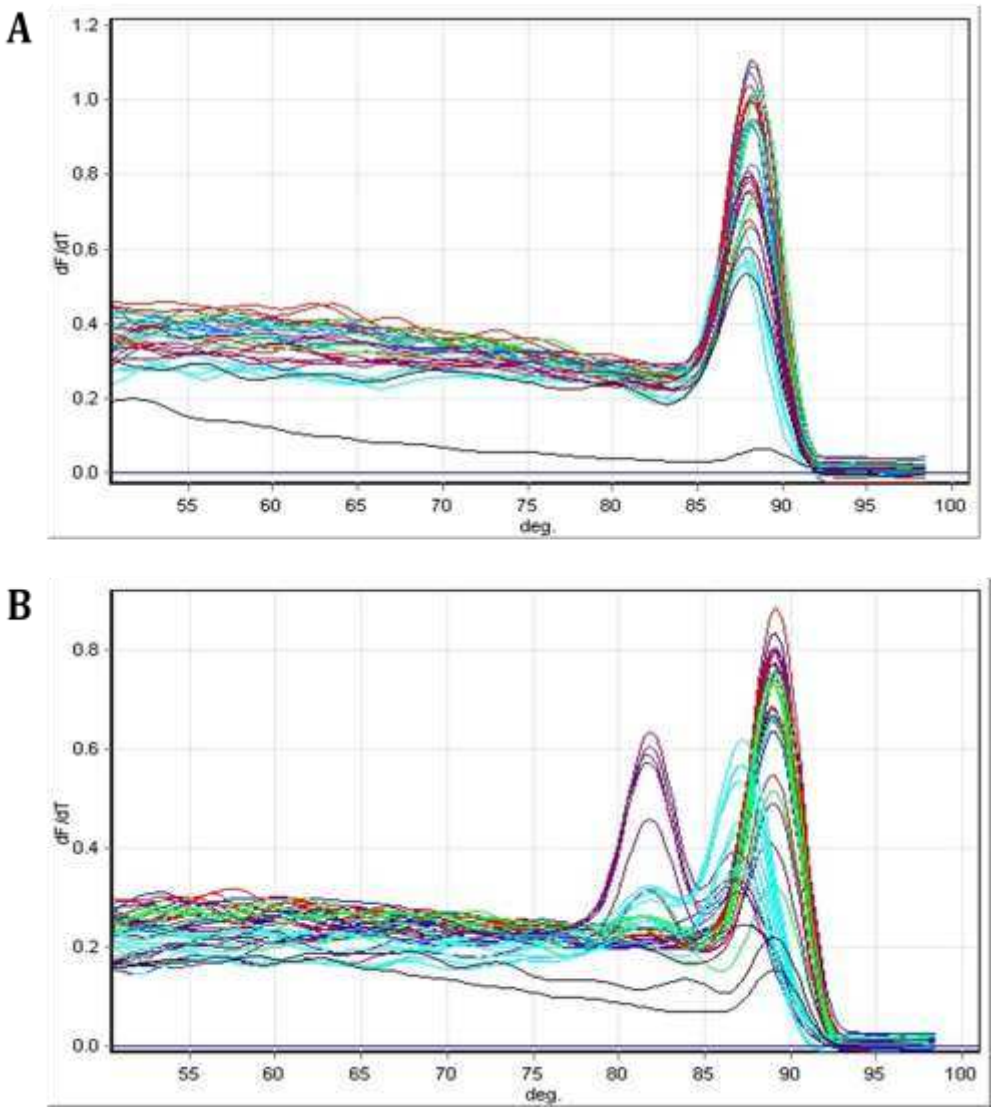


Figure 1.12 Example of a specific and a non-specific melting curve.

The derivative melting peak is the highest at the melting temperature. (A) A specific melting curve with no primer dimer or unwanted products. (B) A melting curve showing three peaks at different melting temperatures indicating significant primer-dimer formation or other non-specific amplification during this qPCR assay.

2. Aims

The focus of this work will be on further investigation of the association of previously identified chromatin structure changes involved in FSHD pathogenesis. In 2009 Zeng et al. reported the reduction of H3K9me3 histone modification at a specific region within the D4Z4 repeats in FSHD patient cell lines. The loss of H3K9me3 at this site was exclusive to FSHD patients and independent of DNA hypomethylation at this region. No significant loss of H3K9me3 was observed in other repeat sequences analysed (Zeng et al. 2009). The first aim of this study will be to repeat and validate the data from Zeng et al. (2009). ChIP analysis will be carried out on a number of patient and control lymphoblastoid cell lines, using antibodies against four different histone modifications, including H3K9me3. The ChIP outcome will be analysed by qPCR using SYBR green as the method of detection and D4Z4-specific primers used by Zeng et al. (2009). Further to the previous findings, other regions within the D4Z4 array will be analysed for any significant change in histone composition.

3. Materials and Methods

3.1 Growth and maintenance of cells

3.1.1 Growth and maintenance of patient and control lymphoblastoid cell lines

Lymphoblastoid cell lines grow in suspension with cells clumped in loose aggregates. These aggregates can be dissociated by gentle trituration with a pipette. All lymphoblastoid lines were grown in 20ml appropriate media (see section 3.1.5) in T25 tissue culture flasks at upright position. The cultures were seeded at a concentration no lower than 200,000 viable cells/ml. To maintain the growth of the cells and avoid the plateau growth level the number of cells were kept below 500,000 viable cells/ml. The cells were kept in an incubator at 37˚C with 5% CO_2. For the full list of the lymphoblastoid cell lines used in this study refer to Table 3.1.

3.1.2 Growth and maintenance of somatic hybrid cells

Several human/rodent somatic cell hybrids were used in this study. These cells were grown as attached cells in 12ml appropriate media (section 3.1.5) in 75cm^2 tissue culture flasks. For passaging cells, 3 to 5 minutes of incubation with 1 x trypsin was performed to detach the cells from the flasks. For the full list of the somatic cell hybrids used in this study refer to Table 3.2.

ID	Disease Status	Additional Information	Array size	Sequence haplotype
GM08729	Unaffected male	Father of an ICF patient		4qB163, 10qA162, CA8CA8CT7+
GM16351	FSHD male	Clinically affected brother of GM16348		
GM16352	Unaffected female	Clinically unaffected sister of GM16348		4qB163, CA9CA8-
GM16348	FSHD female			4qA161, 4qB168, 10qA166 Hc, 10qA166Hb
GM17939	FSHD male		4q: 3/33, 10q: 15/26	4qA161, 4qB163, 10qA166c (2 repeat haplotypes)
GM18207	FSHD male		4q: 3/27, 10q: 18/39	4qA161 (3 distinct repeat haplotypes), 10qA166, CA9CA7+
GM18476	Unaffected female		4q: 25/34, 10q: 8.6/12	4qB163, 10qB161Ta

Table 3.1 Information on the lymphoblastoid cell lines used in this study.

All the cell lines were obtained from Coriell Cell Repositories. Disease status is indicated for all cell lines. The D4Z4 array size is indicated for three the cell lines. The array size is unknown for the other cells. The known haplotypes for each cell line have also been included in this Table. GM16351 has not yet been haplotyped.

ID	Background	Contains human chromosome	Sequence haplotype
4L10	Mouse	#4qB	
CY7	Mouse	#10q, #16, #12	
GM10115 (HHW416)	Chinese hamster	#4 (96% of cells)	
GM10479A	Mouse	#14 (96% of cells)	
GM11418A	Chinese hamster	#15 (80% of cells)	
GM11687	Mouse	#4 (100% of cells)	4qB163
GM11688	Mouse	#10 (92% of cells)	10qB166

Table 3.2 Information on the somatic cell hybrids used in this study.

3.1.3 Long-term storage of cells

The cells were pelleted by centrifugation at 600 x g for 3 minutes. The old media was removed and the cell pellets were resuspended in 1ml of freezing media (appropriate growth media for the cells containing 10% DMSO). The cells were then transferred to one or two cryotubes placed in a freezing rack, wrapped in bubble wrap and transferred to -80˚C for at least 48 hours before storage in liquid nitrogen.

3.1.4 Revival of cells from frozen stock

Appropriate media was warmed to 37˚C and the frozen cells were thawed in room temperature. 1ml of appropriate growth media was added to the cells before transferring them to a flask of media. The flask was then transferred to the incubator and the media replaced the following day.

3.1.5 Media used for growth and maintenance of the cell lines

Cell line	Growth Media
Lymphoblastoid lines	RPMI-1640
	2mM L-glutamine
	15% fetal bovine serum (FBS)
	120U/ml Penicillin
	120µg/ml Streptomycin
Somatic cell hybrids:	
HHW416, 4L10, GM14193	DMEM
	2mM L-glutamine
	15% FBS
	120U/ml Penicillin
	120µg/ml Streptomycin
	0.2mM Proline
GM10479A, GM11687, GM11688	HAM's F12/DMEM, 1:1 mixture
	2mM L-glutamine
	15% FBS
	120U/ml Penicillin
	120µg/ml Streptomycin
	0.25mg/ml Geneticin

CY7	HAM's F12
	2mM L-glutamine
	15% FBS
	120U/ml Penicillin
	120µg/ml Streptomycin
	0.2mM Proline

3.2.1 DNA extraction from cultured cell lines

The salting out procedure (Miller et al. 1988) was used to extract DNA from the cultured cells.

The cells were pelleted by centrifugation at 1000xg for 3 min followed by removal of the culture media. The cell pellet was washed with 10ml PBS (Phosphate Buffer Saline) before it was resuspended in 3ml of nuclei lysis buffer (10mM Tris-HCl, 400mM NaCl, 2mM EDTA, pH 8.2). 0.2ml of 10% SDS and 0.5ml of protease K solution (2mg/ml Protease K in 1% SDS and 2mM EDTA) were added to the cell lysates and then incubated overnight at 37˚C. After the digestion was complete, 1ml of saturated NaCl solution (~6M) was added followed by 15 seconds of vigorous shaking before centrifugation at 2500rpm for 15 min. The supernatant containing the DNA was transferred to a new tube. Exactly two volumes of absolute ethanol was added and the tube was inverted several times. The precipitated DNA strands were removed to 100-200µl of TE buffer (10mM Tris-HCl, 0.2mM EDTA, pH 7.5). The DNA was allowed to dissolve for 2 hours at 37˚C then stored at 4˚C.

3.2.2 Phenol/chloroform extraction of DNA

An equal volume of phenol/chloroform (1:1, v:v) was added to the DNA sample. The sample was then centrifuged in a microfuge at 16000 x g for 5 min. The aqueous phase was removed to a new tube and an equal volume of chloroform was added. The sample was then centrifuged at 1600 x g for 5 min. The upper aqueous phase was removed to a new tube after the centrifugation. 1µl glycogen (20mg/ml) and 10% 3M NaAc, pH 5.5 was added prior to the addition of two volumes of absolute ethanol. After incubation at -20˚C for 30 minutes, the sample was centrifuged at 16000 x g for 10 minutes. The supernatant was discarded and the pellet was washed with 70% ethanol. The pellet was left to air dry then dissolved in 10µl of distilled water.

3.3 Electrophoresis

Agarose gels (1%, 2%, 3% or 4% w:v) were prepared using agarose powder and 1x TAE buffer. Ethidium bromide (0.5µg/ml) was added to the gel to enable visualization of the DNA with UV light. Electrophoresis was carried out in 1% TAE buffer at 5V/cm until the required resolution was achieved.

3.4 Haplotyping of the cell lines

PCR was carried out on the genomic DNA of the lymphoblastoid cells to amplify the SSLP or p13-E11 regions. For the primer pairs and the conditions used for each PCR refer to Table 3.3.

3.5 Subcloning

Subcloning of the SSLP PCR products was carried out using the CloneSmart Blunt Cloning Kit (Cambridge Bioscience #40856-1):

3.5.1 ligation

The PCR products were purified using a GeneElute PCR Clean-Up Kit (#NA1020) and eluted in 30µl of distilled water. The PCR products were phosphorylated by treatment with 1µl T4 polynucleotide kinase (PNK) (Promega) and 5µl of T4 ligase buffer in total volume of 50µl. End-labeling the PCR product with a phosphate group is necessary to enable ligation into the vector. The PNK treatment was carried out at 37˚C for 30min followed by incubation at 65˚C for 20min to deactivate the enzyme. Phosphorylated fragments were then purified using the GeneElute PCR Clean-Up Kit in order to remove PNK from the solution. The presence of the PCR products in the elute was verified by electrophoresis at this stage.

100-200ng of blunt and phosphorylated DNA was ligated into pSMART LCKan vector by following the CloneSmart Blunt Cloning Kit protocol (Lucigen Corporation). Each reaction was incubated at room temperature for 2 hours. The ligase was inactivated by 15min incubation at 70˚C.

3.5.2 Transformation

E.Cloni 10G chemically competent cells (Lucigen Corporation) were used for transformation. 1µl of ligations was added to the cells followed by incubation on ice for 30 minutes. The cells were then incubated at 42˚C for 45 seconds and placed back on ice for a further 2 minutes. 960µl of room temperature Recovery Medium was added to the cells. The cells were then incubated at 37˚C for 1 hour in a shaking incubator at 250rpm. 200µl of cells was plated on a nutrient agar plate containing kanomycin (30µg/ml) and incubated over night at 37˚C.

3.6 Selection of positive clones

3.6.1 Colony PCR

Colony PCR was carried out to select for the clones carrying the correct insert. A small amount of a colony was added to 12µl of sterile distilled water with 12.5µl of BIOMIX Red (Bioline) and appropriate primers (2µM). Standard PCR conditions consisted of 30 cycles of denaturation at 96˚C for 15sec, annealing at different temperatures depending on the primer pair (refer to Table 3.3) for 30sec, elongation at 72˚C for 30sec/Kb. The clones carrying the correct sized insert were selected after analysis by electrophoresis.

3.6.2 DNA purification - Plasmid DNA miniprep

The positive clones were grown over night in growth media containing kanomycine. The Gene Elute plasmid Miniprep kit (Sigma) was used. The manufacturer's protocol was followed except that the DNA was eluted from the purification columns in 50µl sterile distilled water.

3.6.3 Restriction digestion

The plasmid DNA minipreps were further verified as positive clones by EcoR1 digestion. DNA was incubated with 10µl of EcoR1 solution containing 1µl EcoR1 and 2µl buffer H (Promega #R6011) at 37˚C for 2 hours. The clones containing the right sized fragments were sequenced.

3.7 Sequencing

50ng of DNA was used per sequencing reaction. Big dye (diluted 1:12 with 5 x sequencing buffer) and appropriate primers (10pM) were used to amplify cloned region. The cycling conditions were as follows: 5min at 95˚C, followed

by 45 cycles of 95˚C for 30sec, 50˚C for 20sec and 60˚C for 4min. The products were purified using CleanSeq and sent to Durham sequencing. Editing of sequence traces was carried out using Sequencher 4.9.

3.8 Chromatin Immunoprecipitation (ChIP)

ChIP procedure was carried out using EZ-Magna ChIP™ A (Millipore). All the reagents used were provided by the kit unless otherwise stated.

3.8.1 Crosslinking of chromatin

Patients and control lymphocytes were cultured and counted using a haemocytometer. Approximately 1 x 10^7 cells were used per reaction. Formaldehyde solution (38%) (Sigma 05-690) was added to a final concentration of 0.4% and the sample was incubated at room temperature for 10 min on a rotator. 1M glycine was added to a final concentration of 125mM to quench unreacted formaldehyde and stop over-crosslinking. The sample was inverted to mix and incubated at room temperature for a further 5 min. The sample was kept on ice from this point on. After centrifugation at 800 x g for 5 min at 4˚C, the supernatant was removed and the pellet was resuspended in 10ml of chilled PBS and Protease inhibitor Cocktail II solution (96% 1x PBS, 4% Protease inhibitor Cocktail II). This wash was then repeated once. The crosslinked cell pellets were snap frozen in liquid nitrogen and stored at -80˚C.

Primer name	Sequence 5'-3'	Product size (bp)	PCR Conditions						
			Denaturation		Annealing		Elongation		No. of cycles
			Temp. (˚C)	Time	Temp. (˚C)	Time	Temp. (˚C)	Time	
D4F104S F1	CCCAGTTACTGTTCTGGGTGA	500	98	10sec	60	5sec	72	30sec	30
D4F104S R1	GAAAGCCCCCTGTGGGAG								
SSLP F1	GGTGGAGTTCTGGTTTCAGC	160	95	15sec	59	15sec	72	15sec	30
SSLP R	CCTGTGCTTCAGAGGCATTTG								

Table 3.3 Primers used for SSLP and p13-E11 amplification by PCR.

The sequence of each primer in addition to the PCR conditions used for amplification is also shown in this Table. D4F104S primer pair was used for p13-E11 amplification. SSLP primer pair was used to amplify the SSLP region.

3.8.2 Preparation for sonication

The crosslinked cell pellet was resuspended in 500µl of Cell Lysis Buffer (CLB) containing 1% Protease inhibitor Cocktail II. This was followed by incubation on ice for 15 min, vortexing every 5 min. The sample was then centrifuged at 800xg for 5 min at 4˚C. The supernatant was removed and the pellet was resuspended in 500µl of Nuclear Lysis Buffer (NLB) containing 1% Protease inhibitor Cocktail II.

3.8.3 Optimisation of the sonication conditions

Different power settings of the sonicators in addition to different duration and number of pulses were tested. 10% of the samples were removed prior to the sonication for analysis of the unsheared DNA. During the sonication 10% aliquots were removed after a set number of cycles. After the sonication the sample volumes were made up to 50µl with ChIP elution buffer and 0.5mg of RNase A (Sigma #R6148) was added to each sample. This was followed by 30 minutes incubation at 37˚C. 20µg Proteinase K (Qiagen) was added to each sample and incubated over night at 62˚C to denature the crosslinked proteins. The DNA fragments were purified by phenol/chloroform extraction (section 3.2.2) and analysed on a 2% gel.

3.8.3.1 Sonication using Diagenode Biorupture™ 200

The cross-linked sample in NLB was split equally between two 1.5ml eppendorf tubes as 250µl aliquots. Each sample was sonicated at high power for 14 cycles of sonication (30 sec on/30 sec off). The ice was changed every 5 cycles to avoid heating of the sample. After sonication, the sample was centrifuged at 12000 x g for 10 min at 4˚C. 10% of the supernatant was

removed to assess sonication efficiency by performing standard phenol/chloroform extraction (section 3.2.2) and electrophoresis. The remainder of the supernatant was divided into 50µl aliquots, each containing approximately $1x10^6$ cells, which is sufficient for one immunoprecipitation. Aliquots were either used immediately or were snap frozen in liquid nitrogen and stored at -80˚C.

3.8.3.2 Sonication using Covaris™ S-series

The sample in NLB was transferred to a Covaris™ TC13 glass tube-13mm x 65mm (#520025). The sample was sonicated for 20 to 50 cycles (30 seconds on/30 second off) with the intensity ranging from 3 to 8. Duty cycle was set to 20% with 200 cycles per burst. After the sonication the sample was transferred to a 1.5ml eppendorf tube and centrifuged at 12000 x g for 10 min at 4˚C. 10% of the supernatant was removed to analyse the sonication efficiency and the rest of the supernatant was aliquoted and stored as described above.

3.8.4 Immunoprecipitation (IP)

For each immunoprecipitation reaction 50µl of sonicated material was combined with 450µl of Dilution Buffer (DB) containing 2.25µl of Protease Inhibitor Cocktail II. 5µl was removed from one of the IPs for input control, which represents the total amount of chromatin in the sonicated samples before IP. An appropriate amount of antibody (described in section 3.8.6) and 20µl of fully re-suspended protein A magnetic beads were added to each IP. The IPs were incubated overnight at 4˚C on a rotating wheel.

Protein A magnetic beads were pelleted using a magnetic separator and the supernatant was discarded. The beads were then washed consecutively with

500µl of Low Salt Immune Complex Wash Buffer, High Salt Immune Complex Wash Buffer, LiCl Immune Complex wash buffer and TE Buffer. The samples were incubated for 5 min at 4˚C on a rotating wheel for each wash.

3.8.5 Reversal of crosslinking and DNA purification

100µl of ChIP Elution Buffer (warmed to room temperature to ensure the SDS was dissolved) containing Proteinase K (10µg/µl final concentration) was added to the washed and precipitated Protein A magnetic beads as well as the input control. The samples were then incubated for 2 hours at 62˚C in a rotating incubator. After incubation, the samples were incubated for a further 10 minutes at 95˚C before cooling down to room temperature. The magnetic separator was used to pellet the beads and the supernatant was transferred to a new tube. 500µl of Bind Reagent "A" was added to each 100µl of the supernatant. The sample was then transferred to a spin filter in collection tube and centrifuged for 30 seconds at 13500 x g. The flow-through in the collection tube was discarded and 500µl of Wash Reagent "B" was added to the spin filter in the collection tube. The sample was centrifuged at 13500 x g for 30seconds. The flow-through was discarded and the spin filter column was centrifuged for a further 30 seconds at 13500 x g. The spin filter was then transferred to a clean collection tube and 50µl of Elution Buffer "C" was added directly to the centre of the spin filter membrane. The column was centrifuged for 30 seconds at 13500 x g. The purified DNA was then either used immediately for qPCR analysis or stored at -20˚C.

3.8.6 Antibodies used for ChIP

Control antibodies, normal Rabbit IgG and H3Ac antibody were supplied by the EZ-Magna ChIP A kit and 5µg of each was used for each IP. Other antibodies used are listed in Table 3.4.

Antibody	Source	Cat no.	Lot no.	Amount used per IP (µg)
H3K4me2	Millipore	07-030	DAM1570816	4
H3K9me3	Millipore	17-625	DAM1595626 DAM1682222	4
H3K9me3	Diagenode	pAb-013-050	001	2.4 or 4.8
H3K27me3	Millipore	17-622	DAM1647850	4

Table 3.4 Test antibodies used in this study.

3.9 Detection by quantitative real-time PCR (qPCR)

Real-time PCR was used to quantify enrichment levels for different histone modifications using the purified DNA obtained by ChIP. All qPCR experiments in this study were carried out on a Corbett Rotor-Gene 6000 (on standard settings) and the results were analysed using Rotor-Gene 6000 Software 7.1. The threshold value was calculated using the instrument-specific algorithm, which was carried out by the software after each run. All qPCRs were run in triplicates (duplicates at primer optimization stage) and the mean value from the triplicates was used in calculations and to create the standard curves after each run. SYBR green was used as the method of detection in this study. Maxima™ SYBR Green/ROX qPCR Master Mix (2X) (Fermentas Life

Sciences) was used for all qPCR experiments. The typical qPCR conditions used were recommended by Fermentas for "Two-step cycling protocol":

95˚C	10 minutes	
95˚C	15 seconds	} 40x
60˚C	60 seconds	

qPCR reactions were carried out in a total volume of 25µl, containing 12µl SYBR green master mix and 5µl of 10-fold diluted elute from ChIP purification columns. The appropriate amounts of forward and reverse primers were added for efficient and specific qPCR reaction (discussed in section 3.9.1).

3.9.1 Calculating the correct primer concentration for qPCR

A primer concentration test was performed for each primer pair used in this study (see Table 3.5 for the full list of primers). The aim of this test was to identify a final primer concentration whereby the qPCR was the most efficient but still specific to the target sequence. The efficiency of the qPCR at different primer concentrations was determined by the slope of the standard curve produced from a set of DNA serial dilutions at different primer concentration (Section 3.9.2). For some primer pairs, the chromosomal specificity was tested

by using genomic DNA from the somatic cell hybrids carrying a single human chromosome.

Primers were tested at four final concentrations: 0.08pM, 0.4pM, 0.8pM and 1.2pM. Each primer concentration was tested on four 10-fold serial dilutions of 5ng/µl genomic DNA extracted from lymphocytes according to the salting out procedure (Section 3.2.1).

3.9.2 Creation of standard curve

For each qPCR run and for each primer set, a standard curve was produced. Standard curves were created by serial dilution (10-fold) of genomic DNA. Serial dilutions covered the range of 5ng/µl to 5x10^3 ng/µl which covered the data range obtained from the ChIP samples. Mean Ct value from these serial dilutions were plotted against the DNA concentration of each sample (Figure 3.1).

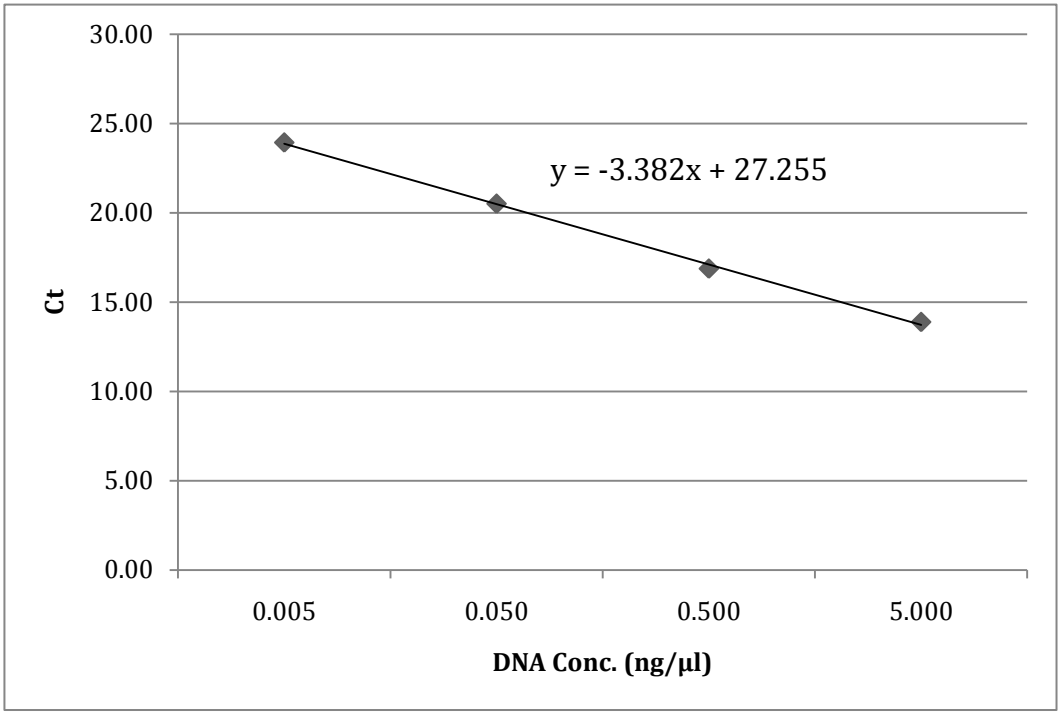


Figure 3.1 An example of a standard curve using the Q-PCR primers.

The slope in this graph is -3.38, which represents a qPCR run with less than 100% efficiency.

Plotting the standard curve shows the consistency of the relationship between the DNA concentration and the Ct value, it also determines the efficiency, the sensitivity and the reproducibility of the qPCR assay. The value of the slope of the line (x) defines the efficiency of the reaction. A value of -3.32 for the slope represents a theoretically 100% efficient qPCR as this is the number of cycles required for a 10-fold increase in product. A more negative value than -3.32 represents a qPCR reaction which is less efficient than 100%, whereas a value less negative than -3.32 indicates a reaction operating at a greater than the maximum efficiency, which may suggests the presence of more DNA in the standard curve serial dilution than thought.

Primer name	Sequence 5'-3'	Product size (bp)
4qA161 F	AGACCCTTGTCATGCCATTT	196
4qA161 R	GGGCAGGTTGGGAGAC	
4qA161 F2	CTTCTGGAGACCCTTGTCAT	191
4qA161 R2	TGGGAGACCCCCTCTGCCGT	
4qB163 F	AGACCCTTGTCATGCCATTA	195
4qB163 R	GGGCAGGTTGGGAGAT	
alphaSYBR F	GGGTGCGGGCTGACTTT	65
alphaSYBR R	CAGCGCCACCCTTTCCT	
DIST F	GAGATGCTGGAGTCAGGACCAT	91
DIST R	AGGAGTCAGGAGCAGCAGTCA	
DXZ4 F	GCCTACGTCACGCAGGAAG	210
DXZ4 R	TATGTTTGGGCAGGAAGATCG	
GAPDH F	TACTAGCGGTTTTACGGGCG	166
GAPDH R	TCGAACAGGAGGAGCAGAGAGCGA	
LUC7L F	CCACGATGGCATAAGGATAATCT	75
LUC7L R	CATACTTTCCGTGCCCTTGTG	
Q-PCR F	CCGCGTCCGTCCGTGAAA	107
Q-PCR R	TCCGTCGCCGTCCTCGTC	
Q-PCR R2	GGGGTCCAAACGAGTCTCC	
Q-PCR F3	CACAGTCCGGCTGAGGTG	113
Q-PCR R3	GTGCTGTCCGAGGGTGTC	

Table 3.5 Sequences of all the primers used for qPCR analysis.

The qPCR conditions used were the same for all the primer pairs (see section 3.9).

3.10 Data analysis and normalization of ChIP-qPCR data

3.10.1 Percent input method

This method represents the ChIP-qPCR results of the IPs as a percentage of the total amount of chromatin that was present before the immunoprecipitation step.

Firstly, the ΔCt between the IP samples and the input sample was calculated. The input Ct value was corrected before the subtraction because the input sample represented one hundredth of the total chromatin in the IP samples before immunoselection. A 10-fold increase in product is represented as -3.32 cycles when the reaction has 100% efficiency. Hence a 100-fold increase in products must be represented by -6.64 cycles.

Corrected input Ct value = $Ct_{(input)} - 6.64$

Since majority of the qPCR reactions carried out in this study showed less than 100% efficiency, a specific correction value was calculated for each primer pair (Section 3.3).

dCt = $Ct_{(IP)}$ – corrected input Ct value

Then, the %input value was calculated separately for each of the 3 replicates of an IP sample.

%input = $2^{(dCt)}$

The %input value of the triplicates was averaged. The data was then corrected for any background signals indicated in the no template control (NTC).

Av %input – Av %input NTC

The standard deviation between the triplicate %input values for each IP was calculated using Microsoft Excel software. The standard error was then calculated from the standard deviation.

$SE = SDEV/\sqrt{(3)}$

This value indicated the standard error of the average taken from the triplicates for each IP.

3.10.2 Normalization to a reference gene - GAPDH

All the qPCR data obtained from different primers were normalized to GAPDH ChIP-qPCR results. In the concept of ChIP-qPCR data, this method represents the histone enrichment in the target regions relative to the histone enrichment at the GAPDH region. As GAPDH is a housekeeping gene and therefore should be expressed in all cell lines Hence, theoretically, the chromatin arrangements at this region should be euchromatic and similar in all the cells. This method enables the direct comparison of the histone enrichment levels at different regions of the genome in a cell line or a number of different cell lines. GAPDH normalisation also corrects for the experimental variation between separate ChIP experiments.

4. Results

4.1 Haplotyping of BN cell line

The BN lymphoblastoid cell line was provided by Dr. Cristina Tufarelli. This cell line was previously used as a control by Dr. Tufarellie's research group and thought to potentially be a suitable control cell line for FSHD. To determine whether the BN cell line was a suitable FSHD control the haplotype of this cell line at the 4q35 allele had to be determined. Ideally, an FSHD control cell line should contain the disease "permissive" haplotype, 4qA161, on at least one of the 4q alleles. In order to determine the haplotype of the BN cell line at the FSHD locus, the SSLP and p13-E11 regions were amplified and sequenced.

The SSLP region was amplified by PCR and cloned in a sequencing vector and 11 positive clones were sequenced. The p13-E11 PCR products were sequenced directly. The sequencing results were analysed and haplotyped according to Lemmers et al (2010). Only four haplotypes are expected within each cell line (two from the 4q and 2 from the 10q allele). Due to the identical sequence variants between certain haplotypes at the SSLP and the p13-E11 region, some haplotypes cannot be distinguished using the sequencing data from these two regions alone (see Table 4.1). Hence, the sequencing data for the BN cell line indicated the presence of 2 possible haplotypes for the 10q allele: 10qA166 (the sequencing did not reach far enough to distinguish between a, b and the c variants of 10qA166 haplotype) or 10qA166Ha and 4 possible haplotypes for the 4q: 4qA166, 4qA166Ha, 4qA166Hb or 4qC166H.

Nevertheless, it was indicated that the BN cell line does not have a 4qA161 allele.

	SSLP								D4F104S1 (p13-E11)																		
HAPLOTYPE	CA1	A	CA2	G/T	CT1	CT2	CT3	8bp	C G	T C	C T	A G	C T	T C	A G	T G	T A	G A	T C	T G	A T	A C	G C	CHROMOSOME	DISTAL A/B
4A161	10	A	10	G	5	1	6	x	C	T	C	A	C	T	A	T	T	G	T	T	T	A	G	4	A
4A166	9	A	8	T	5	1	7	✓	G	C	T	G	T	C	G	G	A	A	T	G	T	A	G	4	A
4A166Ha	9	A	8	T	5	1	7	✓	G	C	T	G	T	C	G	G	A	A	T	G	T	A	G	4	A
4A166Hb	9	A	8	T	5	1	7	✓	G	C	T	G	T	C	G	G	A	A	T	G	T	A	G	4	A
4C166H	9	A	8	T	5	1	7	✓	G	C	T	G	T	C	G	G	A	A	T	G	T	A	G	4	C
10A166Ha	9	A	8	T	5	1	7	✓	G	C	T	G	T	C	G	G	A	A	T	G	T	A	G	10	A
10A166a	9	A	8	T	5	1	7	✓	G	C	T	G	T	C	G	G	A	A	T	G	T	A	G	10	A
10A166b	9	A	8	T	5	1	7	✓	G	C	T	G	T	C	G	G	A	A	C	G	T	A	G	10	A
10A166c	9	A	8	T	5	1	7	✓	G	C	T	G	T	C	G	G	A	A	T	G	T	C	G	10	A
Position in AF117653	1551	1571	1573	1595	1597	1619	1623	1644	4409	4499	4519	4530	4534	4576	4592	4643	4666	4674	4717	4754	4761	4790	4796		

Table 4.1 The sequence variants within the SSLP and the p13-E11 regions. Only the haplotypes discussed in section 4.1 are included in this table. The disease-permissive allele, 4qA161, is shown in the first row and the possible 4q and 10q haplotypes for the BN cell line are shown in the remaining rows. The table was adapted from Joanne Pollington.

In order to ensure that no preferential amplification of one allele had occurred a PstI digestion was carried out on the p13-E11 PCR products from the BN cell line, along with a previously haplotyped lymphoblastoid cell line (GM16348) used as positive control for the digestion. There is an A/G variant within the p13-E11 region that distinguishes between the 10qA166/4qA166 and 4qA161/4qB163 haplotypes (see Table 4.1 position 4592). PstI digestion of 10qA166 and 4qA166 alleles produces 4 fragments (266bp, 183bp, 39bp and 35bp). Whereas, Pst1 digestion of 4qA161/4qB163 alleles results in only two fragments (488bp and 35bp).

The results of the PstI digestion of p13-E11 PCR products from the BN cell line and the control (GM16348) is shown in Figure 4.1. The control displayed 3 bands representing the 488bp fragments corresponding to 4qA161/4qB163 and the 266bp and 183bp fragments corresponding to 10qA166/4qA166. The 35bp and 39bp fragments were below the resolution of the agarose gel electrophoresis under the conditions used. This was in agreement with the known haplotypes assigned to GM16348. The digestion results from the BN cell line only displayed the 266bp and 183bp fragments corresponding to 10qA166/4qA166. This supports the conclusion form the SSLP data that the BN cell line does not carry the 4qA161 haplotype. Hence there was a concern that the BN cell line was not a suitable control for FSHD and therefore other control cell lines were obtained.

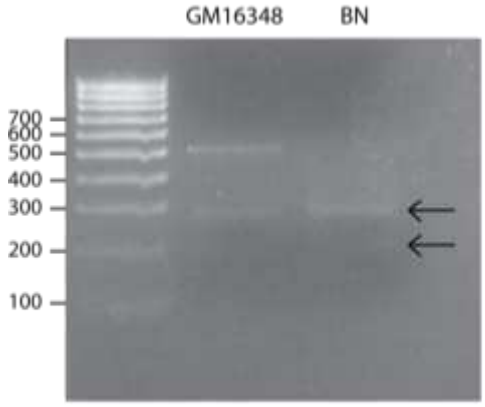


Figure 4.1 PstI digestion of the p13-E11 region.

PstI digestion was done on the PCR products from a FSHD lymphoblastoid cell line carrying the permissive 4qA161 haplotype (GM16348) and the PCR product from the BN cell line. Three bands at 488bp, 266bp and 183bp were produced from digestion of the p13-E11 of the GM16348. The digestion of the p13-E11 region of the BN cell lines resulted in two bands at 266bp and 183bp (arrows). Absence of the band at 488bp indicated that this cell line does not carry a 4qA161 or 4qB163 allele.

4.2 qPCR optimisation

As explained in section 1.5.1, the primer specificity to the region of interest is crucial when using SYBR green as the method of detection for qPCR analysis. In this section, the optimisation of the qPCRs used in the subsequent ChIP assay is described. The aim of the qPCR optimisation was to determine conditions that enable efficient, sensitive and the reproducible PCRs.

4.2.1 Primer concentration optimisation

In addition to the primer sequence, which is the most important factor in primer specificity, the final concentration of the primers in the PCR reaction is also important. Having excessive amount of primers in the reaction increases

the level of non-specific binding by the primers. Therefore, each primer pair used in this study was first tested in a primer concentration titration assay to determine the lowest final primer concentration without compromising the qPCR efficiency. The qPCR efficiency was measured by the quality of the standard curve obtained (section 3.9.2).

The titration assay consisted of testing four different primer concentrations against three 10-fold dilutions of the template DNA. The four final primer concentrations used were 80nM, 0.4µM, 0.8µM and 1.2µM. A melt curve was obtained after each qPCR run to check for primer dimers. The 80nM concentration proved to be inefficient with low reproducibility for most primers (Figure 4.2). However, 0.4µM, 0.8µM and 1.2µM final primer concentrations usually gave very similar results to each other. Some primer pairs, especially GAPDH, did not work very well at 0.4µM (Figure 4.2A). In addition, 0.4µM final concentration gave inconsistent results for other primer pairs such as Q-PCR (Figure 4.2C, D).

Therefore, 0.8µM final concentration was chosen for all the primer pairs used in this study, as this final concentration proved to be the lowest concentration that produced efficient and reproducible PCR reactions for all primer pairs. One standard final concentration was chosen for all primer pairs because choosing different concentration for each primer pair introduces an additional variable step in the ChIP-qPCR assay. In addition, changing the primer concentration may mask the biological differences between the samples when comparing the ChIP-qPCR results of different regions in the genome.

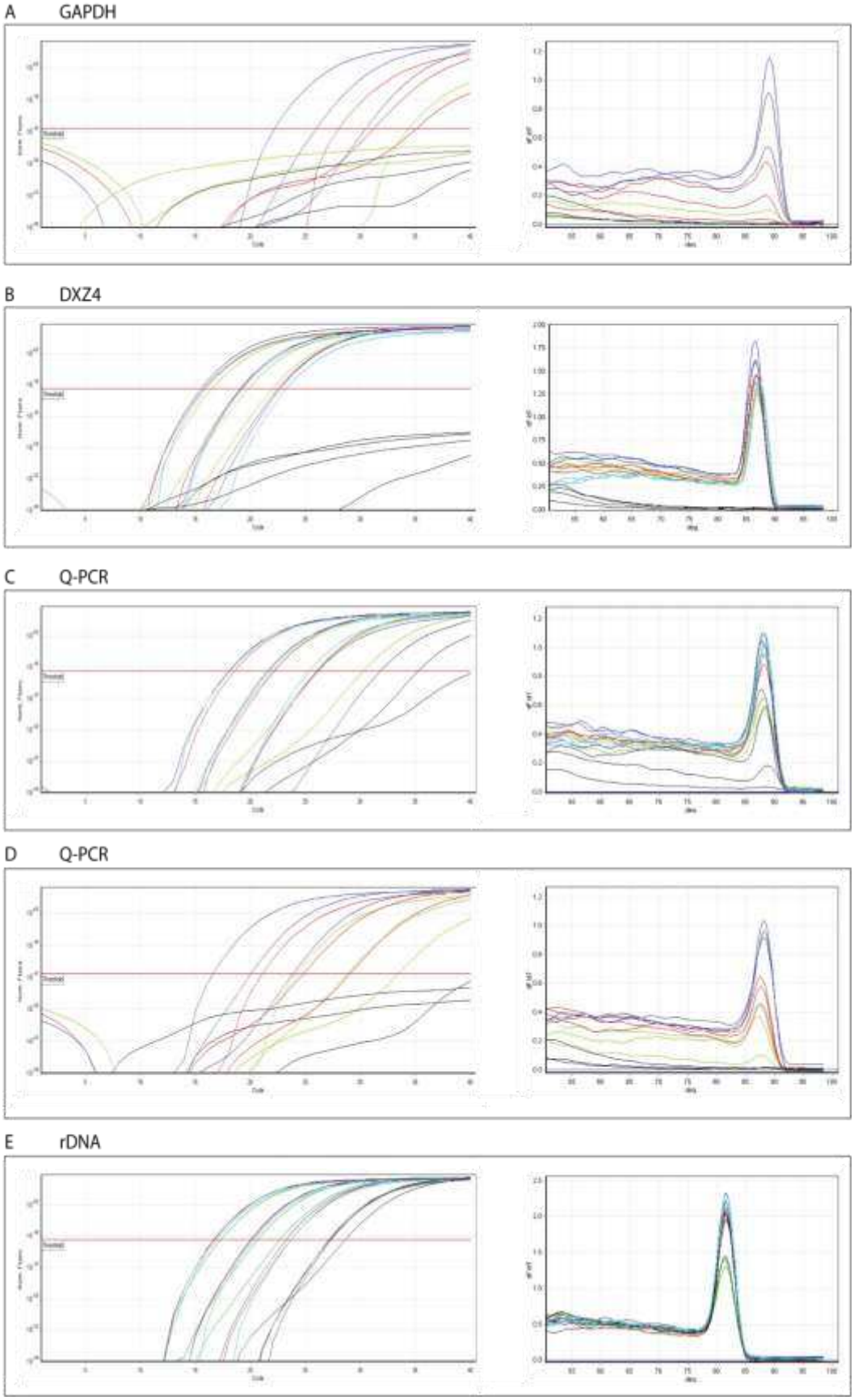


Figure 4.2 Primer concentration titration assay.

Legend on next page

Data from qPCRs with four different primer concentrations on three 10-fold dilutions of the template DNA. The four final concentration used were 80nM (green), 0.4µM (red), 0.8µM (blue) and 1.2µM (not shown on these graphs). The diagram shows the qPCR run and the melt curve for four different primer pairs. Black lines represent the no template control at the three different primer concentrations. No primer dimers formation was observed on any of the melt curves. (A) Primer concentration titration test on GAPDH primer pair. GAPDH primer pair did not work efficiently at 80nM or 0.4µM. (B) Primer concentration titration test on DXZ4 primer pair. DXZ4 primer pair produced an efficient qPCR at all three primer concentrations. (C, D) Primer concentration titration test on Q-PCR primer pair. (C) Q-PCR primer pair did not work efficiently at 80nM. Both the 0.4µM and 0.8µM primer concentrations produced efficient and very similar results. (D) Results obtained for Q-PCR primer pair at 0.4µM were not consistent as they showed much less efficiency compared to the first test (C). (E) Primer concentration titration test on rDNA primer pair. rDNA primer pair produced an efficient qPCR at all three primer concentrations. However, amplification in no template control (black) was observed every time these primers were used and so they were not used for the remainder of this project.

4.2.2 Primer specificity

The specificity of some of the primer pairs to their target sequences was tested by using somatic cell hybrids containing only one or a limited number of human chromosomes. From these experiments it was concluded that the Q-PCR and 4qA161 primers used by Zeng et al (2009) were not as specific as suggested by these authors.

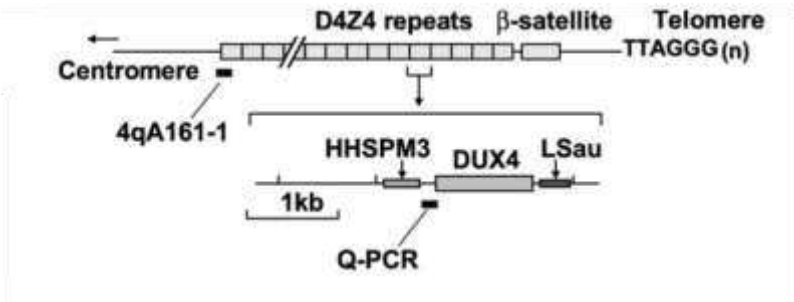


Figure 4.3 A schematic diagram of the Q-PCR and the 4qA161 amplification regions.

Diagram shows the 4q allele and a single D4Z4 repeat. PCR products for Q-PCR and 4qA161 primer pairs are shown with black bars. Figure modified from Zeng et al (2009).

4.2.2.1 Q-PCR primer pair

The Q-PCR primer pair sequences were obtained from Zeng et al. (2009). This primer pair amplifies a 107bp region within the D4Z4 repeats that overlaps with the start of the DUX4 ORF (Figure 4.3). Zeng et al. (2009) reported that this primer pair are specific to the D4Z4 repeats on chromosomes 4q and 10q. They showed this using (end state) PCR on DNA from a number of mono-

chromosomal somatic cell hybrids (2009). In the current study, the specificity of the Q-PCR primer pair to 4q and 10q was tested using qPCR and genomic DNA from a FSHD lymphoblastoid cell line (GM18207) as positive control and five different somatic cell hybrids containing specific human chromosome(s): HHW416 (chromosome 4), GM11688 (chromosome 4), CY7 (chromosomes 10, 12 and 16), GM10479A (chromosome 14) and GM11418A (chromosome 15). Somatic cell hybrids containing acrocentric chromosomes were used because D4Z4-like repeats are known to be present on these chromosomes (Lyle et al. 1995). GM10479A and GM11418A were used as negative controls as they contain chromosomes 14 and 15, respectively. Details of the somatic cell hybrids used in this study are provided in Table 3.2. The primers were tested on 10-fold dilutions of DNA from the selected cell lines. A melt curve was produced after each run to investigate the presence of any primer dimers or amplification of alternative products in the samples.

Using a 60˚C annealing temperature, amplification was observed in the chromosome 4 and chromosome 10 hybrids (Figure 4.4). GM18207 produced the lowest Ct values, which is to be expected as this cell line contains a full set of chromosomes and therefore many target sequences for the Q-PCR primer pair. The Ct values from HHW416, which carries only human chromosome 4, were similar to the Ct values from GM18207. However, amplification also occurred in the hybrids containing only chromosome 14 or 15, indicating non-specific amplification (Figure 4.4). Increasing the annealing temperature did not improve the specificity of this primer pair (Figure 4.4). The melt curves showed one peak for all the samples. This indicated formation of a product in the negative controls with similar size and C/G content to the product from the

positive controls. This non-specific binding was not due to excess of primers, since lowering the primer concentration to 0.4µM did not reduce amplification from the negative controls. The Ct values from the negative controls were higher than that of GM18207 and HHW416. Higher Ct value can indicate lower amount of template or target sequence. This is consistent with the Q-PCR primer pair preferential amplification from 4q and 10q repeats.

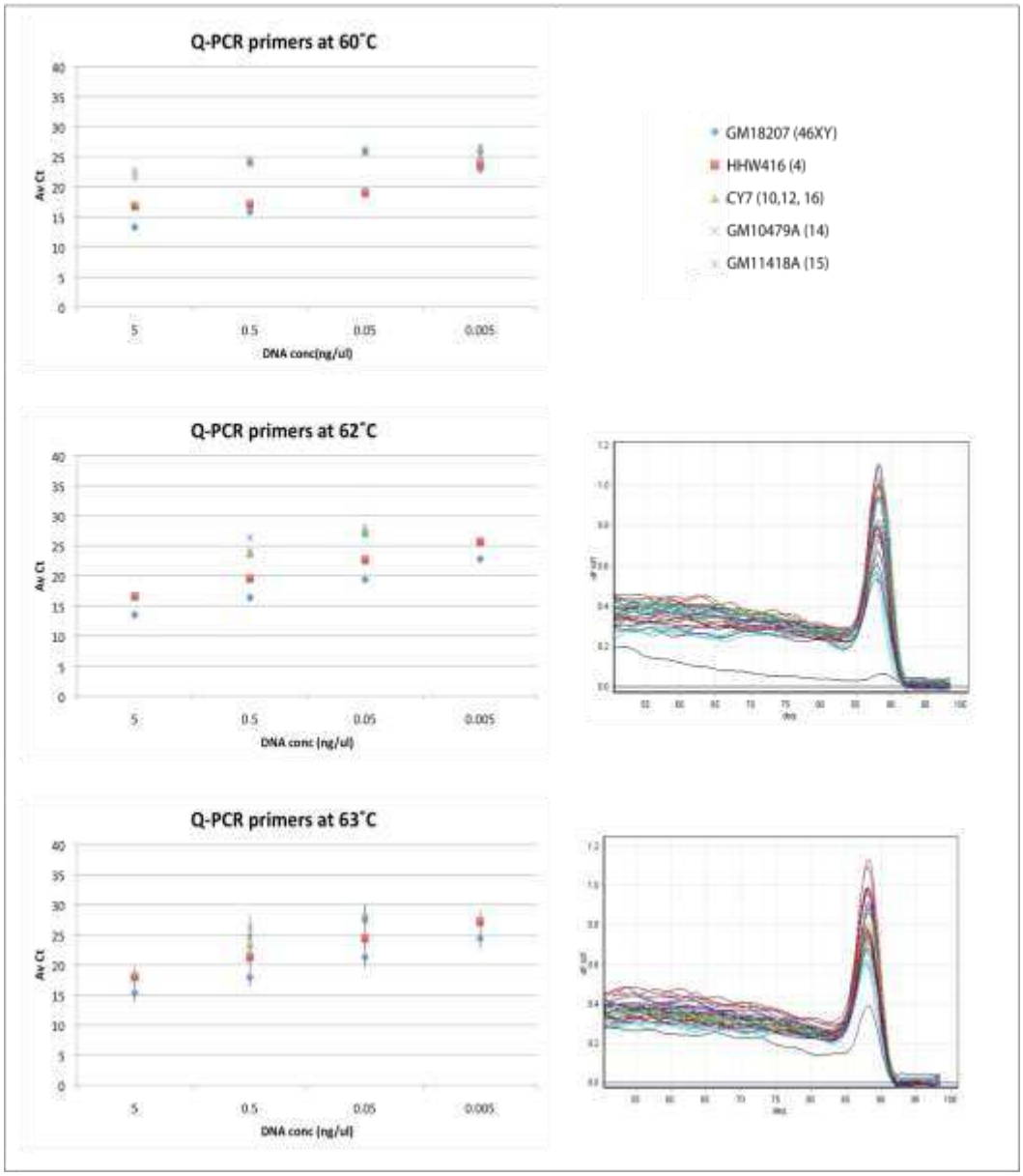


Figure 4.4 Specificity of Q-PCR primers to 4q and 10q at different annealing temperatures.

Legend on next page

Q-PCR primers were tested on four 10-fold dilutions of DNA from a human lymphoblastoid cell line (GM18207) carrying a complete set of chromosomes (46XY) and DNA from four different somatic cell hybrid. The somatic cell hybrids with human chromosome/s (in brackets) are indicated. The average Ct values were plotted against their relevant DNA concentration to produce the graphs on the left. The error bars represent the standard error of the triplicate Ct values. The melt curves obtained at 62˚C and 63˚C are also shown. Raising the annealing temperature did not improve the specificity of the Q-PCR primer pair, as amplification was still observed in all the negative controls. The Ct values of all samples increased as the annealing temperature increased, indicating less efficient amplification during these qPCR runs at higher temperatures than 60˚C. The melt curves show formation of a single peak indicating that the products from the negative controls have a similar size and C/G content to the product from the positive controls.

To investigate whether the Q-PCR primer pair show tendency to target 4q over the other sites, the qPCR products after a full qPCR run were digested with XapI restriction digest. The Q-PCR amplification region contains a C/G polymorphism that can determine whether the amplicons are from 4q (C variant) or 10q (G variant) (Zeng et al. 2009). A XapI digestion site is situated at the site of this polymorphism. Hence, XapI digestion can be used to investigate whether the majority of the Q-PCR amplification is from 4q rather than other regions. XapI digestion of the Q-PCR products from 4q would be expected to produce an 88bp and a 19bp fragment. Q-PCR amplicons from 10q and other chromosomes do not contain a XapI site and therefore no digestion is expected. XapI digestion of the qPCR products from a 46XY lymphoblastoid cell line (GM17939) and the mono-chromosomal somatic cell hybrid (HHW416) is shown in Figure 4.5. Genomic DNA from HHW416 was used as a positive control for XapI digestion, since all Q-PCR amplifications from this template are expected to be from chromosome 4q. However, the Q-PCR products from HHW416 were only partially digested (Figure 4.5), indicating the presence of non-4q amplicons. XapI digestion was negative for the sample where the genomic DNA from GM17939 (46XY) was used as template (Figure 4.5). This indicated that the majority of the qPCR products amplified by the Q-PCR primers were not from the 4q repeats.

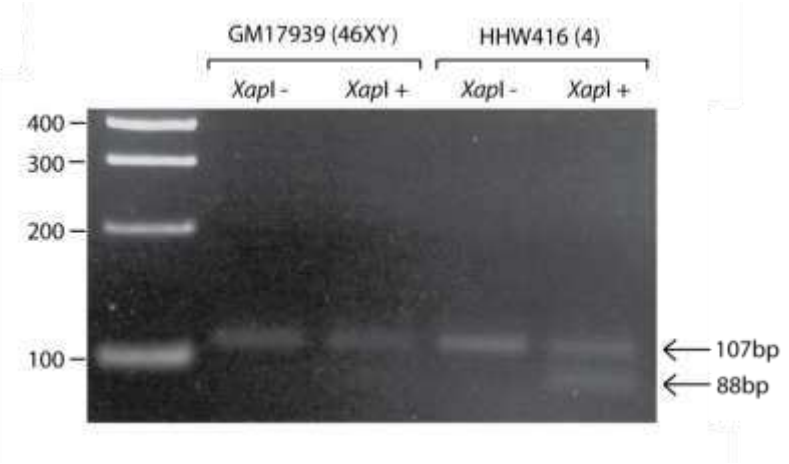


Figure 4.5 XapI digestion of qPCR products with Q-PCR primers.
qPCR was carried out on genomic DNA obtained from a patient lymphoblastoid cell line, GM17939 (46XY), and a somatic cell hybrid (HHW416) carrying human chromosome 4. XapI did not digest Q-PCR products from GM17939. Q-PCR products from HHW416 were only partially digested by XapI.

4.2.2.2 Alternative Q-PCR primers

An alternative Q-PCR reverse primer sequence was designed (Q-PCR R2) and tested for specificity to the 4q allele (Q-PCR.2 primer pair) (Figure 4.6A). High levels of amplification were observed in all the samples at both annealing temperatures tested. However, the melt curve showed formation of a different product in the mono-chromosomal somatic cell hybrids carrying chromosomes 14 and 15 to the cells carrying 4 or 10. Interestingly, the melt curve peak from the FSHD lymphoblastoid cell line (46XX) grouped with the peak observed for hybrid cell lines with chromosomes 4 and 10 (Figure 4.6A). It was concluded that this primer pair showed a tendency to amplify D4Z4 repeats on chromosome 4 and 10 over other regions. Similar results were observed when both forward and reverse primer sequences were altered (Q-PCR.3 primer pair) (Figure 4.6B).

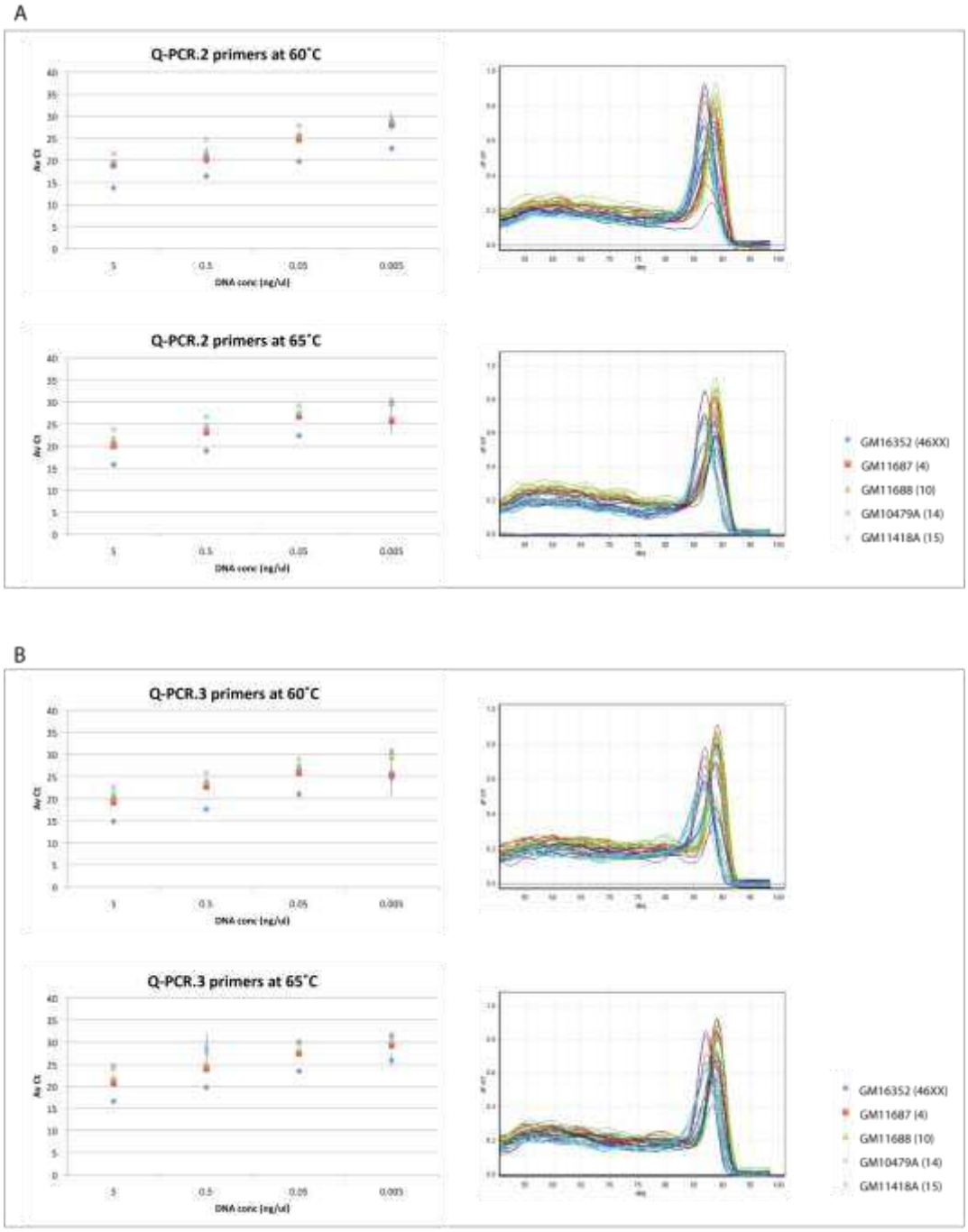


Figure 4.6 Specificity of the Q-PCR.2 and Q-PCR.3 primer pairs.

Legend on next page

The primer pairs were tested on four 10-fold dilutions of DNA from a human lymphoblastoid cell line (GM16352) carrying a complete set of chromosomes (46XX) and DNA from four different mono-chromosomal somatic cell hybrids. The consequent melt curve from each qPCR run is shown on the left. (A) Specificity of the Q-PCR.2 primer pair (Q-PCR F and Q-PCR R2). High levels of amplification were observed in all the samples at both annealing temperatures tested. However, the melt curve showed formation of a different product from the mono-chromosomal somatic cell hybrids carrying chromosomes 14 and 15 to the cells carrying 4 or 10 or the lymphoblastoid cell line (46XX). (B) Specificity of the Q-PCR.3 primer pair (Q-PCR F3 and Q-PCR R3). Similar results were observed for this primer pair as the old forward and the new reverse primer sequences.

4.2.2.3 Testing 4qA161 primer pair

Zeng et al (2009) reported a qPCR primer pair that amplified a region within the p13-E11 region (proximal to the D4Z4 repeats) and was specific for the 4qA161 haplotype (Figure 4.3). This primer pair was designed using sequence variants in the p13-E11 region that are present in the 4qA161 and not in the 4qB163 haplotype (Figure 4.7). The specificity of this primer pair to this regions was tested by performing a primer concentration titration on four 10-fold dilutions of genomic DNA from a positive and a negative control. Genomic DNA from a FSHD lymphoblastoid cell line containing the 4qA161 haplotype (GM18207) was used as positive control and the genomic DNA from a mono-chromosomal somatic cell hybrid containing human chromosome 4qB163 (HHW416) was used as negative control. A standard curve was produced after each run to investigate the efficiency of the qPCR.

The 4qA161 primer pair did not show specificity to the target region at any primer concentration as shown by formation of an alternative product in the negative control (Figure 4.8). Since the qPCR was most efficient with primer concentration of 0.8µM, and less non-specific amplification was observed at this concentration than the higher concentrations, this was the final concentration chosen for this primer pair. Increasing the annealing temperature from 60˚C to 61˚C, 62˚C and 63˚C did not increase the specificity of the 4qA161 primer pair. The specificity of these primers was also tested on DNA from three somatic cell hybrids carrying D4Z4-containing human acrocentric chromosomes. Amplification was observed in all the cell lines tested, this indicated that this primer pair was not specific to the 4qA161 locus (Figure 4.8).

Figure 4.7 Alignment of the 4qA161 primers sequences against part of the p13-E11 region from five different cell lines.

The mismatches of the primer sequences with 4qB163, 10qA166, 4qA166 and 4qB166 haplotypes are highlighted. Sequencing data provided by Joanne Pollington.

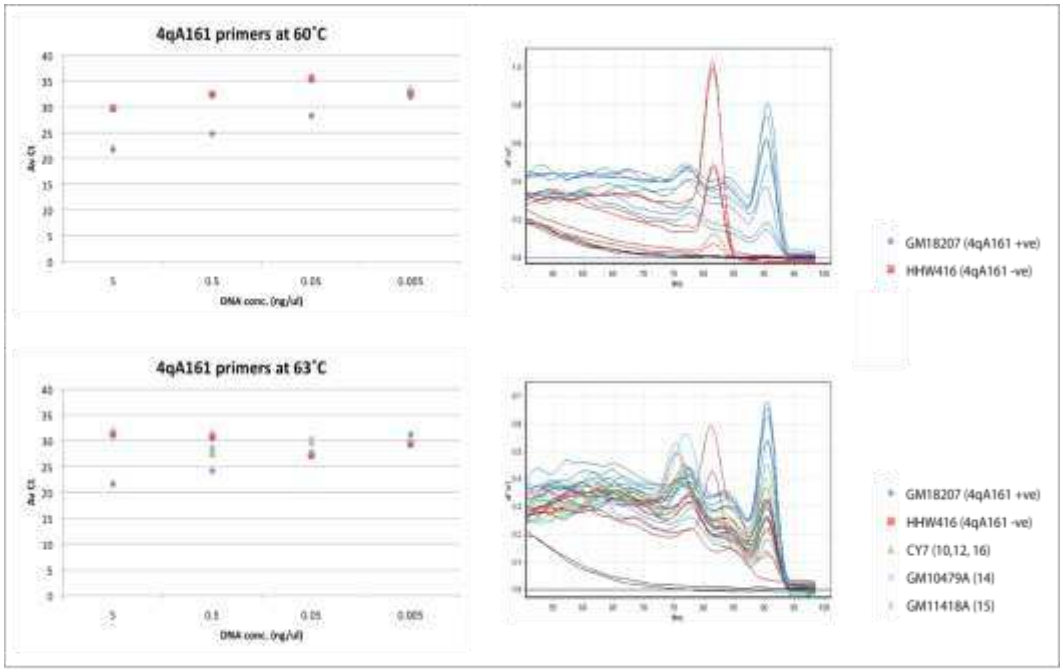


Figure 4.8 Testing the specificity of 4qA161 primers in qPCR.

(A) 4qA161 specificity at 60˚C annealing temperature. Amplification was observed in the negative control, HHW416, although with higher Ct values than the Ct values for the positive control, GM18207. The melt curve shows the formation of an alternative product in the negative control (red peak). (B) 4qA161 specificity at 63˚C annealing temperature. 4qA161 specificity was tested on five different somatic cell hybrids as negative controls. Amplification was seen in all the negative controls indicating a high level of non-specific binding by the 4qA161 primer pair.

4.2.2.4 Alternative 4qA161 primers

A new 4qA161 primer pair was designed with slightly altered sequences (4qA161.2 primer pair) (Figure 4.9). This primer pair was tested for specificity as above, using GM18207 DNA as a positive and HHW416 DNA as a negative control (Figure 4.10A). Amplification was still observed for the negative control, although with higher Ct values than the 4qA161 primer pair, was still

observed, and the melt curve also indicated formation of an alternative product in the negative control (Figure 4.10A).

Different combinations of primers for the 4qA161 region were then tested (Figure 4.10). No improvement was observed using the 4q161.3 primer pair (4qA161 F2 and 4qA161 R) (Figure 4.10B). The 4qA161.4 primer pair (4qA161 F and 4qA161 R2) also resulted in amplification in the negative control. However, the melt curve produced from this qPCR run did not show formation of the alternative product, unlike the other 4qA161 primer combinations (Figure 4.10C). Since the 4qA161 primers showed high non-specificity to the 4qA161 allele, these primer pairs were not used in the ChIP-qPCR assay.

Figure 4.9 Alignment of the 4qA161.2 primers sequences against part of the p13-E11 region from five different cell lines.

The miss-matches of the primers with non-4qA161 haplotypes have been highlighted. Sequencing data provided by Joanne Pollington.

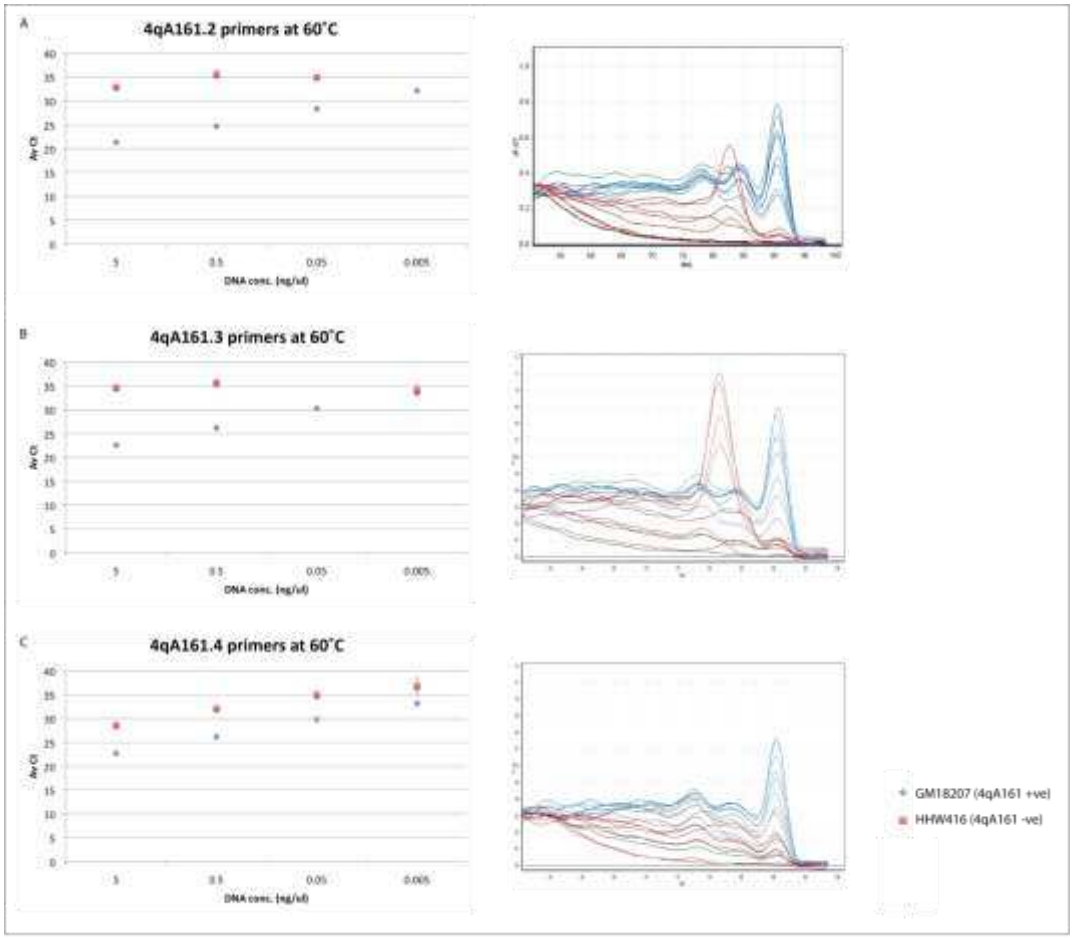


Figure 4.10 Specificity of 4qA161.2, 4qA161.3 and 4qA161.4 primer pairs.
(A) Specificity of the 4qA161.2 primer pair. Some amplification, was still observed in negative control, although with higher Ct values than the 4qA161 primer pair, was observed in the negative control. The melt curve indicated formation of an alternative product in the negative control samples with the highest DNA concentration. (B) Specificity of the 4q161.3 primer pair (4qA161 F2 and 4qA161 R). Some level of amplification was still being observed in the negative control. The melt curve indicated the formation of an alternative product in the negative control. (C) Specificity of the 4qA161.4 primer pair (4qA161 F and 4qA161 R2). A higher level of amplification was observed from the negative control. However the melt curve did not show any evidence of an alternative product formation in the negative control.

4.2.2.5 4qB163 primer pair

This primer pair was designed using sequence variants in the p13-E11 region that are present in the 4qB163 haplotype but not in 4qA161 (Figure 4.11). The specificity of this primer pair was tested as above by performing a primer concentration titration on four 10-fold dilutions of genomic DNA. DNA from a FSHD lymphoblastoid cell line with the 4qB163 haplotype (GM17939) and DNA from a mono-chromosomal somatic cell hybrid containing human chromosome 4qB163 (HHW416) were used as positive controls. Genomic DNA from a FSHD lymphoblastoid cell line with the 4qA161 haplotype (GM18207) was used as negative control for this primer pair. A standard curve was produced after each run to investigate the efficiency of the qPCR.

The 4qB163 primers were most efficient at 0.8µM final concentration (Figure 4.12A). This primer pair showed specificity to the 4qB allele as it was found to also amplify the 4qB168 haplotype. Amplification of 4qB168 haplotype was not surprising as there are no mismatches between the 4qB163 R primer and this alleles and there is only one mismatch at the 3' end of the 4qB163 F primer sequence. However, some amplification was also observed for the 4qA161 haplotype at higher DNA concentrations. To improve the specificity further the annealing temperature was raised from 60˚C to 62˚C (Figure 4.12B), which resulted in a more specific amplification (Figure 4.12B). However, increasing the temperature did compromise the qPCR efficiency, as the Ct values obtained for the positive control were higher than the Ct values obtained at 60˚C. For this reason, 60˚C annealing temperature was used for this primer pair in all the subsequent ChIP-qPCR assays. Non-specific amplification by 4qB163 primers was only observed in samples containing

5ng/µl and 0.5ng/µl of DNA. The amount of DNA recovered from ChIP was usually lower than these concentrations. In addition, keeping the annealing temperature consistent with the other qPCR analyses allowed direct comparison of the data obtained by 4qB163 primer with data obtained using the other primer pairs in this study.

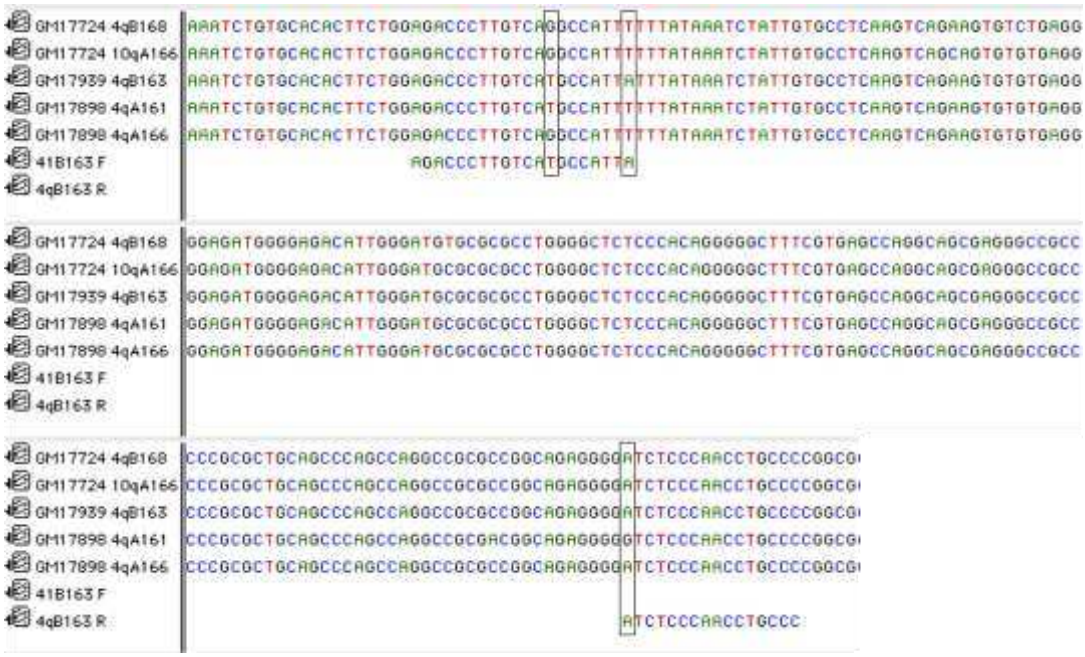


Figure 4.11 Alignment of the 4qB163 primers against part of the p13-E11 region from five different cell lines.

The mismatches of this primer pair with 4qA161, 4qB166, 4qA166 and 10qA166 haplotypes have been highlighted. Sequencing data provided by Joanne Pollington.

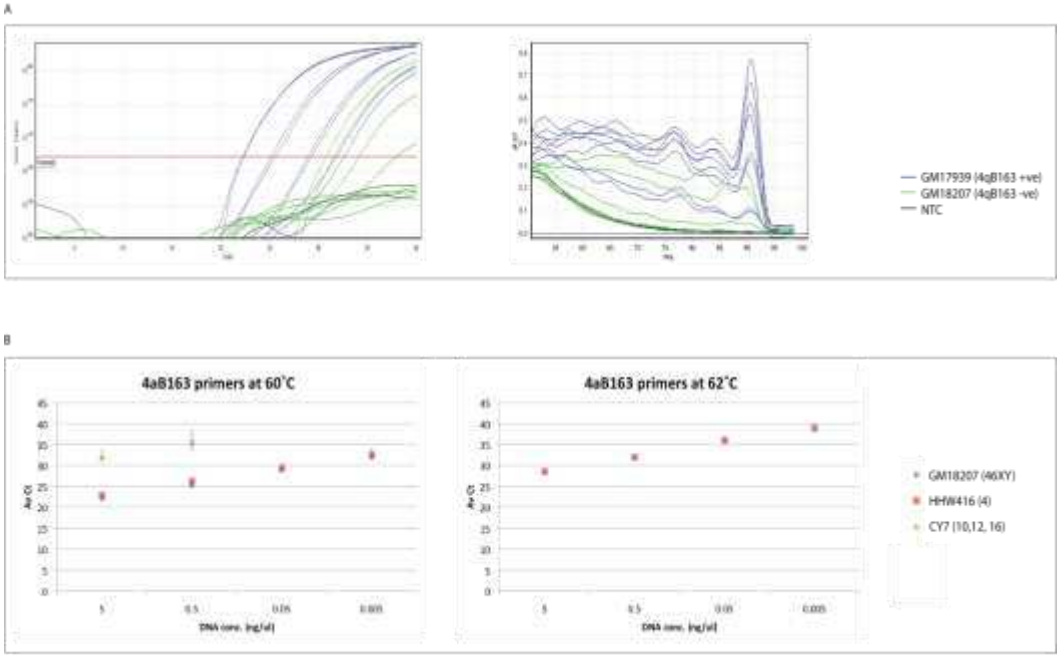


Figure 4.12 Specificity of 4qB163 primers for the 4qB163 allele.

(A) 4qB163 specificity at 0.8pM final primer concentration at 60˚C annealing temperature. The qPCR run is shown on the left. The melt curve obtained from the same run is shown on the right. Some amplification is observed in the negative control samples (green lines) with the highest DNA concentration. However, the melt curve indicates the formation of a different product in the negative control (green) to the products formed in the positive control (blue). (B) Average Ct values are plotted against their relevant DNA concentration. The error bars represent the standard error of the triplicate Ct values. Amplification is observed in the negative control samples at both 5ng/µl and 0.5ng/µl DNA with an annealing temperature of 60˚C. Raising the temperature resulted in increase in the Ct values for the positive control (GM18207 was not used in this assay). No amplification was observed in the negative control with an annealing temperature of 62˚C.

4.3 Calculating the corrected input value

Almost all the standard curves obtained from the different qPCR runs with each primer pair illustrated less than 100% efficiency (i.e. slope value was not 3.32). Since the data analysis of the ChIP-qPCR data relied on this value to correct the input Ct value (section 3.9.2 and 3.10.1), a specific input correction value was calculated for each primer pair. At least three different standard curves for each specific primer pair were compared and the slope values were averaged and then doubled to obtain the input correction value specific for that primer pair. These values are shown in Table 4.1.

Primer pair	Average slope value	SD of the average slope value	Input correction value
4qB163	3.46	0.0702	6.91
alphaSYBR	3.29	0.0624	6.58
DIST	3.31	0.168	6.63
DXZ4	3.29	0.0982	6.58
GAPDH	3.48	0.0906	6.96
LUC7L	3.48	0.309	6.95
Q-PCR	3.38	0.121	6.76

Table 4.2 Average slope value and the input correction value for each of the primer pairs that were used in ChIP-qPCR analysis.

4.4 Sonication optimization

The aim of the sonication step is to break down the chromatin into small fragments without damaging the DNA-binding proteins. Keeping the proteins intact ensures efficient co-precipitation. The size of the chromatin fragments has a direct effect on the resolution of the ChIP such that the smaller the fragments are after the sonication, the higher the resolution of the ChIP. The regions of DNA bound by nucleosomes show some protection from breakage by sonication. Therefore, the smallest fragment size obtained by sonication without damaging the nucleosomes is around 200bp. If the chromatin is broken down to 200bp fragments, when a certain region is found to be bound by a certain histone modification after the ChIP analysis, it can be said that this histone modification was most likely present in the nucleosome bound directly to this region. However, having the DNA fragments too small (i.e. one nucleosome long), may compromise amplification of the target sequence, which may stretch to the regions in between two nucleosomes. Hence, the aim was to obtain 200-700bp fragments with the majority being between 300-500bp, approximately 2-3 nucleosomes long. A minimum size of two nucleosomes ensures that most target regions stay intact. Then if a particular histone modification is found to be associated with a particular region, it can be said that this modification was present either in the nucleosome directly bound to this region, or in one of the neighboring nucleosomes which may be ~100-300bp away.

To check the level of shearing of the chromatin after sonication, the samples were analysed by agarose gel electrophoresis after de-crosslinking. The sonicated samples appear as a smear on the gel indicating a range of different

sized DNA fragments within the samples. For this study, two sonicators of different makes and models were compared for their efficiency to shear the chromatin with less damage to the DNA binding proteins:

4.4.1 Covaris™ S-series

Originally the sonicating conditions recommended by Covaris were used for chromatin shearing. Sonicator conditions were set to duty cycle of 20% with 200 bursts at intensity 8 for 30 continuous 60sec cycles. The chromatin was sheared very efficiently to 300-700bp fragments. However, there were some concerns about the increase of sample temperature that occurred during the course of sonication, which may have resulted in damaging the protein-DNA complexes. Hence, 30sec intervals between each 30sec burst were introduced. As the sample still increased in temperature, the intensity of the sonicator was reduced to 5.

The sonicator conditions were set to duty cycle of 20% with 200 bursts at intensity 5 for 40 cycles (30sec on/30sec off). Aliquots were taken out after cycles 15, 20, 25, 30, 35 and 40 for gel electrophoresis analysis (Figure 4.13B). As the desired fragment size was obtained after 30 cycles without over-heating, these conditions were used to obtain samples for the immunoprecipitation step. The result of qPCR analysis of these samples is shown in Figure 4.13D.

4.4.2 Diagenode Biorupture™ 200

The samples were sonicated at high power for 22 cycles (30sec on/ 30sec off). The ice was changed every 5 cycles in order to keep the samples cool. Aliquots were taken out after 6, 10, 14, 18 and 22 cycles for gel electrophoresis analysis (Figure 4.13A). As the desired sized fragments were obtained after 14 cycles

without over heating, these conditions were used to obtain samples for immunoprecipitation. The result of qPCR analysis of these samples is shown in Figure 4.13C.

As mentioned above, the aim of the sonication step is to break down the chromatins into 200-700bp fragments without damaging the DNA-binding proteins. The appropriate size-range was successfully obtained using both of the sonicators. However, there were much lower levels of DNA recovery after the immunoprecipitation stage from the samples sonicated with the Covaris S-series sonicator compared to the samples from Diagenode Biorupture 200 (Figure 4.13C, D). The qPCR results from the two ChIPs were comparable because the starting material (i.e. crosslinked samples) for both sonicators were from the same cell line and were crosslinked at the same time. In addition, the non-immune IgG control result is similar for both sonicated samples (Figure 4.13C, D). As a result of this comparison, the Diagenode Biorupture 200 was used to prepare the chromatin for all the subsequent ChIP analyses performed in this study.

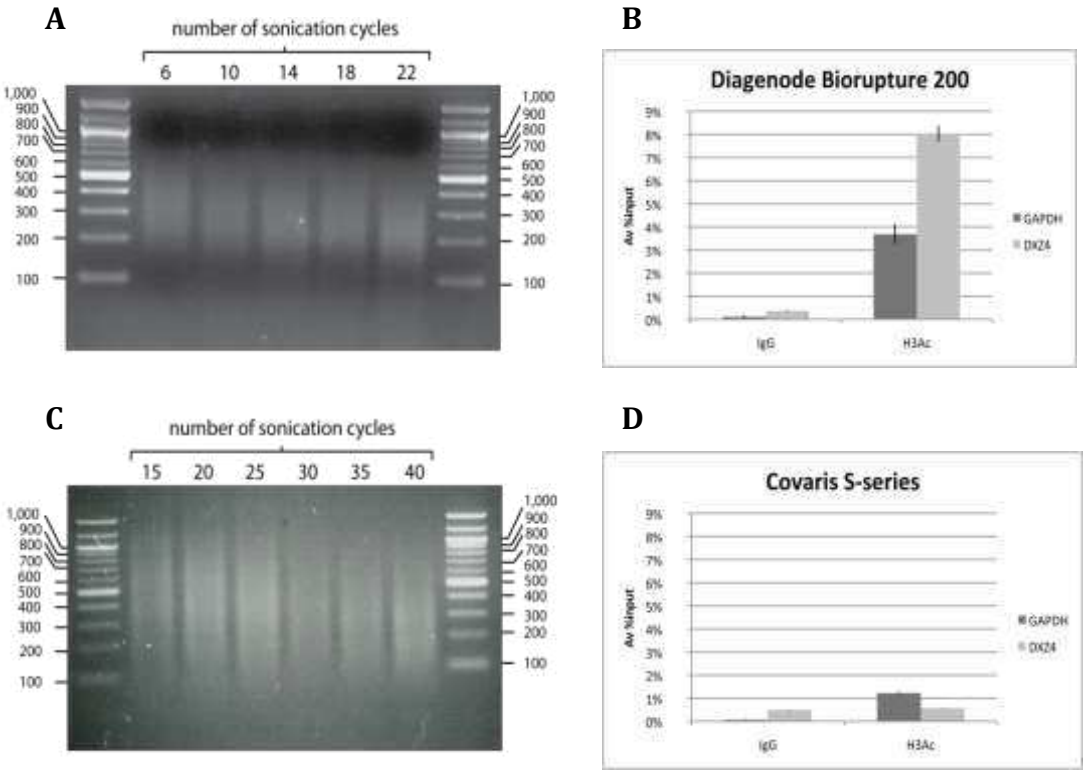


Figure 4.13 Sonication conditions optimisation.

(A) Gel electrophoresis analysis of the sonicated material after 6, 10, 14, 18 and 22 sonication cycles, using Diagenode Biorupture 200. (C) Gel electrophoresis analysis of sonicated material after 15, 20, 25, 30, 35 and 40 sonication cycles, using Covaris S-series. (C) ChIP-qPCR analysis of the samples sonicated for 14 cycles with Diagenode Biorupture 200. (D) ChIP-qPCR analysis of the samples sonicated for 30 cycles with Covaris S-series sonicator. The qPCR results were normalised to the input. The error bars represent the standard error between the triplicate qPCR reactions for each sample.

4.5 Optimisation of the ChIP protocol

ChIP was carried out according to the EZ-Magna ChIP™ A kit (Millipore #17-408) protocol. As mentioned above ChIP results are often subject to some variability (section 1.4). Hence a number of controls were chosen to indicate the level of variability in some of the steps. Before the addition of the antibodies, 10% of the sonicated material was removed as a reference sample (input). This indicates the amount of starting material in each sample before immunoprecipitation (IP) with an antibody. Rabbit IgG antibody was used as the negative control for the immunoprecipitation stage. Any level of DNA recovered from the IgG IP is counted as background signals, which indicate the level of non-specific binding by IgG or the protein A magnetic beads, or other possible contaminant. Acetylated histone H3 (H3Ac) antibody was used as the positive control for the ChIP protocol. H3Ac is known to be enriched in GAPDH gene promoter; hence H3Ac binding at the GAPDH promoter region was used as a positive control for the ChIP protocol. H3K4me2, H3K9me3 and H3K27me3 antibodies were the test antibodies used in this study. These histone antibodies were used in the original study by Zeng et al. (2009).

ChIP was carried out on two control lymphoblastoid cell lines (GM16352, GM08729) and three FSHD lymphoblastoid cell lines (GM16348, GM17939, GM16351). H3K4m2, H3K9me3 and H3K27me3 antibodies were used alongside the control antibodies H3Ac and rabbit IgG. The DIST and LUC7L primer pairs amplify two regions within human chromosome 16p13.3 flanking the alpha-globin gene locus and were previously reported to be negative for H3K4me2 occupancy in EBV-transformed cell lines after ChIP-qPCR and ChIP-chip analyses (De Gobbi et al. 2007). Hence, these regions were used as

negative controls for H3K4me2 in this study. H3K4me2 binding at the DXZ4 region has been shown to be sex-specific, with females having higher levels of H3K4me2 modification at DXZ4 site than males (Chadwick 2008). DXZ4 is a tandem array on the X chromosome, which is similar to the D4Z4 repeats in C/G content. The DXZ4 primer pair was also used as positive control for H3K9me3 (Zeng et al. 2009). The alphaSYBR primer pair amplify a region within the alpha-globin promoter region, which was shown to be negative for H3K9me3 histone modification in the EBV-transformed cell lines. Hence, this region was used as negative control for H3K9me3 binding.

The ChIP samples were analysed by qPCR and the data obtained was then processed according to the normalisation method described in section 2.10. The %input data obtained from each qPCR with a specific primer pair was then used to construct a column graph comparing the %input obtained from each IP. This graph represented the level of each histone modification studied (H3Ac, H3K4me2, H3K9me3 and H3K27me3) in that specific region which was amplified by the primer pair.

H3K9me3, H3K27me3 and in most cases H3K4me2 %input values appeared to be notably higher than that of H3Ac in all genomic regions analysed. This in addition to positive results obtained from the negative control regions, alphaSYBR, DIST and LUC7L, raised the suspicion that the antibodies used in these ChIP analyses have some non-specific binding. Non-specificity of an antibody in ChIP analysis could result in immunoprecipitation of regions of the genome that may not have been bound by the specific protein of interest. Attempts were therefore made to asses the specificity of antibody binding.

4.5.1 H3K9me3 titration

To investigate whether anti-H3K9me3 non-specificity was due to excessive amounts of the antibody in the IP samples, H3K9me3 antibody concentration titration was carried out on both a control and a FSHD lymphoblastoid cell line (Figure 4.14). Prior to this, H3K9me3 antibody was used at the manufacturer's recommended concentration of 4µg/IP. The amount of antibody was reduced from 4µg to 3µg, 2µg and 1µg for the titration experiment. H3Ac and rabbit IgG antibodies were used as positive and negative controls for the ChIP procedures. The data from the qPCR analyses were used to calculate the %input value for all the IPs (Figure 4.14).

The data showed that reducing the amount of the antibody resulted in lowering of the %input value, which represents the DNA recovery after the immunoprecipitation step. A sharp drop in %input values was observed when 1µg of H3K9me3 antibody was used on the FSHD sample. This was not true when the titration ChIP was repeated on a control cell line, where the decrease observed in %input value had more of a gradual decrease. The sharp drop in %input value observed in the FSHD sample was thought to have occurred due to a pipetting error, which resulted in a lower concentration of protein A magnetic beads in this IP compared to the other five IPs in this ChIP.

The %input values obtained from the negative control region, alphaSYBR, stayed close to the %input values obtained from the DXZ4 and Q-PCR regions regardless of the anti-H3K9me3 concentration. It was concluded that the decrease in H3K9me3 antibody concentration does not result in more specific binding.

Interestingly, the same pattern of related %input values at different regions was seen for all anti-H3K9me3 concentrations; H3K9me3 %input value was the lowest at the GAPDH region and the highest at the Q-PCR region (Figure 4.14). Maintenance of this pattern throughout the titration assay indicates that the H3K9me3 antibody shows some specificity to its epitope(s), which was maintained at different concentrations.

The impact of the antibody concentration difference between the different IPs in the titration assay was undermined when the %input data was normalized to the GAPDH data (Figure 4.15). The %input values obtained with DXZ4, Q-PCR and alphaSYBR were normalized to the %input value obtained with GAPDH. This method accounted for some of the differences in the DNA recovery after the immunoprecipitation stage. Note that some level of variability is still observed between the IPs even after the normalisation (Figure 4.15). This highlights the importance of repeating the ChIP-qPCR analyses for obtaining more accurate results.

Thus, it was concluded that as long as the data obtained for an IP is represented as a %input value relative to the %input value of GAPDH, it is comparable to the other IPs with the same antibody. Thus, GAPDH normalisation was used as a method to reduce the effects of the variability between different ChIP assays induced by the IP step.

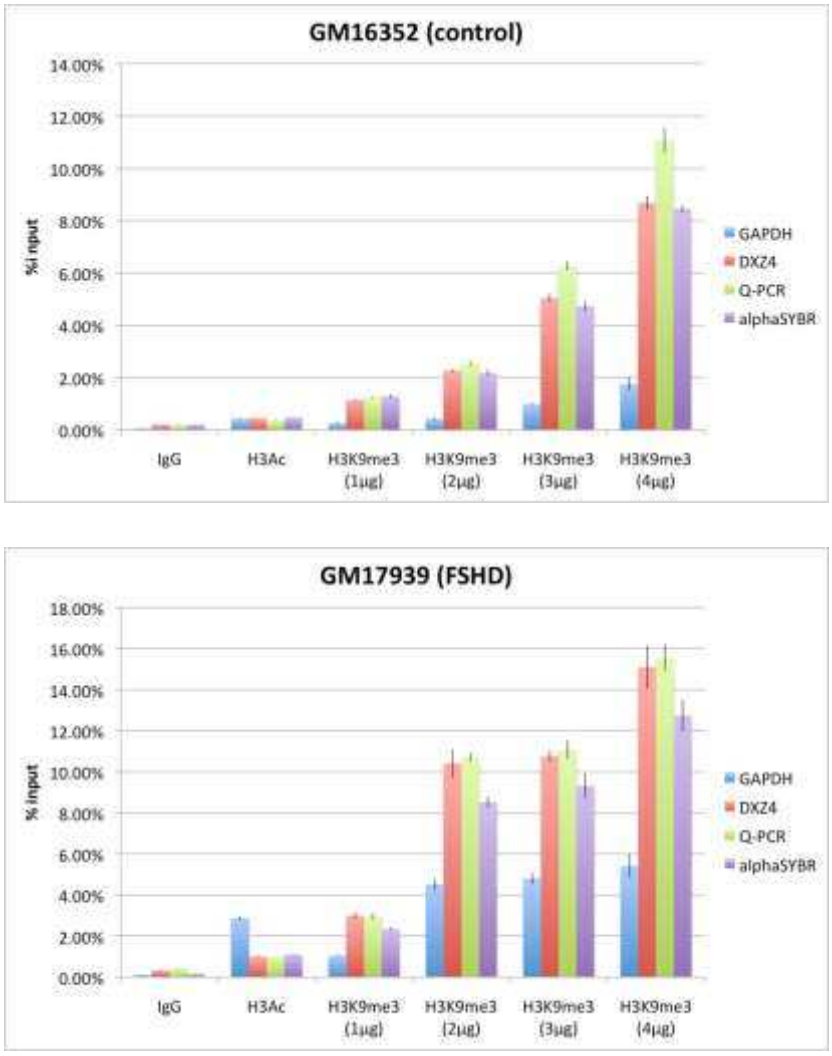


Figure 4.14 Antibody titration analysis.

Parallel ChIP assays using H3K9me3 where performed on a control (GM16352) and a patient (GM17939) lymphoblastoid cell line. The amount of H3K9me3 antibody was reduced from 4µg to 3µg, 2µg and 1µg per IP. H3Ac along with the GAPDH primers was used as positive control for ChIP. Rabbit IgG was used as negative control to indicate any background signals. alphaSYBR primers were used as negative control for H3K9me3 binding. The error bars represent the standard error between the three Ct values obtained for each IP.

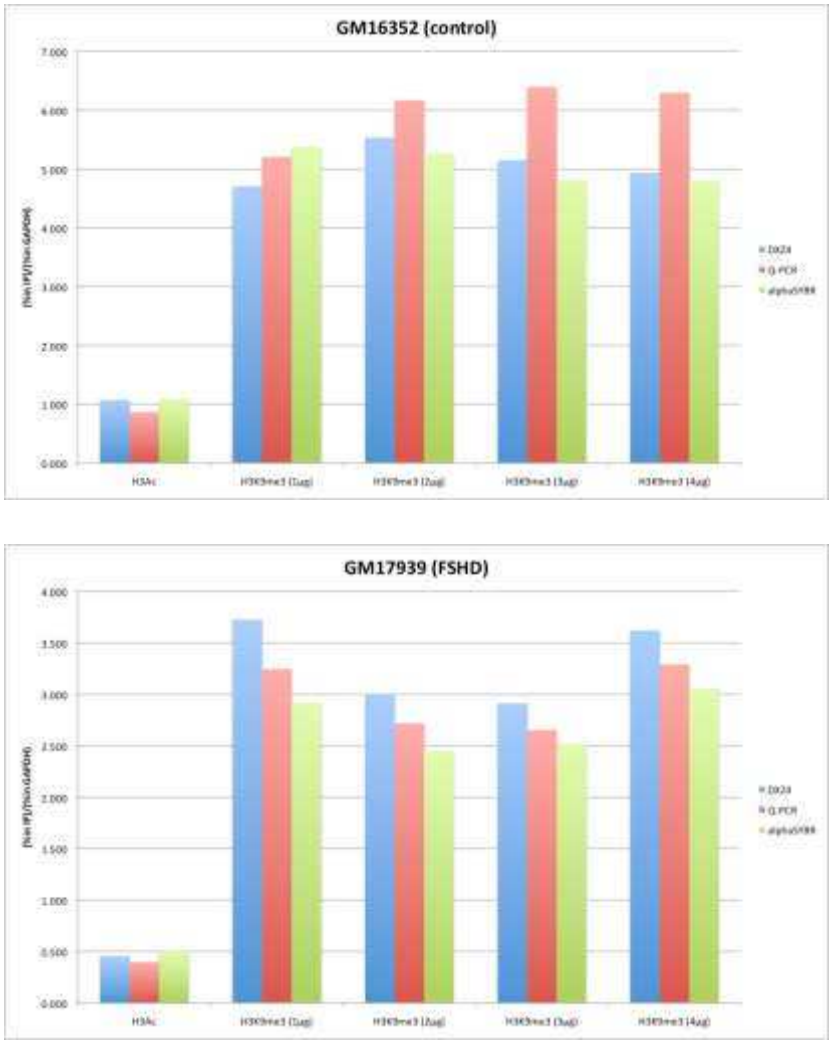


Figure 4.15 GAPDH normalisation of the antibody titration analysis data.
The %input data from DXZ4, Q-PCR and alphaSYBR was normalised to the %input data from GAPDH

4.5.2 Comparison of H3K9me3 antibodies from Diagenode and Millipore

New batches of H3K9me3 antibody were obtained from Millipore (Cat #17-625 Lot #DAM1682222) and Diagenode (Cat#pAb-013-050 Lot #001) in order to compare the specificity during the ChIP procedure. ChIP and qPCR were carried out in parallel using these new antibodies. IgG and H3Ac were used as negative and positive controls for this ChIP. H3K4me3 antibody (Millipore), which was used in all the previous ChIP assays, was also used here as a comparison for %input values that were obtained with the first batch of Millipore H3K9me3 antibody. However, both batches of antibodies failed to recover any significant amount of DNA, even when the ChIP experiments were repeated independently. The %input values obtained from H3Ac and H3K4me2 antibodies were similar to previous ChIP experiments, which indicates that there was unlikely to be a problem with the ChIP or the qPCR procedures. The failure to recover DNA was therefore due to problems with the new batches of antibodies. This problem precluded further analysis of H3K9me3 modification during this project.

4.5.3 Normalisation of the ChIP-qPCR data to GAPDH

The %input data obtained for each antibody using different primer pairs were combined to compare the values at different regions. The data from the two repeat ChIPs on each cell line showed a very similar pattern. However, there was a great difference between %input values, with the second ChIP having much lower recovery than the first ChIP. One possibility is a reduction in the efficiency of the antibodies in the time between the first and the second ChIP. However, normalisation to GAPDH made the %input values more comparable

between the two ChIPs. The mean of the data from the two ChIPs are presented in Figures 4.16 - 4.19.

4.6 Result of the ChIP-qPCR analysis

4.6.1 H3Ac

The normalised ChIP-qPCR data for H3Ac is shown in Figure 4.16. There was a large variability between the cell lines regarding the H3Ac occupancy in the genomic regions that were analysed, with one of the control cell lines (GM08729) yielding the lowest levels in most regions analysed. This was closely followed by one of the FSHD cell lines (GM17939), which had very low levels of H3Ac in all regions except at DXZ4 and 4qB163. The control cell line (GM16352) showed very small variability in H3Ac content between different sites. The H3Ac content for the two remaining FSHD cell lines (GM16348, GM16351) were similar at DXZ4, Q-PCR and alphaSYBR sites. GM16348 showed a lower H3Ac content at DIST, LUC7L and 4qB163 regions compared to GM16351. The highest level of H3Ac was observed at the 4qB163 region of the GM16351 cell line. This high value was obtained from one of the ChIP assays and was not true for the other two. This is indicated by the large error bar on the graph (Figure 4.16).

Overall, there was no significant pattern observed for H3Ac histone modification between control and FSHD cell lines amongst the regions analysed.

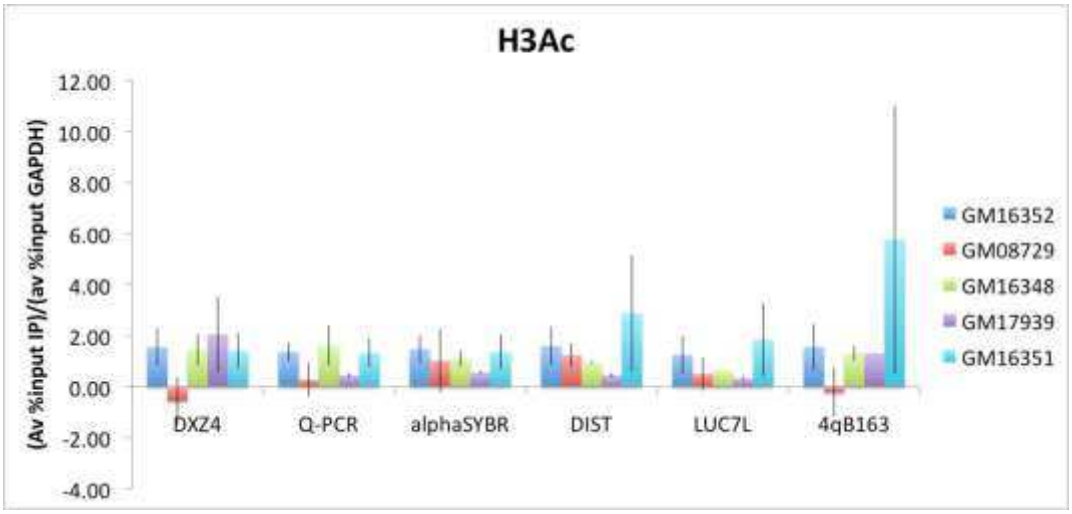


Figure 4.16 ChIP-qPCR results for H3Ac histone modification.

Data is from two control lymphoblastoid cell lines (GM16352, GM08729) and three FSHD lymphoblastoid cell lines (GM16348, GM17939, GM16351). The %input value for each region is normalised %input value of GAPDH. The error bars represent the standard error between the normalised data from three independent ChIP assays (except GM16348, where the standard error is obtained from two separate ChIP assays).

4.6.2 H3K4me2

The normalised ChIP-qPCR data for H3K4me2 is shown in Figure 4.17. DIST and LUC7L regions were used as negative controls for H3K4me2 (see Section 4.5). Our data did not show much difference in H3K4me2 occupancy at these two regions when compared to other sites.

Comparing the different cell lines for the H3K4me2 levels at different genomic regions revealed some level of reduction in H3K4me2 occupancy at the Q-PCR region in FSHD patients compared to the control cell lines. This level of decrease was not observed in any other regions analysed, with exception to DXZ4. The H3K4me2 levels at DXZ4 were previously shown to be sex-specific (Chadwick 2008). A small difference was observed between the H3K4me2 levels of the two controls, with the male cell line (GM08729) yielding lower levels of this histone modification at this site. Therefore the level of H3K4me2 at DXZ4 in the male FSHD cell lines (GM17939 and GM16351) were compared to the male control (GM08729). Similarly the female FSHD cell line (GM16348) was compared to the female control (GM16352). The male FSHD cell lines showed a reduction in H3K4me2 compared to the male control cell line. The female FSHD cell line did not follow the same pattern and did not show a reduction in H3K4me2 histone modification at DXZ4. However, the H3K4me2 levels at DXZ4 for this cell line was not consistent between the two different ChIP-qPCR assays.

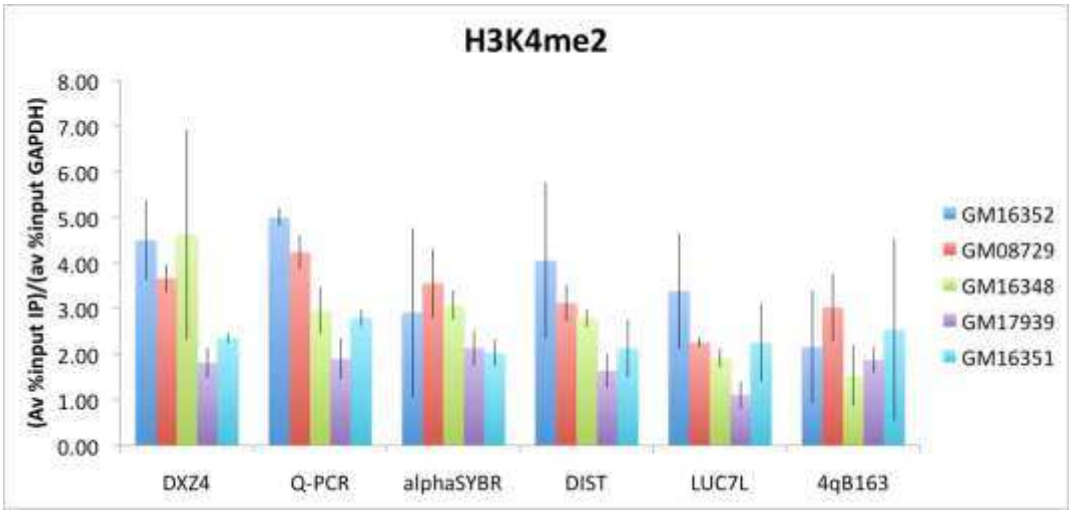


Figure 4.17 ChIP-qPCR results for H3K4me2 histone modification.

Data is from two control lymphoblastoid cell lines (GM16352, GM08729) and three FSHD lymphoblastoid cell lines (GM16348, GM17939, GM16351). The %input value for each region is normalised %input value of GAPDH. The error bars represent the standard error between the normalised data from two independent ChIP assays.

4.6.3 H3K9me3

The normalised ChIP-qPCR data for H3K9me3 is shown in Figure 4.18. The H3K9me3 ChIP-qPCR results from the second ChIP done on GM17939 were excluded from the final analysis. This was because the second batch of Millipore H3K9me3 antibody (Cat #17-625 Lot #DAM1682222) was used for this ChIP, which resulted negative %input values when normalised and were not comparable with the previous data. alphaSYBR region was used as negative control for H3K9me3 binding. The H3K9me3 levels obtained for alphaSYBR region were, in most cases, very similar to the H3K9me3 levels in DXZ4. This pattern was also maintained during the H3K9me3 titration experiment (section 4.5.1). LUC7L region showed lower H3K9me3 levels compared to other regions in all cell lines except for GM16351 (FSHD) where 4qB163 contained the lowest levels of H3K9me3. This, further to the H3K9me3 titration assay, verifies the non-specificity of the H3K9me3 antibody. It is possible that this H3K9me3 antibody also binds to another histone modification at the alphSYBR region that maybe absent from or is present at lower levels at the LUC7L region. This can explain the difference between the H3K9me3 data at alphaSYBR and LUC7L.

The Q-PCR region contained the highest levels of H3K9me3 in all the cell lines analysed. Comparing the H3K9me3 data from different cells revealed a decrease in H3K9me3 levels in FSHD cell lines when compared to the control cell lines at all the genomic regions tested. This difference was less profound between the control cell line, GM16352 and the FSHD cell lines at alphaSYBR, DIST and LUC7L regions compared to the difference observed at DXZ4, Q-PCR and 4qB163 regions. H3K9me3 levels for the second control

cell line, GM08729 were higher than those of the GM16352 control cell line at all the regions analysed. Consequently, all FSHD cell lines demonstrated a reduction in H3K9me3 at all regions compared to this control cell line. ChIP-qPCR was not carried out on an adequate number of control cell lines to enable a statistical analysis of the difference observed between the H3K9me3 occupancy in the control and the FSHD cell lines.

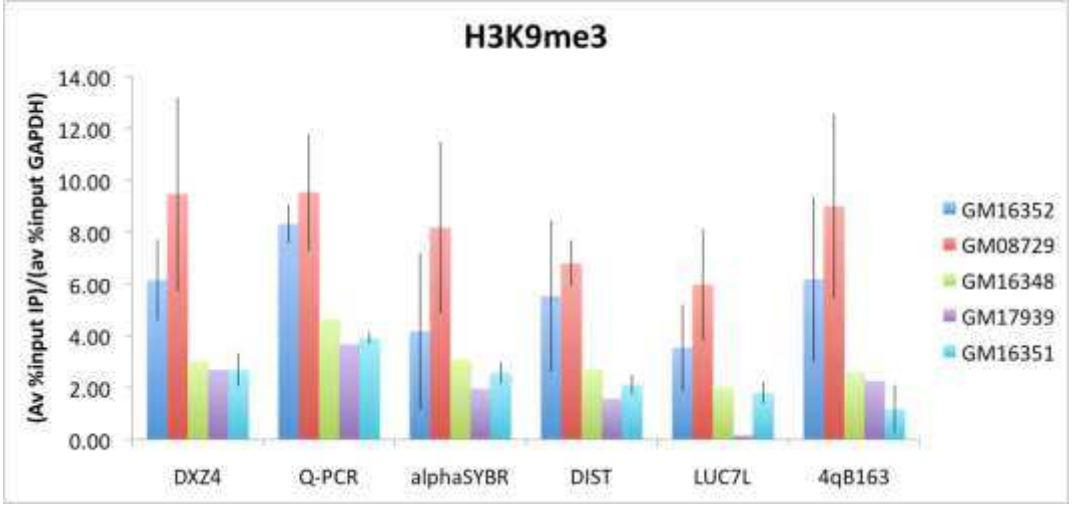


Figure 4.18 ChIP-qPCR results for H3K9me3 histone modification.
Data is from two control lymphoblastoid cell lines (GM16352, GM08729) and three FSHD lymphoblastoid cell lines (GM16348, GM17939, GM16351). The %input value for each region is normalised %input value of GAPDH. The error bars represent the standard error between the normalised data from two independent ChIP assays (excluding GM16348 and GM17939).

4.6.4 H3K27me3

The normalised ChIP-qPCR data for H3K27me3 is shown in Figure 4.19. The highest levels of H3K27me3 were observed at the alpha region for all the cell lines. The lowest levels were observed at the LUC7L region, except for the GM08729 control cell line where 4qB163 region had the lowest levels of H3K27me3.

Comparing the H3K27me3 levels between the cell lines revealed a reduction in H3K27me3 levels at all the regions in FSHD cell lines compared to the control cell lines. H3K27me3 levels were very similar between the two control cell lines at all the regions except at 4qB163. The control cell line, GM08729, showed low levels of H3K27me3 at the 4qB163 region compared to GM16352 (control). The H3K9me3 levels at 4qB163 for this control cell line (GM08729) were similar to the H3K27me3 levels for the FSHD cell lines.

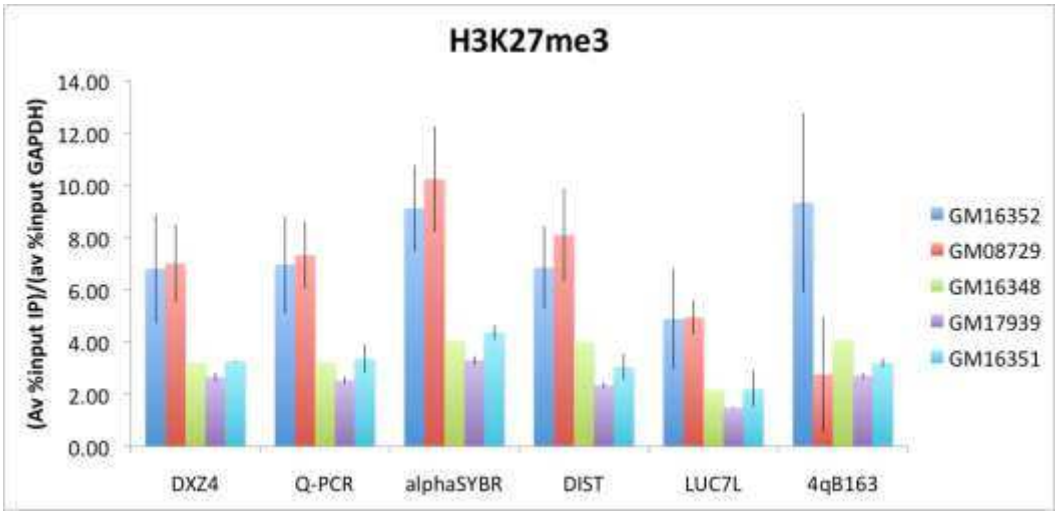


Figure 4.19 ChIP-qPCR results for H3K27me3 histone modification.

Data is from two control lymphoblastoid cell lines (GM16352, GM08729) and three FSHD lymphoblastoid cell lines (GM16348, GM17939, GM16351). The %input value for each region is normalised %input value of GAPDH. The error bars represent the standard error between the normalised data from two independent ChIP assays (excluding GM16348).

5. Discussion

5.1 Q-PCR primer pair specificity to 4q and 10q

The Q-PCR primer pair was designed for qPCR analysis and was shown to amplify products exclusively from chromosome 4 and 10 and not from any other chromosome by Zeng et al. (2009). DNA from somatic cell hybrids carrying individual human chromosomes was used to verify the specificity of this primer pair (Zeng et al. 2009). However, in this study, the specificity of these primers to chromosome 4 and 10 was not confirmed.

The somatic cell hybrids used in this study were obtained from the same company (Coriell Repositories) as the cells used by Zeng et al. (2009), three of which were the same cell lines (GM11687, 11688 and 10479, carrying chromosomes 4, 10 and 14 respectively). Chromosomes 14 and 15 are two of the chromosomes that are known to contain D4Z4-like repeats and for that reason, the Q-PCR primer pair was tested on cell hybrids carrying these two chromosomes (GM10479 and 11418A) as negative controls. Amplification of a product was detected from chromosomes 14 and 15, as well as 4 and 10, which suggests that this primer pair was not as specific as suggested by Zeng et al. (2009), at least not under the qPCR conditions used in this study. Zeng et al. (2009) fail to provide any PCR or qPCR conditions in the published article.

It was speculated that the Q-PCR primer pair show some tendency to chromosome 4 over the other chromosomes. This was based on the observation that the Ct values from the hybrid cell line carrying chromosome 4 (HHW416)

were very close to the Ct values from the human lymphoblastoid cell line containing all chromosomes (GM18207) (Figure 4.4).

5.1.1 XapI digestion of the region amplified by Q-PCR primer pair

The region amplified by the Q-PCR primer pair contains a C/G variant that allows distinction of 4q-derived products from 10q-derived products (Zeng et al. 2009). This C/G variant creates a XapI site on chromosome 4 (C variant), which is absent on chromosome 10 (G variant).

In order to investigate whether the majority of the qPCR products amplified by the Q-PCR primer pair were from chromosome 4, XapI digestion was carried out. There was no digestion of the PCR products from the human lymphoblastoid cell line (GM17939), indicating a majority presence of non-4q-derived products. Partial digestion of the positive control (HHW416, carrying chromosome 4) indicated the presence of some non-4q-derived products (Figure 4.5). Since the HHW416 hybrid cell line is advertised as a mono-chromosomal somatic cell hybrid carrying human chromosome 4 only, it is possible that the Q-PCR primers may amplify a region within the Chinese hamster genome. However, it is also likely that this somatic cell hybrid also contains parts of other human chromosomes.

GM11687 is another mono-chromosomal somatic cell hybrid carrying chromosome 4 only. The Q-PCR primer pair was not tested on the DNA from the GM11687 cell line. The XapI digestion can be repeated on the qPCR products from this cell line to further investigate the specificity of the Q-PCR primer pair.

5.1.2 Alternative Q-PCR primer sequences

The Q-PCR primer sequences were altered in order to increase the specificity of amplification of the D4Z4 repeats on chromosome 4q. However, amplification was still detected from chromosomes 14 and 15, which resulted in the formation of different products to those formed from chromosomes 4 and 10. The different products could be distinguished from the consequent melt curves produced from these samples. The melt curve obtained from the human lymphoblastoid cell line (GM16352) closely resembled the melt curve obtained from chromosomes 4 and 10 (Figure 4.6). It can therefore be concluded that these primer pairs show higher affinity to the D4Z4 repeats on 4q and 10q over chromosomes 14 and 15. Further analysis using other mono-chromosomal somatic cell hybrids is needed to conclude that these primer pairs always preferentially amplify regions from chromosome 4 and 10.

5.2 Antibody specificity

All the antibodies in this study were used at the manufacturer's recommended final concentration for ChIP analysis. The concentration of the antibody used in the ChIP procedure should be as such that it is enough to saturate all target sites but not too in excess to cause non-specific binding. Low amounts of antibody used in IP can cause preferential binding to some regions over others and hence affect the ChIP outcome. H3K9me3, H3K27me3 and in most cases H3K4me2 %input values appeared to be notably higher than that of the control antibody H3Ac in all genomic regions analysed. This, in addition to positive results obtained from the negative control regions, alphaSYBR, DIST and Luc7L, raised the suspicion that the antibodies used in these ChIP analyses had

some non-specific binding. Hence, the H3K9me3 titration assay was carried out.

The titration results showed that reducing the amount of the antibody resulted in lowering of the %input value, which represents the DNA recovery after the immunoprecipitation step (Figure 4.14). At first glance the results from the H3K9me3 antibody titration was to be expected, but one can argue that if the H3K9me3 antibody was highly specific to its epitope, raising the concentration should not have resulted in continuous increase in binding and hence DNA recovery. In theory, the increase in binding should stop once all target sites are saturated. Assuming the saturation step was at some stage before the manufacturer's recommended antibody concentration, our results suggest that the H3K9me3 antibody (Millipore) is not very specific to its epitope.

In the support of this conclusion, the %input values obtained from the negative control region, alpha, stayed close to the %input values obtained from the DXZ4 and Q-PCR regions regardless of the H3K9me3 antibody concentration. This also indicated that the non-specificity of the H3k9me3 antibody was not due to excess of this antibody in each IP.

Interestingly, the pattern obtained from the related %input values at different regions was maintained throughout the change in the anti-H3K9me3 concentration (Figure 4.14). Maintenance of this pattern throughout the titration assay indicates that the H3K9me3 antibody shows specificity to one or a few other epitopes that is maintained at different concentrations.

A study in 2004 (Perez-Burgos et al. 2004) investigated the specificity of commercially available methyl-lysine histone antibodies. This study

investigated the cross-reactivities between these antibodies and other methyl-lysine histone modifications by peptide spotting analysis (dot blots). Their data indicated significant discrepancies in the specificity and avidity of the available methyl-lysine histone antibodies (Perez-Burgos et al. 2004). The main focus of the study was on antibodies against methylated H3K9 and H3K27. This is because the lysine 9 and lysine 27 positions on histone H3 are embedded within the same amino acid sequence "-ARKS-", therefore allowing for considerable cross-reactivities of respective methyl-lysine histone antibodies (Perez-Burgos et al. 2004). The affinity of these antibodies towards different linear peptides was also determined by using a serial dilution of these peptides. The H3K9me2 antibody from Upstate Biotech (now supplying to Millipore) was found to show a high affinity to H3K9me2, but also show minor cross-reactivities with H3K4me2 and H3K27me3. This was different from the H3K9me2 antibody from Abcam, which showed minor cross-specificity to H3K36me2, H4K20 and H3K27me3 (Perez-Burgos et al. 2004). It was proven that these cross-reactivities could be minimised by using a 2-branched peptide that encompasses the H3K9me2 position by five amino acids on either side (Perez-Burgos et al. 2004).

According to the Millipore product information, the H3K27me3 antibody is a 2-branched antibody. However, this is not specified for H3K9me3 or H3K4me2 antibodies. Therefore, it is possible that the H3K9me3 and H3K4me2 antibodies used in this study have some cross-reactivities with other histone modifications. Nevertheless, the aim of this study was to replicate and verify the data from Zeng et al. (2009) who used the same antibodies as what has been used in this study.

A simple method for improving the specificity of methyl-lysine histone antibodies has been suggested (Connor et al. 2010). This method involves pre-incubation of the antibody with the peptide of the histone modification for which the antibody shows non-specificity (Connor et al. 2010). This clean-up procedure resulted in significant improvement in the specificity of the antibodies in ChIP (Connor et al. 2010). This method can be used on polyclonal antibodies prior to ChIP to purify the solution from any antibodies that show any cross-reactivity.

5.3 ChIP-qPCR results

Zeng et al. (2009) reported that there was a reduction in H3K9me3 histone modification in a region within the D4Z4 repeats (Q-PCR region) in FSHD patients. The results obtained from the ChIP-qPCR analysis in this study confirm that there is indeed a reduction in H3K9me3 occupancy at the Q-PCR region in FSHD patients. However, contradictory to Zeng et al. (2009), the reduction of H3K9me3 in FSHD patients was also observed at DXZ4 and other regions analysed (Figure 4.18).

There was a difference in H3K9me3 histone modification occupancy between the two control cell lines (GM16352 and GM08729). This is not unexpected since there are numerous variations within the genome and consequently variation in the epigenome is inevitable. This demonstrates the importance of having more than one control cell line in epigenetic studies. This issue highlights a weakness in the ChIP-qPCR data from Zeng et al. (2009). The ChIP-qPCR data from Zeng et al. (2009) are shown as relative to a control. Our data shows that normalisation to different controls can result in different

conclusions (Figure 5.1). If the data is normalised to GM16352 control cell line, it can be concluded that there is about a 50% reduction in H3K9me3 histone modification at the Q-PCR region in FSHD patients. Whereas, if the data is normalised to GM08729 control cell line, it can be concluded that there is a 70% reduction in H3K9me3 histone modification at the Q-PCR region, which is more similar to the data from Zeng et al. (2009). Therefore, it can be concluded that a larger number of control and FSHD cell lines is needed to validate these findings. A larger number of cell lines would allow for calculation of an average level of a specific histone modification at the region of interest in control cell lines, which could then be compared to individual FSHD cell lines or to an average of multiple FSHD cell lines. This would also allow for statistical analysis in order to determine whether there is a significant difference between the FSHD and control cell lines.

The ChIP-qPCR analysis also indicated a reduction in H3K27me3 in FSHD patients amongst all the regions analysed (Figure 4.19). This is contradictory to the data from Zeng et al. (2009), where no difference in H3K27me3 at the Q-PCR region between FSHD and controls was reported. Since the reduction of H3K27me3 histone modification was not exclusive to a specific region, it can be speculated that there is a global reduction in this histone modification in FSHD patients and the D4Z4 repeats are one of the many regions affected. In order to investigate this further, ChIP-qPCR analysis of a wider range of sites within the genome is required. Coupling of ChIP with microarray (ChIP-CHIP) and high throughput sequencing (ChIP-seq) would allow a more precise genome-wide analysis of the H3K27me3 histone modification in FSHD patients.

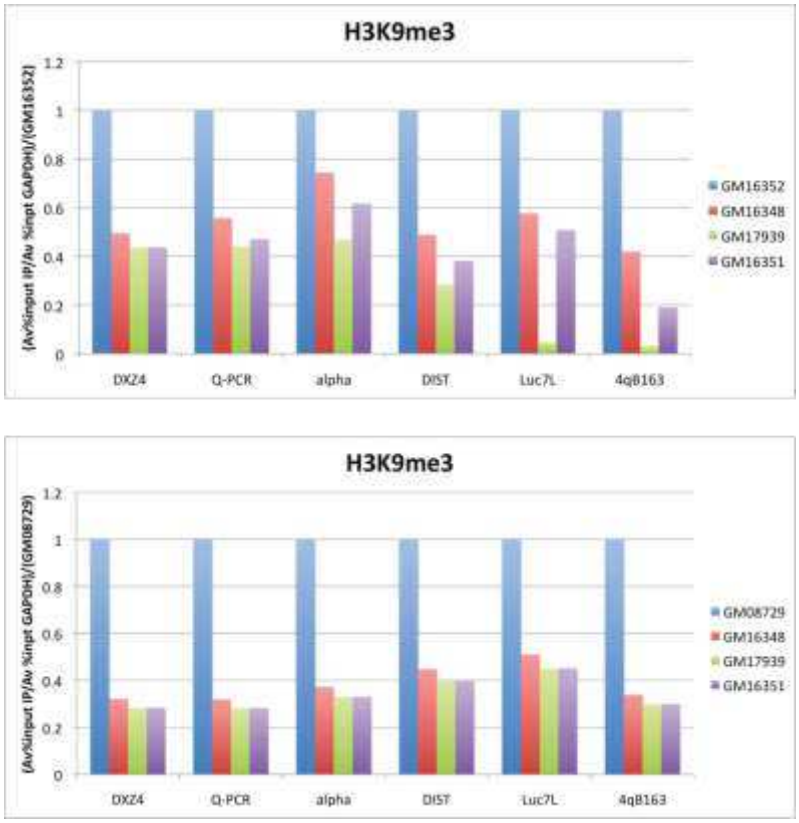


Figure 5.1 Normalisation to a control cell line.

The ChIP-qPCR data was normalised to either GM16352 control cell line (top image) or to the GM08729 control cell line (bottom image).

5.4 Future work

As discussed above, a larger number of control and patient cell lines is needed for statistical analysis and validation of the ChIP-qPCR data obtained so far. In addition, the use of Taq-Man probe instead of SYBR green would increase the specificity of the qPCR analysis. The probes could be designed using the sequence variants that differentiate between the D4Z4 repeats on 4q, 10q and other D4Z4-like sequences on other chromosomes. Haplotype-specific probes could also be designed using the sequence variants within the p13-E11 region to compare the ChIP-qPCR data at the proximal regions of the disease permissive and non-permissive alleles.

Recently, a paper was published reporting the presence of two polymorphisms in the pLAM sequence of the disease permissive haplotypes that create a polyadenylation site for the distal DUX4 transcript (Lemmers et al. 2010). The pLAM region is immediately distal to the last repeat and is absent from the non-permissive 4qB haplotype. The polyadenylation site in the pLAM region is thought to stabilise the transcripts from the distal D4Z4 repeat (Dixit et al. 2007; Snider et al. 2009). This, in combination with the chromatin relaxation at the D4Z4 repeats, may lead to increase in DUX4 transcript levels and therefore it is hypothesised that FSHD is caused by toxic gain of function of DUX4 (Lemmers et al. 2010). It would be extremely interesting to study and compare the histone modification changes at the last repeat and the pLAM region between controls carrying the permissive haplotypes, FSHD2 patients and FSHD1 patients. However, it will be quite challenging to design primers and probes specific to the last repeat and the pLAM region, which consists of two

CG-rich LSau regions (similar to LSau within each repeat) and a 68bp region of Sau 3A repeats (van Geel et al. 2002).

In conclusion, the ChIP-qPCR data from this project was in agreement with the data published by Zeng et al. (2009) on the reduction of H3K9me3 histone modification at the D4Z4 repeats in FSHD patients. However, contradictory to their data, the reduction of this histone modification was also observed on other regions analysed. A global reduction of H3K27me3 was also observed in FSHD patients. Further studies using a larger set of control and FSHD cell lines with more precise detection methods (ChIP-CHIP or ChIP-seq), coupled with DNA methylation studies, should lead to a deeper understanding of the change in the chromatin structure, especially at the distal region of the 4q allele, in FSHD patients.

References:

Alexiadis, V., Ballestas, M. E., Sanchez, C., Winokur, S., Vedanarayanan, V., Warren, M. and Ehrlich, M. (2007). RNAPol-ChIP analysis of transcription from FSHD-linked tandem repeats and satellite DNA. Biochimica Et Biophysica Acta-Gene Structure and Expression **1769**: 29-40.

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U. and Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nature Genetics **23**: 185-188.

Ansseau, E., Laoudj-Chenivesse, D., Marcowycz, A., Tassin, A., Vanderplanck, C., Sauvage, S., Barro, M., Mahieu, I., Leroy, A., Leclercq, I., Mainfroid, V., Figlewicz, D., Mouly, V., Butler-Browne, G., Belayew, A. and Coppee, F. (2009). DUX4c Is Up-Regulated in FSHD. It Induces the MYF5 Protein and Human Myoblast Proliferation. PLoS One **4**: e7482.

Bakay, M., Wang, Z., Melcon, G., Schiltz, L., Xuan, J., Zhao, P., Sartorelli, V., Seo, J., Pegoraro, E., Angelini, C., Shneiderman, B., Escolar, D., Chen, Y. W., Winokur, S. T., Pachman, L. M., Fan, C., Mandler, R., Nevo, Y., Gordon, E., Zhu, Y., Dong, Y., Wang, Y. and Hoffman, E. P. (2006). Nuclear envelope dystrophies show a transcriptional fingerprint suggesting disruption of Rb-MyoD pathways in muscle regeneration. Brain **129**: 996-1013.

Barski, A., Cuddapah, S., Cui, K. R., Roh, T. Y., Schones, D. E., Wang, Z. B., Wei, G., Chepelev, I. and Zhao, K. J. (2007). High-resolution profiling of histone methylations in the human genome. Cell **129**: 823-837.

Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., 3rd, Gingeras, T. R., Schreiber, S. L. and Lander, E. S. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell **120**: 169-181.

Bestor, T. H. and Tycko, B. (1996). Creation of genomic methylation patterns. Nature Genetics **12**: 363-367.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & Development **16**: 6-21.

Bodega, B., Ramirez, G. D., Grasser, F., Cheli, S., Brunelli, S., Mora, M., Meneveri, R., Marozzi, A., Mueller, S., Battaglioli, E. and Ginelli, E. (2009). Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. Bmc Biology **7**: 41.

Bosnakovski, D., Lamb, S., Simsek, T., Xu, Z., Belayew, A., Perlingeiro, R. and Kyba, M. (2008). DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. Experimental Neurology **214**: 87-96.

Cedar, H. and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nature Reviews Genetics **10**: 295-304.

Celegato, B., Capitanio, D., Pescatori, M., Romualdi, C., Pacchioni, B., Cagnin, S., Vigano, A., Colantoni, L., Begum, S., Ricci, E., Wait, R., Lanfranchi, G. and Gelfi, C. (2006). Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre differentiation and a general deregulation of MyoD-dependent genes. Proteomics **6**: 5303-5321.

Chadwick, B. P. (2008). DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. Genome Research **18**: 1259-1269.

Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G. and Li, B. F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science **277**: 1996-2000.

Clapp, J., Mitchell, L. M., Bolland, D. J., Fantes, J., Corcoran, A. E., Scoffing, P. J., Armour, J. A. L. and Hewitt, J. E. (2007). Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. American Journal of Human Genetics **81**: 264-279.

Connor, C., Cheung, I., Simon, A., Jakovcevski, M., Weng, Z. and Akbarian, S. (2010). A simple method for improving the specificity of anti-methyl histone antibodies. Epigenetics **5**.

Cunningham, M. D., Kassis, J. A. and Pfeifer, K. (2010). Chromatin modifiers, cognitive disorders, and imprinted genes. Developmental Cell **18**: 169-170.

De Gobbi, M., Anguita, E., Hughes, J., Sloane-Stanley, J. A., Sharpe, J. A., Koch, C. M., Dunham, I., Gibbons, R. J., Wood, W. G. and Higgs, D. R. (2007). Tissue-specific histone modification and transcription factor binding in alpha globin gene expression. Blood **110**: 4503-4510.

de Greef, J. C., Frants, R. R. and van der Maarel, S. M. (2008). Epigenetic mechanisms of facioscapulohumeral muscular dystrophy. Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis **647**: 94-102.

de Greef, J. C., Lemmers, R., van Engelen, B. G. M., Sacconi, S., Venance, S. L., Frants, R. R., Tawil, R. and van der Maarel, S. M. (2009). Common Epigenetic Changes of D4Z4 in Contraction-Dependent and Contraction-Independent FSHD. Human Mutation **30**: 1449-1459.

de Greef, J. C., Wohlgemuth, M., Chan, O. A., Hansson, K. B., Smeets, D., Frants, R. R., Weemaes, C. M., Padberg, G. W. and van der Maarel, S. M. (2007). Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. Neurology **69**: 1018-1026.

Deidda, G., Cacurri, S., Grisanti, P., Vigneti, E., Piazzo, N. and Felicetti, L. (1995). Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular-

dystrophy locus on chromosome 4qter. European Journal of Human Genetics **3**: 155-167.

Deidda, G., Cacurri, S., Piazzo, N. and Felicetti, L. (1996). Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). Journal of Medical Genetics **33**: 361-365.

Dixit, M., Ansseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., Sauvage, S., Mattotti, C., van Acker, A. M., Leo, O., Figiewicz, D., Barro, M., Laoudj-Chenivesse, D., Belayew, A., Coppee, F. and Chen, Y. W. (2007). DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proceedings of the National Academy of Sciences, USA **104**: 18157-18162.

Doerner, A., Pauschinger, M., Badorff, A., Noutsias, M., Giessen, S., Schulze, K., Bilger, J., Rauch, U. and Schultheiss, H. P. (1997). Tissue-specific transcription pattern of the adenine nucleotide translocase isoforms in humans. Febs Letters **414**: 258-262.

Esteve, P. O., Chin, H. G., Smallwood, A., Feehery, G. R., Gangisetty, O., Karpf, A. R., Carey, M. F. and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes & Development **20**: 3089-3103.

Ferguson-Smith, A. C. and Surani, M. A. (2001). Imprinting and the epigenetic asymmetry between parental genomes. Science **293**: 1086-1089.

Frank, D., Keshet, I., Shani, M., Levine, A., Razin, A. and Cedar, H. (1991). Demethylation of CpG islands in embryonic-cells. Nature **351**: 239-241.

Fuks, F., Hurd, P. J., Deplus, R. and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Research **31**: 2305-2312.

Funakoshi, M., Goto, K. and Arahata, K. (1998). Epilepsy and mental retardation in a subset of early onset 4q35-facioscapulohumeral muscular dystrophy. Neurology **50**: 1791-1794.

Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M. A., Bottinelli, R., Green, M. R. and Tupler, R. (2006). Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. Nature **439**: 973-977.

Gabellini, D., Green, M. R. and Tupler, R. (2002). Inappropriate gene activation in FSHD: A repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell **110**: 339-348.

Gabriels, J., Beckers, M. C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S. M., Padberg, G. W., Frants, R. R., Hewitt, J. E., Collen, D. and Belayew, A. (1999). Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene **236**: 25-32.

Garrick, D., De Gobbi, M., Samara, V., Rugless, M., Holland, M., Ayyub, H., Lower, K., Sloane-Stanley, J., Gray, N., Koch, C., Dunham, I. and Higgs, D. R. (2008). The role of the polycomb complex in silencing alpha-globin gene expression in nonerythroid cells. Blood **112**: 3889-3899.

Gibbons, R. J., McDowell, T. L., Raman, S., O'Rourke, D. M., Garrick, D., Ayyub, H. and Higgs, D. R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nature Genetics **24**: 368-371.

Gilbert, J. R., Stajich, J. M., Wall, S., Carter, S. C., Qiu, H., Vance, J. M., Stewart, C. S., Speer, M. C., Pufky, J., Yamaoka, L. H. and et al. (1993). Evidence for heterogeneity in facioscapulohumeral muscular dystrophy (FSHD). The American Journal of Human Genetics **53**: 401-408.

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., Golic, K. G., Jacobsen, S. E. and Bestor, T. H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science **311**: 395-398.

Gruenbaum, Y., Cedar, H. and Razin, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. Nature **295**: 620-622.

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. and Young, R. A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell **130**: 77-88.

Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G. and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature **460**: 410-U130.

Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M. R. and Gartler, S. M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proceedings of the National Academy of Sciences, USA **96**: 14412-14417.

He, H. H., Meyer, C. A., Shin, H., Bailey, S. T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., Mieczkowski, P., Lieb, J. D., Zhao, K., Brown, M. and Liu, X. S. (2010). Nucleosome dynamics define transcriptional enhancers. Nature Genetics **42**: 343-347.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E. and Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genetics **39**: 311-318.

Hermann, A., Schmitt, S. and Jeltsch, A. (2003). The human Dnmt2 has residual DNA-(Cytosine-C5) methyltransferase activity. Journal of Biological Chemistry **278**: 31717-31721.

Hewitt, J. E., Lyle, R., Clark, L. N., Valleley, E. M., Wright, T. J., Wijmenga, C., Vandeutekom, J. C. T., Francis, F., Sharpe, P. T., Hofker, M., Frants, R. R. and Williamson, R. (1994). Analysis of the tandem repeat locus D4Z4

associated with facioscapilohumeral muscular-dystrophy. Human Molecular Genetics **3**: 1287-1295.

Horike, S., Cai, S. T., Miyano, M., Cheng, J. F. and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nature Genetics **37**: 31-40.

Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A. and Viegaspequignot, E. (1993). An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. Human Molecular Genetics **2**: 731-735.

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. and Cheng, X. D. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature **449**: 248-251.

Jiang, G. C., Yang, F., van Overveld, P. G. M., Vedanarayanan, V., van der Maarel, S. and Ehrlich, M. (2003). Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. Human Molecular Genetics **12**: 2909-2921.

Jiang, Y. H., Bressler, J. and Beaudet, A. L. (2004). Epigenetics and human disease. Annual Review of Genomics and Human Genetics **5**: 479-510.

Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J. and Wolffe, A. P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature Genetics **19**: 187-191.

Kafri, T., Ariel, M., Brandeis, M., Shemer, R., Urven, L., McCarrey, J., Cedar, H. and Razin, A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes & Development **6**: 705-714.

Kawamura-Saito, M., Yamazaki, Y., Kaneko, K., Kawaguchi, N., Kanda, H., Mukai, H., Gotoh, T., Motoi, T., Fukayama, M., Aburatani, H., Takizawa, T.

and Nakamura, T. (2006). Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Hum Mol Genet **15**: 2125-2137.

Kernohan, K. D., Jiang, Y., Tremblay, D. C., Bonvissuto, A. C., Eubanks, J. H., Mann, M. R. W. and Berube, N. G. (2010). ATRX Partners with Cohesin and MeCP2 and Contributes to Developmental Silencing of Imprinted Genes in the Brain. Developmental Cell **18**: 191-202.

Kim, T. H., Barrera, L. O., Zheng, M., Qu, C., Singer, M. A., Richmond, T. A., Wu, Y., Green, R. D. and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature **436**: 876-880.

Kondo, T., Bobek, M. P., Kuick, R., Lamb, B., Zhu, X. X., Narayan, A., Bourc'his, D., Viegas-Pequignot, E., Ehrlich, M. and Hanash, S. M. (2000). Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. Human Molecular Genetics **9**: 597-604.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell **128**: 693-705.

Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C. B., Sauvage, S., Mattotti, C., Arias, C., Corona, E. D., Nufiez, N. G., Leo, O., Wattiez, R., Iglewicz, D. F., Laoudj-Chenivesse, D., BelayeW, A., Coppe, F. and Rosa, A. L. (2007). The DUX4 gene at the FSHDIA locus encodes a pro-apoptotic protein. Neuromuscular Disorders **17**: 611-623.

Kundu, T. K. and Rao, M. R. (1999). CpG islands in chromatin organization and gene expression. Journal of Biochemistry **125**: 217-222.

Lachner, M. and Jenuwein, T. (2002). The many faces of histone lysine methylation. Current Opinion in Cell Biology **14**: 286-298.

Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A. G., Rechavi, G. and Cedar, H. (2007). Role of DNA

methylation in stable gene repression. Journal of Biological Chemistry **282**: 12194-12200.

Laoudj-Chenivesse, D., Carnac, G., Bisbal, C., Hugon, G., Bouillot, S., Desnuelle, C., Vassetzky, Y. and Fernandez, A. (2005). Increased levels of adenine nucleotide translocator 1 protein and response to oxidative stress are early events in facioscapulohumeral muscular dystrophy muscle. Journal of Molecular Medicine **83**: 216-224.

Lee, J. H., Goto, K., Matsuda, C. and Arahata, K. (1995). Characterization of a tandemly repeated 3.3-kb KpnI unit in the facioscapulohumeral muscular dystrophy (FSHD) gene region on chromosome 4q35. Muscle Nerve **2**: S6-13.

Lee, J. T. and Jaenisch, R. (1997). The (epi)genetic control of mammalian X-chromosome inactivation. Current Opinion in Genetics & Development **7**: 274-280.

Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R. and Young, R. A. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell **125**: 301-313.

Lehming, N., Le Saux, A., Schuller, J. and Ptashne, M. (1998). Chromatin components as part of a putative transcriptional repressing complex. Proceedings of the National Academy of Sciences, USA **95**: 7322-7326.

Lehnertz, B., Ueda, Y., Derijck, A. A. H. A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T. P., Li, E., Jenuwein, T. and Peters, A. H. F. M. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Current Biology **13**: 1192-1200.

Lemmers, R., de Kievit, P., Sandkuijl, L., Padberg, G. W., van Ommen, G. J. B., Frants, R. R. and van der Maarel, S. M. (2002). Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nature Genetics **32**: 235-236.

Lemmers, R., Osborn, M., Haaf, T., Rogers, M., Frants, R. R., Padberg, G. W., Cooper, D. N., van der Maarel, S. M. and Upadhyaya, M. (2003). D4F104S1 deletion in facioscapulohumeral muscular dystrophy - Phenotype, size, and detection. Neurology **61**: 178-183.

Lemmers, R., van der Maarel, S. M., van Deutekom, J. C. T., van der Wielen, M. J. R., Deidda, G., Dauwerse, H. G., Hewitt, J., Hofker, M., Bakker, E., Padberg, G. W. and Frants, R. R. (1998). Inter- and intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. Human Molecular Genetics **7**: 1207-1214.

Lemmers, R., Wohlgemuth, M., Frants, R. R., Padberg, G. W., Morava, E. and van der Maarel, S. M. (2004). Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. American Journal of Human Genetics **75**: 1124-1130.

Lemmers, R., Wohlgemuth, M., van der Gaag, K. J., van der Vliet, P. J., van Teijlingen, C. M. M., de Knijff, P., Padberg, G. W., Frants, R. R. and van der Maarel, S. M. (2007). Specific sequence variations within the 4q35 region are associated with Facioscapulohumeral muscular dystrophy. American Journal of Human Genetics **81**: 884-894.

Lemmers, R. J., van der Vliet, P. J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J. G., Snider, L., Straasheijm, K. R., Jan van Ommen, G., Padberg, G. W., Miller, D. G., Tapscott, S. J., Tawil, R., Frants, R. R. and van der Maarel, S. M. (2010). A Unifying Genetic Model for Facioscapulohumeral Muscular Dystrophy. Science **(in press)**.

Lemmers, R. J. L., de Kievit, P., van Geel, M., van der Wielen, M. J., Bakker, E., Padberg, G. W., Frants, R. R. and van der Maarel, S. M. (2001). Complete

allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis. Annals of Neurology **50**: 816-819.

Leonhardt, H., Page, A. W., Weier, H. U. and Bestor, T. H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA-replication in mammalian nuclei. Cell **71**: 865-873.

Li, K., Warner, C. K., Hodge, J. A., Minoshima, S., Kudoh, J., Fukuyama, R., Maekawa, M., Shimizu, Y., Shimizu, N. and Wallace, D. C. (1989). A human-muscle adenine-nucleotide translocator gene has 4 exons, is located on chromosome-4, and is differentially expressed. Journal of Biological Chemistry **264**: 13998-14004.

Lunt, P. W. and Harper, P. S. (1991). Genetic-Counseling in Facioscapulohumeral Muscular-Dystrophy. Journal of Medical Genetics **28**: 655-664.

Lunt, P. W., Jardine, P. E., Koch, M. C., Maynard, J., Osborn, M., Williams, M., Harper, P. S. and Upadhyaya, M. (1995). Correlation between fragment size at D4F10S1 and age at onset or at weelchair use, with a possible generational-effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular-dystrophy (FSHD). Human Molecular Genetics **4**: 951-958.

Lyle, R., Wright, T. J., Clark, L. N. and Hewitt, J. E. (1995). The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. Genomics **28**: 389-397.

Macleod, D., Charlton, J., Mullins, J. and Bird, A. P. (1994). SP1 sites in the mouse Aprt gene promoter are required to prevent methylation of the CpG island. Genes & Development **8**: 2282-2292.

Masny, P. S., Bengtsson, U., Chung, S. A., Martin, J. H., van Engelen, B., van der Maarel, S. M. and Winokur, S. T. (2004). Localization of 4q35.2 to the

nuclear periphery: is FSHD a nuclear envelope disease? Humam Molecular Genetics **13**: 1857-1871.

Masny, P. S., Chan, O. Y., de Greef, J. C., Bengtsson, U., Ehrlich, M., Tawil, R., Lock, L. F., Hewitt, J. E., Stocksdale, J., Martin, J. H., van der Maarel, S. M. and Winokur, S. T. (2010). Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei. European Journal of Human Genetics **18**: 448-456.

Monk, M., Boubelik, M. and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ-cell lineages during mouse embryo development. Development **99**: 371-382.

Mukhopadhyay, A., Deplancke, B., Walhout, A. J. M. and Tissenbaum, H. A. (2008). Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in Caenorhabditis elegans. Nature Protocols **3**: 698-709.

Nan, X. S., Hou, J. H., Maclean, A., Nasir, J., Lafuente, M. J., Shu, X. H., Kriaucionis, S. and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proceedings of the National Academy of Sciences, USA **104**: 2709-2714.

Nan, X. S., Meehan, R. R. and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Research **21**: 4886-4892.

Nan, X. S., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature **393**: 386-389.

Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D. and Bird, A. (1999). MBD2 is a

transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nature Genetics **23**: 58-61.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell **99**: 247-257.

Okano, M., Xie, S. P. and Li, E. (1998). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Research **26**: 2536-2540.

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K. Q., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X. D. and Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature **448**: 714-U713.

Ottaviani, A., Rival-Gervier, S., Boussouar, A., Foerster, A. M., Rondier, D., Sacconi, S., Desnuelle, C., Gilson, E. and Magdinier, F. (2009). The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. PLoS Genet **5**: e1000394.

Ottaviani, A., Schluth-Bolard, C., Rival-Gervier, S., Boussouar, A., Rondier, D., Foerster, A. M., Morere, J., Bauwens, S., Gazzo, S., Callet-Bauchu, E., Gilson, E. and Magdinier, F. (2009). Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF. Embo Journal **28**: 2428-2436.

Pauler, F. M., Sloane, M. A., Huang, R., Regha, K., Koerner, M. V., Tamir, I., Sommer, A., Aszodi, A., Jenuwein, T. and Barlow, D. P. (2009). H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. Genome Research **19**: 221-233.

Perez-Burgos, L., Peters, A. H., Opravil, S., Kauer, M., Mechtler, K. and Jenuwein, T. (2004). Generation and characterization of methyl-lysine histone antibodies. Methods Enzymol **376**: 234-254.

Petrov, A., Pirozhkova, I., Carnac, G., Laoudj, D., Lipinski, M. and Vassetzky, Y. S. (2006). Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. Proceedings of the National Academy of Sciences, USA **103**: 6982-6987.

Pfaffl, M. W., Tichopad, A., Prgomet, C. and Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. Biotechnology Letters **26**: 509-515.

Pirozhkova, I., Petrov, A., Dmitriev, P., Laoudj, D., Lipinski, M. and Vassetzky, Y. (2008). A Functional Role for 4qA/B in the Structural Rearrangement of the 4q35 Region and in the Regulation of FRG1 and ANT1 in Facioscapulohumeral Dystrophy. PLoS One **3**.

Rijkers, T., Deidda, G., van Koningsbruggen, S., van Geel, M., Lemmers, R., van Deutekom, J. C. T., Figlewicz, D., Hewitt, J. E., Padberg, G. W., Frants, R. R. and van der Maarel, S. M. (2004). FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. Journal of Medical Genetics **41**: 826-836.

Roh, T. Y., Cuddapah, S., Cui, K. and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. Proceedings of the National Academy of Sciences, USA **103**: 15782-15787.

Roh, T. Y., Cuddapah, S. and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes & Development **19**: 542-552.

Roh, T. Y., Wei, G., Farrell, C. M. and Zhao, K. (2007). Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Research **17**: 74-81.

Sarraf, S. A. and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Molecular Cell **15**: 595-605.

Siegfried, Z., Eden, S., Mendelsohn, M., Feng, X., Tsuberi, B. Z. and Cedar, H. (1999). DNA methylation represses transcription in vivo. Nature Genetics **22**: 203-206.

Smallwood, A., Esteve, P. O., Pradhan, S. and Carey, M. (2007). Functional cooperation between HP1 and DNMT1 mediates gene silencing. Genes & Development **21**: 1169-1178.

Snider, L., Asawachaicharn, A., Tyler, A. E., Geng, L. N., Petek, L. M., Maves, L., Miller, D. G., Lemmers, R., Winokur, S. T., Tawil, R., van der Maarel, S. M., Filippova, G. N. and Tapscott, S. J. (2009). RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. Human Molecular Genetics **18**: 2414-2430.

Solomon, M. J., Larsen, P. L. and Varshavsky, A. (1988). Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. Cell **53**: 937-947.

Squazzo, S. L., O'Geen, H., Komashko, V. M., Krig, S. R., Jin, V. X., Jang, S. W., Margueron, R., Reinberg, D., Green, R. and Farnham, P. J. (2006). Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Research **16**: 890-900.

Stancheva, I., Collins, A. L., Van den Veyver, I. B., Zoghbi, H. and Meehan, R. R. (2003). A mutant form of MeCP2 protein associated with human Rett syndrome cannot be displaced from methylated DNA by notch in Xenopus embryos. Molecular Cell **12**: 425-435.

Stepien, G., Torroni, A., Chung, A. B., Hodge, J. A. and Wallace, D. C. (1992). Differential expression of adenine-nucleotide translocator isoforms in mammalian-tissues and during muscle-cell diferentiation. Journal of Biological Chemistry **267**: 14592-14597.

Sterner, D. E. and Berger, S. L. (2000). Acetylation of histones and transcription-related factors. Microbiology and Molecular Biology Reviews **64**: 435-459.

Strohman, R. C. (1995). Linear genetics, non-linear epigenetics: Complementary approaches to understanding complex diseases. Integrative Physiological and Behavioral Science **30**: 273-282.

Sun, H. B., Shen, J. and Yokota, H. (2000). Size-dependent positioning of human chromosomes in interphase nuclei. Biophysical Journal **79**: 184-190.

Tam, R., Smith, K. P. and Lawrence, J. B. (2004). The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres. Journal of Cell Biology **167**: 269-279.

Tanabe, H., Habermann, F. A., Solovei, I., Cremer, M. and Cremer, T. (2002). Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. Mutation Research **504**: 37-45.

Tawil, R., Forrester, J., Griggs, R. C., Mendell, J., Kissel, J., McDermott, M., King, W., Weiffenbach, B., Figlewicz, D., Cos, L., Langsam, A., Pandya, S., Martens, B., Brower, C., Herr, B., Downing, K. and Gorell, W. C. (1996). Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. Annals of Neurology **39**: 744-748.

Tawil, R. and Van der Maarel, S. M. (2006). Facioscapulohumeral muscular dystrophy. Muscle & Nerve **34**: 1-15.

Thomas, M. J. and Seto, E. (1999). Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? Gene **236**: 197-208.

Tupler, R., Berardinelli, A., Barbierato, L., Frants, R., Hewitt, J. E., Lanzi, G., Maraschio, P. and Tiepolo, L. (1996). Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. Journal of Medical Genetics **33**: 366-370.

Tupler, R., Perini, G., Pellegrino, M. A. and Green, M. R. (1999). Profound misregulation of muscle-specific gene expression in facioscapulohumeral muscular dystrophy. Proceedings of the National Academy of Sciences, USA **96**: 12650-12654.

Vakoc, C. R., Mandat, S. A., Olenchock, B. A. and Blobel, G. A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Molecular Cell **19**: 381-391.

van der Maarel, S. M. and Frants, R. R. (2005). The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. Am J Hum Genet **76**: 375-386.

van Geel, M., Dickson, M. C., Beck, A. F., Bolland, D. J., Frants, R. R., van der Maarel, S. M., de Jong, P. J. and Hewitt, J. E. (2002). Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. Genomics **79**: 210-217.

van Geel, M., van Deutekom, J. C., van Staalduinen, A., Lemmers, R. J., Dickson, M. C., Hofker, M. H., Padberg, G. W., Hewitt, J. E., de Jong, P. J. and Frants, R. R. (2000). Identification of a novel beta-tubulin subfamily with one member (TUBB4Q) located near the telomere of chromosome region 4q35. Cytogenetics and Cell Genetics **88**: 316-321.

van Koningsbruggen, S., Straasheijm, K. R., Sterrenburg, E., de Graaf, N., Dauwerse, H. G., Frants, R. R. and van der Maarel, S. M. (2007). FRG1P-mediated aggregation of proteins involved in pre-mRNA processing. Chromosoma **116**: 53-64.

van Overveld, P. G. M., Lemmers, R., Sandkuijl, L. A., Enthoven, L., Winokur, S. T., Bakels, F., Padberg, G. W., van Ommen, G. J. B., Frants, R. R. and van der Maarel, S. M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nature Genetics **35**: 315-317.

vanDeutekom, J. C. T., Lemmers, R., Grewal, P. K., vanGeel, M., Romberg, S., Dauwerse, H. G., Wright, T. J., Padberg, G. W., Hofker, M. H., Hewitt, J. E. and Frants, R. R. (1996). Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. Human Molecular Genetics **5**: 581-590.

Vandeutekom, J. C. T., Wijmenga, C., Vantienhoven, E. A. E., Gruter, A. M., Frants, R. R., Hewitt, J. E., Padberg, G. W., Vanommen, G. J. B. and Hofker, M. H. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 Kb tandemly repeated unit. Human Molecular Genetics **2**: 2037-2042.

Wijmenga, C., Frants, R. R., Browwer, O. F., Moerer, P., Weber, J. L. and Padberg, G. W. (1990). Location of Facioscapulohumeral muscular dystrophy gene on chromosome 4. Lancet **336**: 651-653.

Wijmenga, C., Hewitt, J. E., Sandkuijl, L. A., Clark, L. N., Wright, T. J., Dauwerse, H. G., Gruter, A. M., Hofker, M. H., Moerer, P., Williamson, R., Vanommen, G. J. B., Padberg, G. W. and Frants, R. R. (1992). Chromosome-4q DNA rearrangements associated with facioscapulohumeral muscular-dystrophy. Nature Genetics **2**: 26-30.

Wijmenga, C., Padberg, G. W., Moerer, P., Wiegant, J., Liem, L., Brouwer, O. F., Milner, E. C. B., Weber, J. L., Vanommen, G. B., Sandkuyl, L. A. and Frants, R. R. (1991). Mapping of Facioscapulohumeral Muscular-Dystrophy Gene to Chromosome 4q35-Qter by Multipoint Linkage Analysis and Insitu Hybridization. Genomics **9**: 570-575.

Winokur, S. T., Barrett, K., Martin, J. H., Forrester, J. R., Simon, M., Tawil, R., Chung, S. A., Masny, P. S. and Figlewicz, D. A. (2003). Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. Neuromuscular Disorders **13**: 322-333.

Winokur, S. T., Bengtsson, U., Feddersen, J., Mathews, K. D., Weiffenbach, B., Bailey, H., Markovich, R. P., Murray, J. C., Wasmuth, J. J., Altherr, M. R.

and Schutte, B. C. (1994). The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: Implications for a role of chromatin structure in the pathogenesis of the disease. Chromosome Research **2**: 225-234.

Wolffe, A. P. and Matzke, M. A. (1999). Epigenetics: Regulation through repression. Science **286**: 481-486.

Wright, T. J., Wijmenga, C., Clark, L. N., Frants, R. R., Williamson, R. and Hewitt, J. E. (1993). Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. Human Molecular Genetics **2**: 1673-1678.

Xu, G. L., Bestor, T. H., Bourc'his, D., Hsieh, C. L., Tommerup, N., Bugge, M., Hulten, M., Qu, X. Y., Russo, J. J. and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature **402**: 187-191.

Yang, F., Shao, C. B., Vedanarayanan, V. and Ehrlich, M. (2004). Cytogenetic and immuno-FISH analysis of the 4q subtelomeric region, which is associated with facioscapulohumeral muscular dystrophy. Chromosoma **112**: 350-359.

Yant, S. R., Zhu, W., Millinoff, D., Slightom, J. L., Goodman, M. and Gumucio, D. L. (1995). High-affinity YY1 binding motifss - identification of 2 core types (ACAT and CCAT) and distribution of potential binding-sites within the human beta-globin cluster. Nucleic Acids Research **23**: 4353-4362.

Yoder, J. A., Walsh, C. L. and Bestor, T. H. (1996). Control of cytosine methylation and retrovirus silencing in early development. Molecular Biology of the Cell **7**: 47-47.

Zeng, W. H., de Greef, J. C., Chen, Y. Y., Chien, R., Kong, X. D., Gregson, H. C., Winokur, S. T., Pyle, A., Robertson, K. D., Schmiesing, J. A., Kimonis, V. E., Balog, J., Frants, R. R., Ball, A. R., Lock, L. F., Donovan, P. J., van der Maarel, S. M. and Yokomori, K. (2009). Specific Loss of Histone H3 Lysine 9

Trimethylation and HP1 gamma/Cohesin Binding at D4Z4 Repeats Is Associated with Facioscapulohumeral Dystrophy (FSHD). Plos Genetics **5**.

Zhang, X. Y., Loflin, P. T., Gehrke, C. W., Andrews, P. A. and Ehrlich, M. (1987). Hypermethylation of human DNA-sequences in embryonal carcinoma-cells and somatic tissues but not in sperm. Nucleic Acids Research **15**: 9429-9449.

Zhang, Y., Forner, J., Fournet, S. and Jeanpierre, M. (2001). Improved characterization of FSHD mutations. Annales de gentique **44**: 105-110.