



Universiteit
Leiden
The Netherlands

Genetics and epigenetics of repeat derepression in human disease

Thijssen, P.E.

Citation

Thijssen, P. E. (2016, September 1). *Genetics and epigenetics of repeat derepression in human disease*. Retrieved from <https://hdl.handle.net/1887/42675>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/42675>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/42675> holds various files of this Leiden University dissertation.

Author: Thijsen, P.E.

Title: Genetics and epigenetics of repeat derepression in human disease

Issue Date: 2016-09-01

Genetics and epigenetics of repeat derepression in human disease

Peter E. Thijssen

Lay out & cover design: Peter Thijssen

Printing: Publisher BOXPress | | Proefschriftmaken.nl

ISBN: 9789462954250

© **Copyright by Peter E. Thijssen**

All rights reserved. No parts of this thesis may be reprinted or reproduced or utilized in any form or by electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system without written permission of the author.

Publication of this thesis was financially supported by the department of Human Genetics and the Leiden University Medical Center

Genetics and epigenetics of repeat derepression in human disease

Proefschrift

ter verkrijging van de graad van Doctor aan de Universiteit Leiden
op gezag van Rector Magnificus prof.dr.C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 1 september 2016
klokke 13:45

door

Peter Evert Thijssen

geboren te Zoetermeer in 1983

Promotores:

Prof. Dr. Ir. S.M. van der Maarel
Prof. Dr. P.E. Slagboom

Leden promotiecommissie:

Prof. Dr. A.M. Aartsma-Rus
Prof. Dr. J.H.L.M. van Bokhoven¹
Prof. S. Tapscott, MD, PhD²

1: UMC Radboud, Nijmegen, The Netherlands

2: Fred Hutchinson Cancer Research Center, Seattle, WA, USA

A grayscale photograph of a river flowing through a forest. The river is in the center, surrounded by large, dark rocks. The background is filled with tall, thin trees, and the overall scene is dimly lit, creating a somber and quiet atmosphere.

- Voor papa -

Table of Contents

Chapter 1:	General introduction	9
Chapter 2:	Correlation analysis of clinical parameters with epigenetic modifications in the <i>DUX4</i> promoter in FSHD.	35
Chapter 3:	Chromatin remodeling of human subtelomeres and TERRA promoters upon cellular senescence: commonalities and differences between chromosomes.	47
Chapter 4:	Increased <i>DUX4</i> expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4.	67
Chapter 5:	<i>DUX4</i> promotes transcription of <i>FRG2</i> by directly activating its promoter in Facioscapulohumeral muscular dystrophy.	89
Chapter 6:	Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD.	105
Chapter 7:	Mutations in <i>CDCA7</i> and <i>HELLS</i> cause immunodeficiency, centromeric instability and facial anomalies syndrome.	135
Chapter 8:	General discussion	149
Chapter 9:	Appendices	171
	Summary	
	Samenvatting	
	Curriculum vitae	
	List of publications	
	Dankwoord	



General introduction

1

Epigenetic regulation of the genome

The human diploid genome consists of roughly 6.5 billion base pairs (bp), divided over 23 different chromosome pairs. This huge linear genome, and that of eukaryotes in general, is packed into the cell nucleus in a non-random and organized fashion. In order to store, maintain and use the genetic information in our genome, DNA is folded into a nucleoprotein structure called chromatin. Historically, chromatin is classified into two states: the more accessible state called euchromatin and a more inaccessible state called heterochromatin^{1,2}. Euchromatin allows the DNA to be accessed by protein machineries in the nucleus and is mainly found at actively transcribed loci. In contrast, the more inaccessible heterochromatin is mainly found at repressed and non-transcribed regions of the genome. Although chromatin organization of the genome is not static, it is mitotically heritable and is central in studying epigenetics: *“nuclear inheritance which is not based on differences in DNA sequence”*³. More specifically, epigenetics can be defined as *“the sum of alterations to the chromatin template that collectively establish and propagate different patterns of gene expression and silencing from the same genome”*¹. Thus, epigenetic regulation lies at the heart of establishing and maintaining cell identity, and is achieved by modifying and regulating the chromatin template at multiple levels.

Chromatin, histones and their post-translational modifications

The basic component of chromatin is the nucleosome: an octamer of 4 different histone proteins (H2A, H2B, H3 and H4) wrapped by ~146 bp of DNA (**Fig. 1A**). The globular domains of H2A, H2B, H3 and H4 fold into the histone octamer, whereas the more linear tails of the histone proteins are protruding out of the nucleosome (**Fig. 1A**)¹. Histone tails, and the globular domains to a lesser extent, are subject to a wide variety of post-translational modifications including, but not limited to, acetylation, methylation, phosphorylation, ubiquitylation and sumoylation, which all in some way can affect the organization and/or regulation of the chromatin template. Eu- and heterochromatin are characterized by the presence of specific patterns of histone modifications, which influence the chromatin through directly impacting chromatin structure or acting as a scaffold for regulatory proteins¹.

In general, euchromatin is characterized by high levels of acetylation on lysine residues in histones (**Fig. 1B**). The chromatin structure is directly affected by histone acetylation as it neutralizes the positive charge of lysine residues on nucleosomes, thereby interfering with the interaction of the nucleosome with negatively charged DNA and increasing the binding possibility of transcription factors to DNA⁴. Both eu- and heterochromatic regions are enriched for lysine methylation, which can have different degrees and functionalities depending on the number of methyl groups added to the substrate: mono, di or trimethylation (**Fig. 1B-D**). Both the degree of methylation and the specific histone tail residue used as substrate are associated with different chromatin contexts. For example, active promoters are typically marked by high levels of Histone 3 lysine 4 di- and tri-methylation (H3K4me2/3) (**Fig. 1B**), whereas long distance enhancers are usually marked by H3K4me1⁵. Methylation of H3K4 is thus considered

to mark euchromatin. In contrast, methylation of H3K9 and H3K27 is typically found at heterochromatin, which can be further subdivided in constitutive and facultative heterochromatin. Constitutive heterochromatin, marked by high levels of H3K9me₃, is gene poor, often repetitive in nature and silenced in all somatic cell types (**Fig. 1C**)². Facultative heterochromatin, enriched for H3K27me₃, is often found at gene bodies which need to be transcriptionally silenced in specific cell types or during development and is considered to be more plastic of nature (**Fig. 1D**)².

Histone marks are established, recognized and removed by so called “writer”, “reader” and “eraser” proteins, respectively. Acetylation of histone is catalysed by histone acetyl transferases (HATs) and can be subsequently removed by histone deacetylases (HDACs). Both HATs and HDACs are subdivided into different subclasses based on domain organization of the proteins and substrate specificity¹. Histone acetylation is “read” by proteins containing a bromodomain (BrD), which is found in at least 41 human proteins. Among these 41 proteins are transcription factors, chromatin remodelers and HATs, of which the latter create a positive feedback loop where histone acetylation leads to more histone acetylation (**Fig. 1B**)^{6,7}.

Methylation and demethylation of histones is carried out by different lysine methyl transferases (KMTs) and lysine specific demethylases (KDMs), respectively, which are non-redundant in target residues and degrees of methylation. All except one member of the large group of KMTs contain a SET domain, which catalyses lysine methylation⁸. Different KMTs have different substrate specificity: methylation of H3K4, for example, can be carried out by mixed lineage leukaemia (MLL) proteins, whereas H3K36 methylation is mainly catalysed by SET2 (**Fig. 1B**)⁸. H3K9 methylation can be catalysed by different KMTs, including suppressor of variegation 3-9 homologue 1 (*SUV39H1*) and *SUV39H2* (**Fig. 1C**). Two major H3K27 KMTs are identified to date: enhancer of zeste homologue 1 (*EZH1*) and *EZH2*, both only active in the context of the multi subunit Polycomb repressive complex 2 (PRC2) (**Fig. 1D**)⁸.

Lysine methylation can be “read” by a versatile group of protein domains, including the PHD zinc finger and the chromodomain⁹. As for acetylation, “reading” methylation can create a positive feedback loop. H3K9me creates a binding site for the chromodomain of Heterochromatin protein 1 (HP1) which recruits the H3K9me “writer” SUV39H1 (**Fig. 1C**)^{10,11}. Similarly, the WD40 domain of the PRC2 component embryonic ectoderm development (EED) binds H3K27me₃ and thereby promotes more H3K27me₃ (**Fig. 1D**)^{8,12}.

Next to the establishment of positive feedback loops, “Reader” proteins are also central to the concept of crosstalk between different histone modifications. At euchromatin, for example, the chromodomain of HDAC1, which travels with the transcriptional machinery, binds SET2 mediated H3K36me₃ and leads to histone deacetylation in transcribed gene bodies (**Fig. 1B**)¹³⁻¹⁵. In heterochromatin, PRC2 mediated H3K27me₃ is “read” by the PRC1 complex, which further promotes chromatin compaction and silencing through H2AK119 mono-ubiquitylation (H2AK119Ub) (**Fig. 1D**)¹⁶. In both examples, “reading” of methylation marks leads to the removal or deposition of different modifications,

creating another layer of regulatory complexity on the chromatin template. Altogether, the dynamic nature of histone modifications, their ability to act as a docking platform for effector proteins and their potential crosstalk creates a potent mechanism to organize, maintain and employ the large amount of genetic information in the context of the chromatin template.

Epigenetic regulation on the DNA backbone: CpG methylation

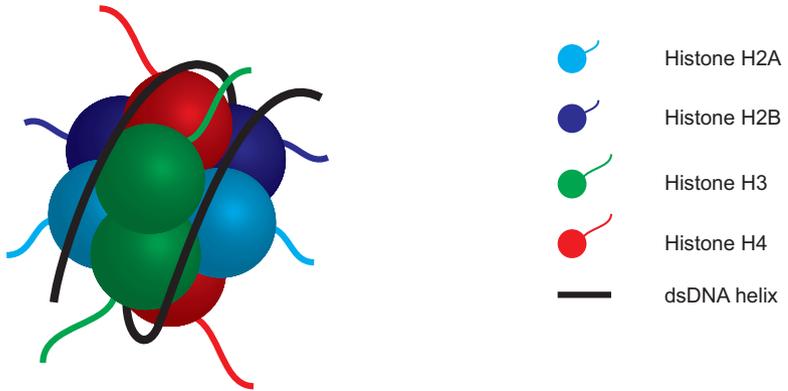
The regulation of chromatin structure is not limited to the modification of histone proteins. In fact, the DNA backbone can be subject to methylation, which affects gene expression and chromatin organization. In mammals, CpG dinucleotides form the main substrate for cytosine methylation¹⁷. CpGs are found dispersed throughout the genome as single CpGs, or as clustered CpG islands (CGIs) in gene promoters. In general, single CpGs throughout the genome are methylated, whereas the majority of CGIs are unmethylated (**Fig 1B-D**)¹⁷. As for histone modifications, the human genome also encodes “writers”, “readers” and “erasers” of DNA methylation to ensure proper regulation and interpretation of this mark.

Methylation of CpGs is “written” by DNA methyltransferases (DNMTs). *DNMT1* primarily acts on hemi-methylated DNA and thereby is pivotal for maintaining CpG methylation patterns during DNA replication^{17, 18}. *DNMT1* is targeted to DNA replication foci by its interaction with proliferating cell nuclear antigen (PCNA). Specific targeting of *DNMT1* to heterochromatic regions is dependent on the H3K9me machinery. *DNMT1* interacts with ubiquitin-like, containing PHD and Ring finger domains 1 (UHRF1), which binds H3K9me3 through its PHD finger, and with H3K9 KMTs directly (**Fig. 1C**)¹⁷. Binding through UHRF1 is mediated by ubiquitylation of H3K23, another example of crosstalk between epigenetic marks.

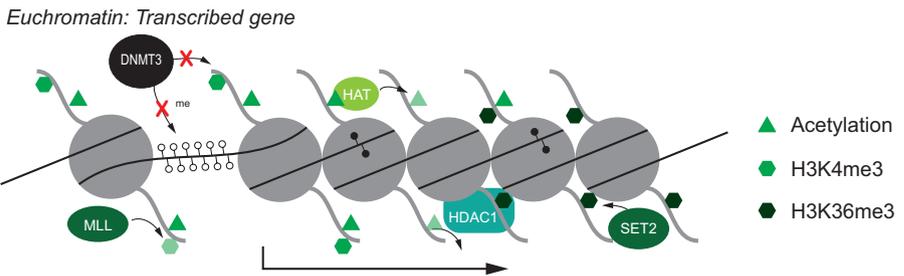
Figure 1: Schematic representation of histone proteins, chromatin and chromatin modifications

A) The double stranded DNA helix (thin black line) wraps itself around an octamer of 4 histone proteins -H2A (cyan), H2B (dark blue), H3 (green) and H4 (red)- to form the nucleosome, the basic component of the chromatin template. The linear tails of the histone proteins, subject to a wide variety of post-translational modifications, are protruding out of the nucleosome. **B**) schematic representation of euchromatin at actively transcribed regions. Euchromatin is generally characterized by high levels of histone acetylation (green triangles) and trimethylation of H3K4 and H3K36 (green hexagons). MLL proteins trimethylate H3K4, whereas H3K36 is mainly methylated by SET2. Histone acetylation is “written” and “erased” by HATs and HDACs respectively. Active gene expression, indicated by the arrow, associates with CpG island hypomethylation, as H3K4 methylation inhibits de novo CpG methylation by DNMT3. **C**) Regions of constitutive heterochromatin are generally characterized by high levels of H3K9me3 and CpG methylation. HP1 proteins can bind H3K9me3 and recruit the SUV39H1 methyltransferase, creating a positive feedback loop. CpG methylation is “read” by i.a. MeCP2 which promotes heterochromatin formation by recruitment of HDACs. Upon DNA replication, *DNMT1* is localized to sites of heterochromatin through UHRF1 in order to maintain methylation levels. **D**) Polycomb repressive complexes 1 and 2 play a major role in silencing gene expression at facultative heterochromatin. PRC2 catalyses H3K27me3 (yellow hexagons) which is “read” by PRC1 to establish H2AK119Ub (red circles) which further compacts the chromatin. TET enzymes, not necessarily at facultative heterochromatin, catalyse active demethylation of meCpG through a series of oxidative reactions.

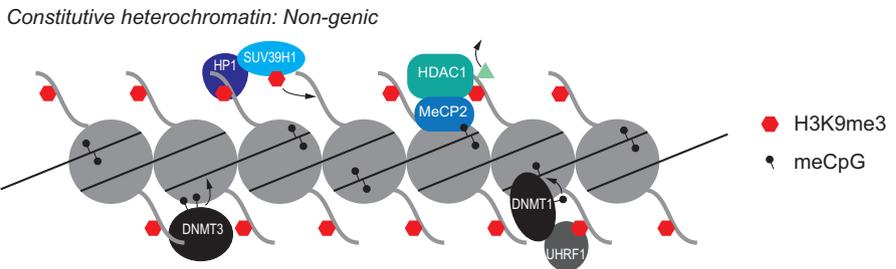
A



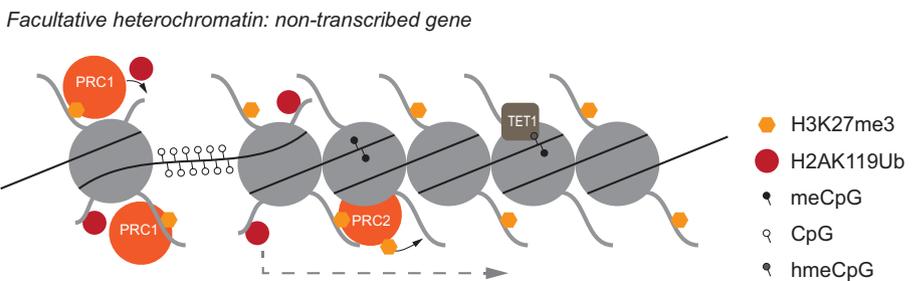
B



C



D



DNMT3A and *DNMT3B* encode de novo methyltransferases which, together with the non-catalytic *DNMT3L* co-factor, establish the genome wide pattern of DNA methylation during early development^{17, 19}. Establishment of DNA methylation in mammals is at least in part dependent on crosstalk with histone modifications (or the lack thereof). *DNMT3A/B* enzymes contain an ATRX-DNMT3-DNMTL (ADD) domain which efficiently binds unmethylated H3K4¹⁷. However, H3K4me3, highly enriched at promoters of actively transcribed genes, inhibits binding of DNMT enzymes and as a consequence promoter CGIs are protected from de novo methylation (**Fig. 1B, D**). In contrast, methylation of H3K9 has a strong positive correlation with CpG methylation. At a subset gene promoters, which are silenced during differentiation, H3K9me (in) directly recruits *DNMT3A* and/or *DNMT3B* and thereby promotes CpG methylation. De novo methylation at sporadic, non-genic CpG sites can occur either dependent or independent of H3K9me machineries, reliant on the genetic context. CpG methylation at these sites is an important mechanism to maintain genomic integrity and preserve the heterochromatic conformation of non-transcribed loci¹⁷.

DNA methylation can be read primarily by proteins containing a methyl binding domain (MBD), which was first identified in methyl-CpG binding protein 2 (*MeCP2*)²⁰. *MeCP2*, as well as other MBD containing proteins, interacts with HDACs and KMTs to maintain a heterochromatic structure and thereby bridges two layers of epigenetic regulation (**Fig. 1C**)²⁰. A possible direct link between CpG methylation and repressive histone methylation exists through SET domain and bifurcated 1 (*SETDB1*) and *SETDB2*, two H3K9 KMTs that have a putative MBD²⁰.

More recently, a class of enzymes was discovered that can “erase”, or better “edit”, CpG methylation. Active removal of CpG methylation is carried out through stepwise oxidation of the methyl group to hydroxymethyl, formyl and carboxyl which finally can be removed and subsequently repaired. This oxidation, and removal of mCpG, is carried out by ten eleven translocation 1 (*TET1*), *TET2* and *TET3* proteins (**Fig. 1D**)²¹. Next to active removal of CpG methylation, *TET* enzymes create another layer of possible epigenetic regulation: the intermediates formed by the *TET* enzymes may have biological roles themselves²². In support of this, for example, is the observed stable and persistent enrichment of hydroxymethylation at euchromatic regions in cells of the neuronal lineage, which positively correlates with gene expression²². In summary, CpG methylation is established and maintained by DNMTs, interpreted by MBD containing proteins and removed by *TET* enzymes. It correlates with histone modification patterns and together these epigenetic systems dictate the organization of the chromatin template and create a platform to maintain and use genetic information in order to establish heritable patterns of gene expression, which identify cell identity.

Epigenetics and disease

The establishment of stable and heritable patterns of gene expression ensures cell, tissue and organ homeostasis. Therefore, epigenetic dysregulation of the genome is an important risk factor for the development of disease. Indeed, the dysregulation of the epigenome is one of the hallmarks of cancer cells, which generally display hypomethylation of sporadic CpGs, hypermethylation of hundreds of promoter CpG islands and disturbed patterns of histone modifications^{23, 24}. Changes in the epigenetic regulation of the genic part of genome in cancer cells can lead to the activation of oncogenes and/or the silencing of tumor suppressors. Moreover, the globally unbalanced epigenome is believed to result in higher genomic instability, another hallmark of cancer cells²⁴.

Next to cancer, various classes of epigenetic diseases have been recognized, among which imprinting disorders are the classic example. Imprinting is an epigenetic process leading to mono-allelic expression depending on parental origin of a substantial group of human genes and is primarily mediated by epigenetic regulation *in cis* on several levels. Genetic or epigenetic disruption of these imprinted regions leads to aberrant expression of the imprinted genes (biallelic expression or absence of expression) and can lead to human disease^{25, 26}. For example, Beckwith-Wiedemann syndrome (BWS), characterized by overgrowth, and Silver-Russell syndrome (SRS), characterized by undergrowth and asymmetry, both map to an imprinted region on chromosome 11p15. Opposite incorrect epigenetic regulation of the loci that control the imprinting of this imprinted region leads to either increased paternal or maternal expression of the imprinted genes, leading to BWS or SRS respectively^{25, 26}.

Imprinting disorders belong to the group of *in cis* epigenetic disorders, where local changes in the chromatin organization lead to human disease. Several *in cis* epigenetic disorders are known in which non-imprinted loci are involved. For example, genetic mutations in the fragile X mental retardation 1 (*FMR1*) gene lead to the neurodegenerative FXTAS disorder or fragile X syndrome, depending on the type of mutation²⁷. In both cases, a trinucleotide repeat in the 5' untranslated region (UTR) of *FMR1* is expanded to either a pre-mutation allele (55-200 copies, FXTAS) or a full mutation allele (>200 copies, fragile X syndrome)²⁷. The pre-mutation allele leads to transcriptional activation, presumably because the expansion results in the formation of a larger promoter region. Full mutation alleles, on the contrary, result in transcriptional repression of the *FMR1* gene by the recruitment of repressive complexes that silence the locus²⁷. The expanded repeat thus acts *in cis* to control the levels of transcription through epigenetic mechanisms.

The example of fragile X syndrome shows that a gene mutations can have an epigenetic effect *in cis* which leads to disease. The list of disorders where genetic mutations lead to an epigenetic phenotype *in trans* is considerably larger. Mutations in numerous “writers”, “readers” and “erasers” have been identified to underlie syndromes, often characterized by developmental problems and intellectual disability²⁶. An intriguing example of an *in trans* disorder is Kabuki syndrome, characterized by intellectual

disability, facial dysmorphisms and short stature. Kabuki syndrome is caused by mutations in *MLL2* or *KDM6A*, an H3K4 KMT and H3K27 KDM respectively^{28, 29}. By modulating lysine methylation on histones *MLL2* promotes chromatin relaxation whereas *KDM6A* inhibits chromatin repression. This essentially results in the same: a shifted balance of gene expression at target genes of these machineries, which is supported by the indistinguishable phenotype of both patient groups²⁶.

All the above shows that faithful epigenetic regulation of genome is pivotal for cell homeostasis and that disruptions in this system, globally and locally, can result in human disease. In general, studies focus on the effect of epigenetic dysregulation on the genic compartment of the genome. Since the great minority of the human genome is actually protein coding, the effect on non-coding genomic regions should not be underestimated.

The repetitive genome: expand and silence.

With the completion of the human genome project at the beginning of the century, early estimates of the total number of genes in the human genome (around 100.000) were proven wrong³⁰. In fact, the latest numbers indicate that the human genome contains less than 25.000 genes. Compared to the number of genes identified in lower, less complex, eukaryotes like *Saccharomyces cerevisiae* and *Drosophila melanogaster*, it becomes clear that increased organismal complexity does not solely depend on the number of genes (**Fig. 2A**)^{31, 32}. Besides, the number of identified genes in these organisms does not reflect the size of their genomes (**Fig. 2A**). In other words: the complexity of human life, compared to that of budding yeast or fruit fly, cannot be simply explained by an increase in the number of genes.

Rather than the coding part of the genome, the steep increase in non-coding DNA sequences underlies the dramatic expansion of the human genome compared to that of other eukaryotes. With increasing genome sizes in *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Homo sapiens*, there is a concomitant decrease in the percentage of protein coding basepairs (**Fig. 2B-D**)³³. The vast majority of the human genome, more than 97%, is actually non-protein coding and was referred to as non-functional “junk DNA”³⁴. The publication of the “Encyclopedia of DNA Elements” (ENCODE), however, revealed that many of these junk DNA regions are actually functional, e.g. by acting as distant gene expression enhancer sites, and contribute to the regulation of gene expression patterns³⁵. The increase in non-coding regulatory elements thus creates an additional layer of transcriptional regulation and thereby contributes to organismal complexity.

Apart from the size of the genome, the fraction of repetitive DNA positively correlates to organismal complexity (**Fig. 2B-D**). Up to 45% of the non-coding part of the human genome is repetitive of nature and is typically packed into constitutive heterochromatin. Repeated sequences include large stretches (10-300 kb) of duplicated sequence blocks known as segmental duplications. However, the majority of repetitive DNA is comprised of two main classes of highly repetitive elements: interspersed and tandem repeats³⁶.

Interspersed repeats, including long/short interspersed nuclear elements (LINE/SINE) and long terminal repeats (LTRs), are viral DNA elements which have covered the human genome by retrotransposition and account for 90% of all repetitive elements in the human genome³⁶. Tandem repeats, organised in a head to tail fashion, are polymorphic in length and further classified according to the size of the repetitive unit. Microsatellite repeats, or short tandem repeats (STR), have a repetitive unit of 1-7 bp long and can span up to several hundreds of basepairs³⁶. Telomeric repeats, as well as some centromeric satellite repeats, fall into this category. Minisatellites have a repeat unit size between 8 and 100 bp and are typically found near centromeres and telomeres. Micro- and mini-satellites are often used for DNA fingerprinting in forensic DNA analyses. Macrosatellites are at least 100 bp, but usually several kb per unit and can span up to several megabases in total length^{36, 37}. In total, the repetitive genome comprises a significant proportion of the human genome and in majority has to be in

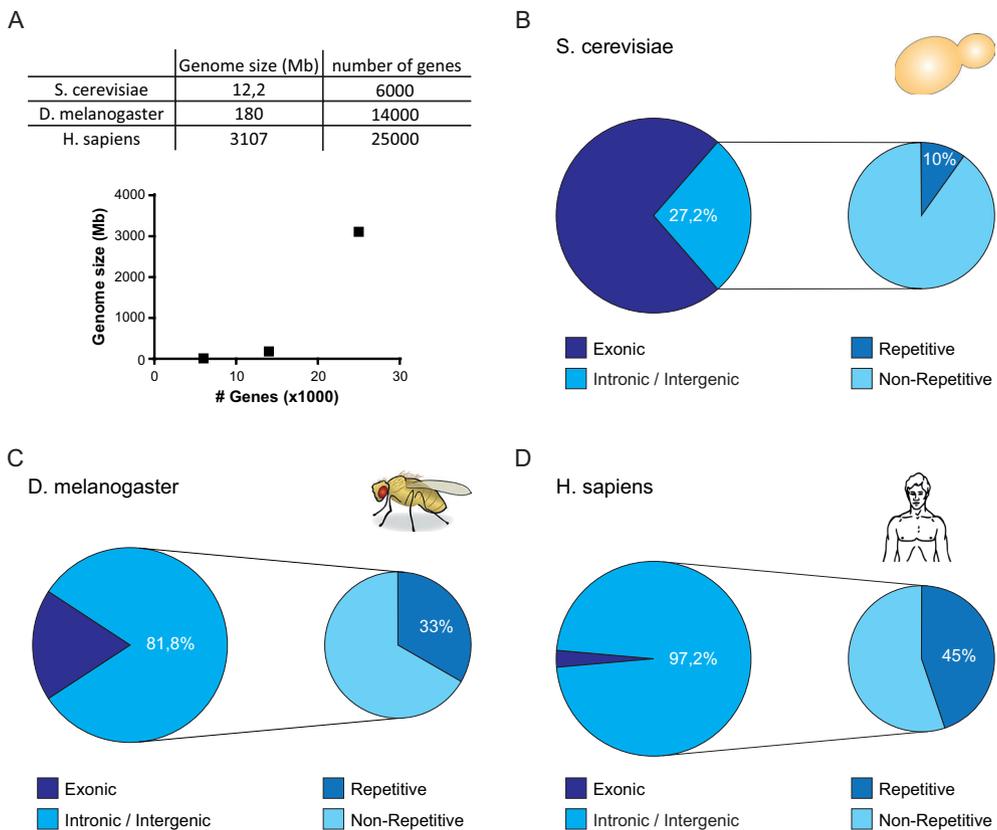


Figure 2: The size of the genome and the fraction of repetitive DNA correlate with organismal complexity.

A) The genome size, rather than the number of encoded genes, correlates with increasing organismal complexity in *S. cerevisiae*, *D. melanogaster* and *H. sapiens*. **B-D)** The relative amount of non-coding DNA (intronic and intergenic) and the relative distribution between non-repetitive and repetitive DNA in *S. cerevisiae* (**B**), *D. melanogaster* (**C**) and *H. sapiens* (**D**) shows a correlation between organismal complexity and the amount of repetitive DNA.

a repressed chromatin conformation in order to maintain genome stability and silence transcription of repeats.

This thesis focuses on the genetic and epigenetic features of facioscapulohumeral muscular dystrophy (FSHD) and immunodeficiency, centromere instability and facial anomalies (ICF) syndrome: epigenetic disorders in cis and in trans, respectively. Common to both diseases is the epigenetic dysregulation of repetitive DNA. In FSHD this is most often confined to the D4Z4 macrosatellite repeat, whereas in ICF syndrome the epigenetic dysregulation of repeat sequences occurs genome wide, including D4Z4 and centromeric satellites. Both disorders will be further introduced below.

FSHD: derepression of a macrosatellite repeat

FSHD (OMIM 158900/158901) is a progressive muscular dystrophy first described by Landouzy and Dejerine, with recent estimates to affect approximately 1 in 8000 individuals^{38, 39}. Patients suffer from progressive weakening of the facial, shoulder and proximal limb muscles and often show asymmetric involvement of muscles⁴⁰. With disease progression, also other muscles may become affected. FSHD mostly shows an age at onset in the second decade of life, but is however characterized by a high inter- and intra-familial variability in onset, progression and severity⁴⁰. Extreme cases show muscle weakness at birth, whereas some individuals remain asymptomatic throughout life. Eventually, 20% of FSHD patients above the age of 50 years become wheelchair bound. A minority of patients shows respiratory and cardiac involvement (atrial arrhythmia), of which the latter is rarely symptomatic. Extra muscular symptoms have been reported and mainly involve retinal vasculopathy and progressive hearing loss⁴⁰.

FSHD is linked to the subtelomeric D4Z4 repeat on chromosome 4q35

In most cases, FSHD is inherited in an autosomal dominant manner, with a high frequency (10-30%) of de novo cases^{39, 41}. In the early nineties, linkage analysis revealed that FSHD segregates with marker loci in the subtelomere of chromosome 4q35, which harbours the D4Z4 macrosatellite repeat (**Fig. 3A**)⁴²⁻⁴⁴. Each D4Z4 repeat unit is 3.3 kb in size and the number of repeats per allele is highly polymorphic. The D4Z4 array consists of 1 to over 100 units, leading to a possible size difference of more than three megabasepairs between individual alleles (**Fig. 3A**)³⁷. The 4q subtelomere exists in two equally frequently occurring haplotypes (4qA and 4qB), and FSHD uniquely associates with the A variant⁴⁵⁻⁴⁸. Using restriction enzyme analysis, it was found that partial deletion of D4Z4 on 4qA alleles, resulting in a repeat array of less than 11 but more than 1 units, leads to the development of FSHD type 1 (FSHD1)⁴⁹⁻⁵². The number of residual repeats shows a rough positive correlation with age at onset and wheelchair use^{53, 54}. Only contraction of 4qA alleles is pathogenic since D4Z4 repeat arrays of less than 10 units in the control population can be observed on 4qB chromosomes⁴⁶.

The contraction of the D4Z4 repeat is diagnostic for the vast majority of FSHD patients. However, a small remaining group of patients, classified as FSHD2, shows an indistinguishable phenotype, but carries a D4Z4 repeat in the lower size range of control individuals^{55, 56}. As seen for FSHD1, the disease relies on the presence of the 4qA

haplotype, as all FSHD2 patients carry at least one such allele⁵⁵.

The D4Z4 macrosatellite repeat is located in the subtelomere of chromosome 4q35, which is immediately adjacent to the intact telomeric [TTAGGG] repeats. Subtelomeres are characterized by the presence of repetitive DNA and segmental duplications and are packed in a constitutive heterochromatin structure like the adjacent telomeres^{57, 58}. Subtelomeric segmental duplications have occurred both intra- and inter-chromosomally and the duplicons can also be identified in non-subtelomeric regions of the genome, such as pericentromeres⁵⁷. Indeed, the subtelomere of chromosome 4q, including the D4Z4 repeat array, is duplicated to the subtelomere of chromosome 10q (**Fig. 3A**), but contractions of the 10q copy of D4Z4 are typically not pathogenic^{59, 60}. Additionally, single, often incomplete, D4Z4 copies can be found dispersed throughout the genome, but were never linked to pathogenicity⁶¹⁻⁶³.

Together, genetic analyses put the partial deletion of the D4Z4 macrosatellite repeat at 4q35 at the centre of FSHD pathology. Each D4Z4 unit encodes a copy of the *DUX4* retrogene, a member of the double homeobox transcription factor gene family which has only been identified in placental mammals⁶⁴. *DUX4* is most likely a retrotransposed copy of the ancestral and intron containing *DUXC* gene which is lost in the primate lineage^{64, 65}. *DUX4* does not have a rodent orthologue, however a paralogue has been identified: the rodent specific *Dux* array identified in mouse and rat suggests divergent evolutionary events leading to conservation of a tandemly repeated *Dux* gene⁶⁴⁻⁶⁶. Remarkably, the organization of *DUXC/DUX4/Dux* like genes into a tandem repeat array is conserved in mammals⁶⁶. By ectopic expression, *DUX4* was found to be a pro-apoptotic protein and an inhibitor of muscle cell differentiation, however its expression or dysregulation in FSHD muscle could for a long time not be established⁶⁷⁻⁷⁰. The non-detectable dysregulation of *DUX4*, together with the fact that only partial deletion of the heterochromatic D4Z4 repeat causes FSHD, suggested an epigenetic component in FSHD disease aetiology⁵⁶.

The complex interplay of chromatin regulators at D4Z4

The D4Z4 repeat, as most macrosatellites, is transcriptionally silenced and organized into heterochromatin in somatic cells. D4Z4, characterized by a high density of CpG dinucleotides, is highly but inhomogeneously methylated and it is marked by H3K9me3 in somatic cells, consistent with its heterochromatic nature (**Fig. 3B**)^{55, 56, 71-75}. Remarkably, histone markers for euchromatin (acetylation), as well as facultative heterochromatin (H3K27me3) were also found to be enriched at D4Z4 (**Fig. 3B**)^{75, 76}. In FSHD individuals, the chromatin organization is disrupted in somatic cells as CpG methylation levels are reduced at D4Z4 (**Fig. 3B**)^{55, 56, 72}. Moreover, using primary myoblasts and fibroblasts, it was shown that H3K9me3, mediated by SUV39H1, is decreased at D4Z4 in FSHD patient derived cell lines compared to healthy controls or patients suffering from other muscular dystrophies (**Fig. 3B**)⁷⁵. Furthermore, the downstream “readers” of H3K9me3, HP1 γ and Cohesin, were shown to be reduced at D4Z4⁷⁵. Together this shows that the heterochromatin organization at D4Z4 is disrupted in patients, leading to partial relaxation of the locus.

The epigenetic changes observed at D4Z4 are common to FSHD1 and FSHD2. In fact, in FSHD2 individuals the chromatin changes are observed on both D4Z4 repeat arrays on chromosome 4 as well as the 10q copies, whereas in FSHD1 individuals the effects are restricted to the contracted pathogenic repeat^{55, 56, 75}. FSHD1 is an *in cis* epigenetic disorder: the contraction of the repeat leads to a change in local chromatin structure, similar to fragile-X syndrome. In contrast, FSHD2 is an *in trans* epigenetic disorder as >80% of the FSHD2 patients carry mutations in the structural maintenance of chromosomes flexible hinge domain containing 1 (*SMCHD1*) gene, which is underlying the changes in D4Z4 chromatin structure⁷⁷.

SMCHD1 is structurally related to the SMC protein superfamily, which constitutes core proteins of the Cohesin complex, and was first identified in a screen to identify epigenetic modifiers of variegated expression in a murine model⁷⁸. *Smchd1* has been shown to play a role in X-chromosome inactivation, an epigenetic process ensuring dosage compensation in females by silencing one of the two X chromosomes. A hallmark of X-chromosome inactivation is the expression of a long non-coding RNA (lncRNA) known as *Xist*. *Xist* covers the X-chromosome *in cis* and recruits the PRC2 complex to ensure gene silencing throughout the inactive X-chromosome, with the exception of some genes that escape this process⁷⁹. In female *Smchd1* knockout mice, X-chromosome inactivation is perturbed, with promoter hypomethylation of CpG islands and concomitant upregulation of clustered transcripts normally subject to X-chromosome inactivation, showing a role for *Smchd1* in establishment and/or maintenance of CpG methylation^{78, 80-82}. Furthermore, it has been shown that SMCHD1 is involved in the higher order compaction of the inactivated X-chromosome by interacting with *Xist* and H3K27me3⁸². Next to its role in X-chromosome inactivation, *Smchd1* is involved in the silencing of several mono-allelically expressed autosomal genes, among which the clustered protocadherin genes on mouse chromosome 18^{81, 83}.

In concordance with all these observations, reduced binding of SMCHD1 at D4Z4 correlates with CpG hypomethylation and chromatin derepression in FSHD2 patients (Fig. 3B)⁷⁷. Moreover, SMCHD1 was shown to act as a modifier of disease severity in FSHD1 patients, supporting a role for SMCHD1 in both genetic forms of the disease^{84, 85}. Expression of both long and small non-coding RNAs from D4Z4 have been reported and linked to chromatin repression and/or activation^{86, 87}. A lncRNA starting proximal to the D4Z4 repeat was shown to recruit the chromatin modifier ASH1L, an H3K36 KMT normally associated with euchromatin, resulting in derepression of *DUX4*⁸⁶. Conversely, expression of several different small interfering RNAs (siRNAs) matching the D4Z4 repeat sequence led to repression of D4Z4 in a DICER/AGO dependent fashion⁸⁷. Altogether, a complex interplay of different mechanisms regulating the compaction of chromatin has been shown to act at the D4Z4 macrosatellite repeat, highlighting the epigenetic component of FSHD.

D4Z4 chromatin changes in FSHD lead to the derepression of DUX4

In absence of evidence for *DUX4* expression in FSHD muscle cells, early studies proposed that the changed local chromatin environment at D4Z4 had an effect *in cis* on proximal

genes. This hypothesis relied either on proximal spreading of the altered chromatin structure at D4Z4 and/or changes in higher order chromatin organization and long range interactions. 4q35, in contrast to 10q26, preferentially localizes to the nuclear periphery⁸⁸. This is likely mediated through interactions with the nuclear matrix, which is disturbed upon D4Z4 contraction⁸⁹. Next to disturbed interactions with the nuclear matrix, D4Z4 contractions also lead to an altered higher order chromatin structure at

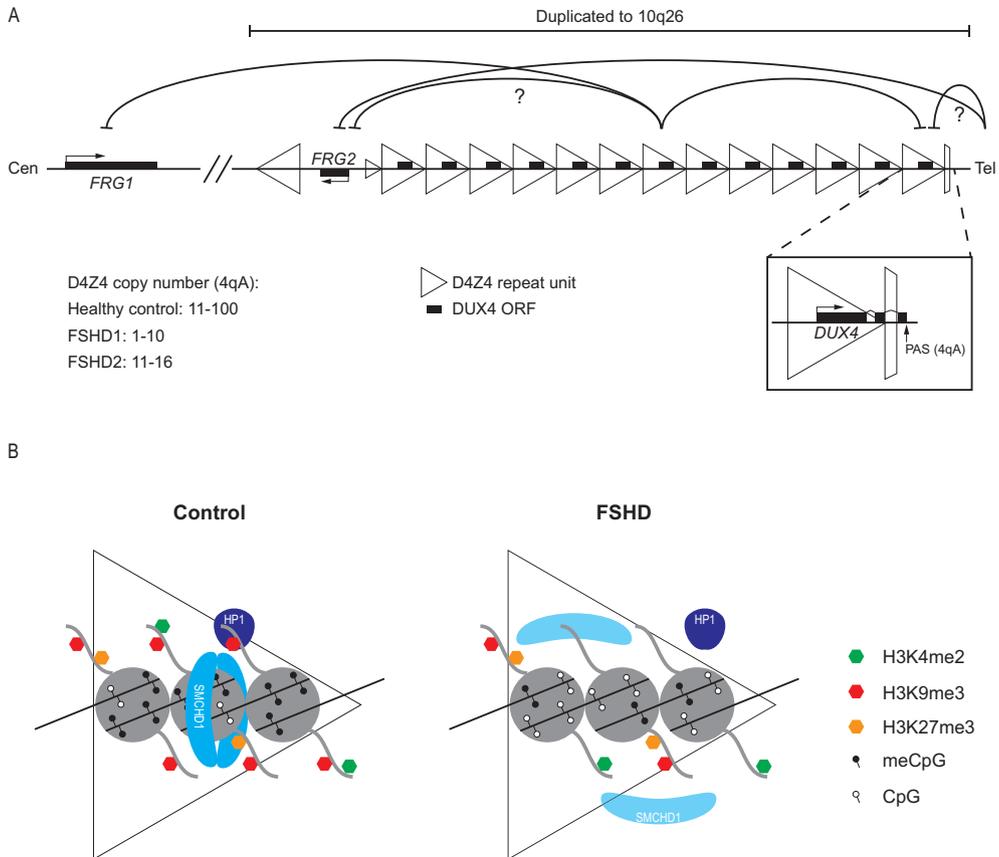


Figure 3: schematic representation of the genetic and epigenetic features of the FSHD locus on chromosome 4q35.

A) The D4Z4 macrosatellite (triangles) and its flanking sequences, including the proximal *FRG1* and *FRG2* genes, is contracted in FSHD1 patients. All patients (FSHD1 and FSHD2) carry the 4qA variant of the p-LAM sequence element distal to the D4Z4, encoding a non-canonical poly adenylation signal allowing the formation of stable DUX4 transcripts. The D4Z4 repeat and some flanking sequences are duplicated to chromosome 10q26. Arches depict the reported long range interactions and/or position effects of the chromatin affected in FSHD. Inset: overview of the DUX4 transcript produced from the most distal D4Z4 unit and the p-LAM sequence. The full *DUX4* open reading frame is included in the first exon and is therefore present in each repeat unit. B) The D4Z4 chromatin structure is characterized by the presence of methylation markers for both eu- and hetero-chromatin (hexagons) and high levels of CpG methylation. D4Z4 compaction is further established by binding of HP1 and SMCHD1. In FSHD patients, D4Z4 is decompacted evidenced by a loss of CpG methylation and H3K9me3 with a concomitant loss of SMCHD1 and HP1 binding.

4q35^{89, 90}. Furthermore, D4Z4 was reported to physically interact with more proximal regions by which it may influence the local chromatin structure of more upstream genes on 4q35^{91, 92}.

Two candidate genes were identified in the region flanking D4Z4 proximally: FSHD region gene 1 (*FRG1*) and *FRG2*, of which the latter is also present on chromosome 10 (**Fig. 3A**)^{93, 94}. Both *FRG1* and *FRG2* were reported to be upregulated in FSHD, suggesting a mechanism of long range interactions and/or spreading of chromatin derepression from the D4Z4 repeat (**Fig. 3A**)^{63, 93, 95-97}. Overexpression of *FRG1*, an Actin bundling and mRNA processing protein, induces a muscular dystrophy phenotype in different animal models, however most follow up studies failed to confirm the upregulation of *FRG1* in FSHD patient material⁹⁶⁻¹¹². In contrast, *FRG2* activation is a robust and reproducible hallmark of FSHD cells, however its function is unknown and overexpression in mice did not lead to a muscle phenotype^{93, 97, 101}.

More recently, few studies revealed a possible link between deregulation of FAT atypical cadherin 1 (*FAT1*), located 3.6 Mb upstream of D4Z4, and FSHD pathology¹¹³⁻¹¹⁵. Mice in which *Fat1* was genetically ablated developed asymmetric muscle wasting reminiscent of FSHD, and genetic analysis of *FAT1* in human patients may suggest a secondary or indirect involvement of *FAT1* in FSHD pathology¹¹³⁻¹¹⁵. Altogether, these studies highlight that genes proximal to the D4Z4 repeat may be deregulated in FSHD. Their involvement in the disease mechanism and the mechanisms behind their deregulation remain unclear at this point.

With the identification of several transcripts produced from D4Z4, including an mRNA encoding the full length DUX4 protein, efforts to identify the FSHD disease mechanism focused on *DUX4* again¹¹⁶. The key to the FSHD disease mechanism lies within the unique association of FSHD with the 4qA haplotype distal to the D4Z4 repeat⁴⁵⁻⁴⁷. This sequence element (pLAM) encodes 1) an additional *DUX4* exon distal to the last repeat unit and 2) a non-canonical *DUX4* polyadenylation signal (PAS). Both these elements are absent in 4qB alleles, while the PAS is absent from 10q alleles^{68, 117}. It was shown that the presence of this PAS can lead to the formation of a stable full length DUX4 transcript and genetically unifies all FSHD patients (**Fig. 3B**)¹¹⁷. Additionally, full length DUX4 was shown to be abundantly expressed in sporadic myonuclei of FSHD derived proliferating myoblasts and differentiated myotubes, but at low or even undetectable levels in control derived material^{118, 119}.

These combined efforts have led to a unifying disease mechanism in which developing FSHD relies on three interdependent prerequisites:

1. *the presence of at least one PAS containing 4qA allele (contracted to 1-10 units in FSHD1);*
2. *chromatin derepression at D4Z4 through an in cis (FSHD1) or in trans (FSHD2) mechanism;*
3. *sporadic DUX4 expression in a myogenic context.*

Upon expression, *DUX4* acts as a potent transcriptional activator. In muscle cells, *DUX4* activates a specific set of genes, through direct binding to a double homeodomain DNA motif¹²⁰. In response to *DUX4* activation, genes involved in germline biology, early stem cell development and innate immunity are deregulated¹²⁰. Furthermore, *DUX4* binds and activates retroelements, mainly of the ERV/MaLR type, which can lead to the formation of alternative transcriptional start sites for flanking genes^{120, 121}. Many of the *DUX4* targets identified by overexpression are deregulated in patient derived material, including fetal muscles, and account for the majority of transcriptional changes between FSHD and control muscle cells and/or biopsies^{112, 120, 122}. More recently, a reporter based approach, allowing transcriptome analysis of individual muscle cells expressing endogenous *DUX4*, confirmed many of these targets and highlighted a role for disrupted RNA metabolism in FSHD pathology¹²³.

Apart from initiating aberrant transcriptional programs in muscle cells, *DUX4* expression has other detrimental effects in various model systems which may or may not depend on its function as a transcriptional activator. In rhabdomyosarcoma cells it was shown that *DUX4* induces cell cycle arrest in a P21 dependent fashion, possibly impacting muscle regeneration¹²⁴. Besides, *DUX4* expression in murine ES cells leads to reduced pluripotency and an imbalance in the formation of the three germ layers upon differentiation¹²⁵. Expression of *DUX4* in mesenchymal stromal cells promotes their differentiation into osteoblasts by an unknown mechanism¹²⁶. *DUX4* was also shown to inhibit RNA degradation through nonsense mediated decay (NMD). Expression of *DUX4* leads to the degradation of the NMD factor UPF1, thereby creating a positive feedback loop as *DUX4* itself is a substrate for NMD¹²⁷. Altogether, this likely accounts for the observed toxicity of *DUX4* expression in skeletal muscle cells.

Although the consequences of *DUX4* expression are extensively studied, the mechanisms underlying the sporadic activation of *DUX4* are not so well understood. What is driving the sporadic bursts of expression? Why is *DUX4* expression increased during myotube differentiation? Few studies have focused on these aspects and how the sporadic expression of *DUX4* leads to the progressive and variable phenotype characteristic of FSHD. For example, *DUX4* activation is repressed by active Wnt/ β -catenin signalling¹²⁸. Next to that, *DUX4* activation was linked to the activity of two enhancers proximal to D4Z4, which show myogenic activity in controls and patients¹²⁹. *DUX4* was also shown to be controlled by a telomere position effect (TPE), a chromatin mediated regulation similar to what was proposed for proximal gene regulation by D4Z4 (**Fig. 3A**)¹³⁰. It was shown that the expression levels of *DUX4*, as well as *FRG2*, inversely correlate with the length of the adjacent 4q telomere. As telomere length naturally declines with (cellular) age and was shown to influence the epigenetic regulation of the adjacent subtelomeres^{131, 132}, this study possibly links the expression levels of *DUX4* to the progressive nature of FSHD.

Determining the mechanism of sporadic *DUX4* activation in skeletal muscle will be key to find targets for therapeutic intervention. Of importance is the identification of the different epigenetic mechanisms regulating the D4Z4 repeat and their relative

contribution to silencing DUX4 in muscle cells as these are potential druggable targets. Furthermore, the epigenetic regulation of D4Z4 can be a determinant of disease severity and variability: both endogenous factors, like epigenetic modifiers of D4Z4 chromatin structure, and environmental factors influencing the epigenome may determine penetrance of FSHD.

Next to new avenues for mechanistic and molecular studies, the firm establishment of the FSHD disease mechanism also paves the road for the generation of animal models to initiate more translational research. The generation of faithful animal models is however challenged by the fact that the D4Z4 macrosatellite and the *DUX4* gene do not have a homologue in a similar genomic context in other, non-primate species⁶⁵. Recapitulation of the FSHD phenotype therefore has to rely on the ectopic expression of DUX4 in animal models, with the potential pitfall that the transcriptional targets of DUX4, and thereby its molecular effects, may not be conserved between different species.

ICF syndrome: an epigenetic disorder in trans

Immunodeficiency, centromere instability and facial anomalies syndrome (OMIM 602900/614064) is a rare autosomal recessive primary immunodeficiency, first described in two independent reports in the late 1970's^{133, 134}. Patients suffer from a triad of phenotypes of which hypo- or a-gammaglobunemia (low or undetectable levels of serum immunoglobulin A (IgA) and IgG) is the most prominent^{135, 136}. Although serum immunoglobulins are drastically decreased in ICF patients, they do have circulating B-cells, suggesting a defect in the final steps of B-cell maturation and immunoglobulin selection and production¹³⁶. A- or hypo-gammaglobunemia in ICF patients results in recurrent infections of the gastro-intestinal and/or respiratory tract, which are often fatal at young age, although some patients show long term survival. These symptoms can be alleviated by immunoglobulin replacement therapy or haematopoietic stem cell transplantation¹³⁵⁻¹³⁷. Nearly all patients present with a distinct but variable spectrum of facial anomalies, of which hypertelorism, flat nasal bridge and epicanthus are the most prevalent^{135, 136}. Further developmental problems include, but are not limited to, a delay in motor and speech development and variable intellectual disability^{135, 136}.

ICF syndrome was one of the first epigenetic disorders to be recognized as such, because of the cytogenetic hallmark of centromere instability on chromosomes 1, 9 and 16^{134, 138}. This instability leads to chromosomal aberrations in cultured patient cells, similar to those observed in cell lines treated with demethylating agents^{134, 138}. The involvement of centromeric chromatin organization was further proven by early reports showing DNA hypomethylation in ICF patients of mainly, but not exclusively, satellite 2 centromeric repeats, highly abundant on chromosomes 1, 9 and 16^{138, 139}.

Three different groups of patients can be recognized based on the genetic defect underlying the syndrome. In the late '90s two papers described mutations in *DNMT3B* to underlie ICF syndrome in approximately half of the patients (ICF1; OMIM 602900)^{140, 141}. The majority of identified mutations affect the catalytic domain of *DNMT3B* and are of

the missense type, resulting in reduced methyltransferase activity of DNMT3B^{136, 142, 143}. ICF1 patients carry at least one partially functional *DNMT3B* copy, as nonsense alleles are only identified in combination with missense alleles in compound heterozygotes¹³⁶. This is in line with the observed phenotypes of mouse models for ICF1. Whereas *Dnmt3b*^{-/-} mice show embryonic lethality, hypomorphic mouse models carrying (patient derived) missense mutations in *Dnmt3b* present with CpG hypomethylation at centromeric satellite repeats, craniofacial abnormalities, runting and an impaired immune system characterized by increased levels of apoptosis in T-cells¹⁴⁴⁻¹⁴⁶. Although some features of ICF syndrome are clearly recapitulated in these models, the most prominent difference is that none of the available models displays impaired B-cell functionality.

In line with a defect in one of the two de novo methyltransferases, the genome of ICF1 patients is characterized by global hypomethylation at both coding and non-coding regions¹⁴⁷⁻¹⁵⁰. Moreover, higher order chromatin structure organization is altered in ICF1 patients, exemplified by a changed nuclear localization of genes on the inactivated X-chromosome¹⁵¹. ICF patients show hypomethylation at various types of repetitive elements throughout the genome, including interspersed LINE repeat elements and tandem repeats like centromeric satellites and the subtelomeric D4Z4 macrosatellite^{139, 147, 152, 153}. CpG hypomethylation at subtelomeres in ICF patients correlates with extensive shortening of the adjacent telomeres¹⁵⁴. Telomeres are transcriptionally active, as they have been shown to produce telomere repeat containing RNAs (TERRA)^{155, 156}. In ICF1 derived patient cell lines an increase in expression levels of TERRA lncRNAs has been observed, most likely in some way linked to the extensive telomere shortening in these cells¹⁵⁴. How these observations contribute to the disease mechanism remains unclear at this point. In general, analyses of methylation at genic and non-genic regions and transcriptional changes in ICF1 patient derived cell lines only showed partial correlations and have not revealed a comprehensive disease mechanism yet^{147-150, 157, 158}.

A second group of patients, negative for mutations in *DNMT3B*, shares all epigenetic and phenotypic characteristics with ICF1 patients, however with additional hypomethylation of alpha-satellite DNA, a centromeric macrosatellite^{135, 159}. Additionally, specific germline genes display hypomethylation and concomitant transcriptional activation in ICF1 derived patient material only¹⁵⁰. Genetic analyses of these DNMT3B mutation negative patients revealed that the majority has mutations in zinc finger and BTB domain containing protein 24 (*ZBTB24*; ICF2; OMIM 614064)¹⁶⁰⁻¹⁶³. In contrast to what has been observed for *DNMT3B*, mutations in *ZBTB24* do not localize to a confined domain of the gene¹³⁶. In fact, mutations in *ZBTB24* are almost exclusively of the nonsense type, most likely leading to complete absence of the full length protein in patients¹³⁶. Thus far, no molecular function has been described for *ZBTB24*, although by homology it is member of the ZBTB family of (hematopoietic) transcription factors¹⁶⁴. The BTB domain, found in the N-terminus of ZBTB24, mediates homo- or hetero-dimerization and may facilitate additional protein-protein interactions¹⁶⁴. The C-terminal zinc finger (ZNF) of ZBTB proteins is thought to mediate the localization to specific DNA sequences. Based hereon, *ZBTB24* is functionally unrelated to *DNMT3B* and discovering its function is of great importance to understand ICF pathology.

The majority of ICF patients is genetically explained by mutations in *DNMT3B* or *ZBTB24*. However, the small group of patients negative for mutations in both genes (ICFX) shows that at least one additional gene defect underlies ICF syndrome. Both the identification of *DNMT3B* and *ZBTB24* have not resulted in a comprehensive pathomechanism for the triad of phenotypes in ICF syndrome yet. Both further characterizing the function of the known ICF genes and identification of the gene(s) underlying ICFX is essential to understand the disease mechanism. The overlapping clinical phenotype of all patients suggests that *DNMT3B*, *ZBTB24*, and any number of additional ICF genes functionally converge at some point.

Outline of the thesis

This thesis focuses on two epigenetic diseases: FSHD and ICF syndrome. Common to both diseases is the epigenetic dysregulation of repetitive DNA, specifically the D4Z4 macrosatellite repeat. We first aimed to better understand the chromatin dysregulation at D4Z4 and its possible correlation to disease severity. In **Chapter 2** we set out to analyse the correlation of the chromatin compaction at D4Z4 in patient derived primary cell lines and the clinical severity. Although trends exist, a significant correlation between clinical severity and chromatin compaction, measured by relative amounts of H3K4me2 and H3K9me3, could not be observed. This study did reveal a clear correlation between muscle pathology in the vastus lateralis muscle and clinical severity. With regard to the known influence of telomeres on the epigenetic regulation of the adjacent subtelomeres, **chapter 3** describes the effect of telomere shortening and cellular aging (senescence) on the epigenetic regulation of subtelomeres. We observed that subtelomeres, including the D4Z4 macrosatellite, are characterized by a shifted balance between markers for constitutive and facultative heterochromatin upon telomere induced senescence.

In **chapter 4**, crosstalk and relative contributions of different epigenetic machineries affecting D4Z4 chromatin structure and *DUX4* activity during muscle differentiation are investigated. *SMCHD1*, the major FSHD2 gene, is found to be a master regulator of the chromatin organization of D4Z4 in both genetic forms of FSHD and forms a barrier between the constitutive heterochromatic nature of D4Z4 and the PRC2 machinery, characteristic of facultative heterochromatin. In **chapter 5** we challenge the position effect hypothesis of telomeres and D4Z4 on the FSHD specific activation of *FRG2*, by showing that *FRG2* is a direct target gene of *DUX4* and follows the expression pattern of other well-established *DUX4* targets. Overall, these chapters highlight the clear epigenetic component in FSHD and the central role of *DUX4* in its pathology.

Chapter 6 describes the generation of two transgenic mouse models. Both models carry human D4Z4 repeats in the size range of either FSHD1 (2.5 copies) or controls (12.5 copies) and our study shows that key (epi)genetic features of D4Z4 are conserved between man and mouse. Where the D4Z4-2.5 mouse recapitulates key features of the disease, including chromatin relaxation of the D4Z4 repeat and sporadic activation of *DUX4*, the D4Z4-12.5 resembles the situation observed in healthy controls. Although this mouse does not show an overt muscular phenotype, it offers great potential in

deciphering the mechanisms underlying DUX4 activation and potential ways to identify drugable targets for FSHD.

Finally, with regards to ICF syndrome, **chapter 7** describes the identification of mutations in the cell division cycle associated 7 (*CDC47*) and the helicase, lymphoid specific (*HELLS*) genes in previously unexplained cases. By doubling the number of ICF disease genes, this work highlights the genetic heterogeneity of the disorder and leaves only few cases unexplained. Molecular characterization of these genes will help to decipher the pathomechanism of ICF syndrome.

References

- Allis,C.D., Jenuwein,T., Reinberg,D., & Caparros,M. Epigenetics(Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2007).
- Saksouk,N., Simboeck,E., & Dejardin,J. Constitutive heterochromatin formation and transcription in mammals. Epigenetics. Chromatin. 8, 3 (2015).
- Holliday,R. Epigenetics: an overview. Dev. Genet. 15, 453-457 (1994).
- Struhl,K. Histone acetylation and transcriptional regulatory mechanisms. Genes Dev. 12, 599-606 (1998).
- Ernst,J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473, 43-49 (2011).
- Sanchez,R., Meslamani,J., & Zhou,M.M. The bromodomain: from epigenome reader to druggable target. Biochim. Biophys. Acta 1839, 676-685 (2014).
- Muller,S., Filippakopoulos,P., & Knapp,S. Bromodomains as therapeutic targets. Expert. Rev. Mol. Med. 13, e29 (2011).
- Mozzetta,C., Boyarchuk,E., Pontis,J., & Ait-Si-Ali,S. Sound of silence: the properties and functions of repressive Lys methyltransferases. Nat. Rev. Mol. Cell Biol. 16, 499-513 (2015).
- Musselman,C.A., Lalonde,M.E., Cote,J., & Kutateladze,T.G. Perceiving the epigenetic landscape through histone readers. Nat. Struct. Mol. Biol. 19, 1218-1227 (2012).
- Lachner,M., O'Carroll,D., Rea,S., Mechtler,K., & Jenuwein,T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature 410, 116-20 (2001).
- Schotta,G. et al. Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J. 21, 1121-1131 (2002).
- Margueron,R. et al. Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461, 762-767 (2009).
- Keogh,M.C. et al. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. Cell 123, 593-605 (2005).
- Joshi,A.A. & Struhl,K. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. Mol. Cell 20, 971-978 (2005).
- Carrozza,M.J. et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123, 581-592 (2005).
- Di,C.L. & Helin,K. Transcriptional regulation by Polycomb group proteins. Nat. Struct. Mol. Biol. 20, 1147-1155 (2013).
- Du,J., Johnson,L.M., Jacobsen,S.E., & Patel,D.J. DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. 16, 519-532 (2015).
- Leonhardt,H., Page,A.W., Weier,H.U., & Bestor,T.H. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell 71, 865-873 (1992).
- Auclair,G., Guibert,S., Bender,A., & Weber,M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biol. 15, 545 (2014).
- Du,Q., Luu,P.L., Stirzaker,C., & Clark,S.J. Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics.1-23 (2015).

21. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930-935 (2009).
22. Hahn, M.A., Szabo, P.E., & Pfeifer, G.P. 5-Hydroxymethylcytosine: a stable or transient DNA modification? *Genomics* 104, 314-323 (2014).
23. Baylin, S.B. & Jones, P.A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer* 11, 726-734 (2011).
24. Kulis, M. & Esteller, M. DNA methylation and cancer. *Adv. Genet.* 70, 27-56 (2010).
25. Lee, J.T. & Bartolomei, M.S. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 152, 1308-1323 (2013).
26. Fahrner, J.A. & Bjornsson, H.T. Mendelian disorders of the epigenetic machinery: tipping the balance of chromatin states. *Annu. Rev. Genomics Hum. Genet.* 15, 269-293 (2014).
27. Usdin, K. & Kumari, D. Repeat-mediated epigenetic dysregulation of the FMR1 gene in the fragile X-related disorders. *Front Genet.* 6, 192 (2015).
28. Ng, S.B. et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* 42, 790-793 (2010).
29. Lederer, D. et al. Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. *Am. J. Hum. Genet.* 90, 119-124 (2012).
30. Venter, J.C. et al. The sequence of the human genome. *Science* 291, 1304-1351 (2001).
31. Engel, S.R. et al. The reference genome sequence of *Saccharomyces cerevisiae*: then and now. *G3. (Bethesda.)* 4, 389-398 (2014).
32. Adams, M.D. et al. The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185-2195 (2000).
33. Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M., & Gerstein, M.B. Annotating non-coding regions of the genome. *Nat. Rev. Genet.* 11, 559-571 (2010).
34. Ohno, S. So much "junk" DNA in our genome. *Brookhaven. Symp. Biol.* 23, 366-370 (1972).
35. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74 (2012).
36. Smit, A.F.A., Hubley, R., & Green, P. RepeatMasker Open - 4.0. www.repeatmasker.org (2015).
37. Schaap, M. et al. Genome-wide analysis of macrosatellite repeat copy number variation in worldwide populations: evidence for differences and commonalities in size distributions and size restrictions. *BMC. Genomics* 14, 143 (2013).
38. Deenen, J.C. et al. Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* (2014).
39. Landouzy, L. & Dejerine, J. De la myopathie atrophique progressive. *Rev Med* 5, 253-366 (1885).
40. Statland, J. & Tawil, R. Facioscapulohumeral muscular dystrophy. *Neurol. Clin.* 32, 721-8, ix (2014).
41. van der Maarel, S.M. et al. De novo facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. *Am. J. Hum. Genet.* 66, 26-35 (2000).
42. Wijmenga, C. et al. Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. *Genomics* 9, 570-575 (1991).
43. Wijmenga, C. et al. Genetic linkage map of facioscapulohumeral muscular dystrophy and five polymorphic loci on chromosome 4q35-qter. *Am. J. Hum. Genet.* 51, 411-415 (1992).
44. Weiffenbach, B. et al. Linkage analyses of five chromosome 4 markers localizes the facioscapulohumeral muscular dystrophy (FSHD) gene to distal 4q35. *Am. J. Hum. Genet.* 51, 416-423 (1992).
45. Lemmers, R.J. et al. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat. Genet.* 32, 235-236 (2002).
46. Lemmers, R.J. et al. Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 75, 1124-1130 (2004).
47. Lemmers, R.J. et al. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am. J Hum. Genet* 81, 884-894 (2007).
48. van Geel, M. et al. Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. *Genomics* 79, 210-7 (2002).
49. Tupler, R. et al. Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. *J Med Genet* 33, 366-370 (1996).

50. van Deutekom, J.C. et al. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* 2, 2037-2042 (1993).
51. Wijmenga, C., Brouwer, O.F., Padberg, G.W., & Frants, R.R. Transmission of de-novo mutation associated with facioscapulohumeral muscular dystrophy. *Lancet* 340, 985-986 (1992).
52. Wijmenga, C. et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 2, 26-30 (1992).
53. Lunt, P.W. et al. Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). *Hum Mol Genet* 4, 951-958 (1995).
54. Tawil, R. et al. Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann Neurol* 39, 744-748 (1996).
55. de Greef, J.C. et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mutat.* 30, 1449-1459 (2009).
56. van Overveld, P.G. et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet* 35, 315-317 (2003).
57. Ambrosini, A., Paul, S., Hu, S., & Riethman, H. Human subtelomeric duplicon structure and organization. *Genome Biol.* 8, R151 (2007).
58. Mefford, H.C. & Trask, B.J. The complex structure and dynamic evolution of human subtelomeres. *Nat Rev Genet* 3, 91-102 (2002).
59. Bakker, E. et al. The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve* 2, 39-44 (1995).
60. Deidda, G. et al. Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the Facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur J Hum Genet* 3, 155-167 (1995).
61. Hewitt, J.E. et al. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* 3, 1287-1295 (1994).
62. Lyle, R., Wright, T.J., Clark, L.N., & Hewitt, J.E. 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 (1995).
63. Winokur, S.T. et al. 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 Res* 2, 225-234 (1994).
64. Clapp, J. et al. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* 81, 264-279 (2007).
65. Leidenroth, A. & Hewitt, J.E. A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC. Evol. Biol.* 10, 364 (2010).
66. Leidenroth, A. et al. Evolution of DUX gene macrosatellites in placental mammals. *Chromosoma* 121, 489-497 (2012).
67. Bosnakovski, D. et al. DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. *Exp. Neurol.* 214, 87-96 (2008).
68. Dixit, M. et al. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci. U. S. A* 104, 18157-18162 (2007).
69. Kowaljow, V. et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord* 17, 611-623 (2007).
70. Wallace, L.M. et al. DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann. Neurol.* 69, 540-552 (2011).
71. Gaillard, M.C. et al. Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. *Neurology* 83, 733-742 (2014).
72. Hartweck, L.M. et al. A focal domain of extreme demethylation within D4Z4 in FSHD2. *Neurology* 80, 392-399 (2013).
73. Huichalaf, C., Micheloni, S., Ferri, G., Caccia, R., & Gabellini, D. DNA methylation analysis of the macrosatellite repeat associated with FSHD muscular dystrophy at single nucleotide level. *PLoS. ONE.* 9, e115278 (2014).
74. Jones, T.I. et al. Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with

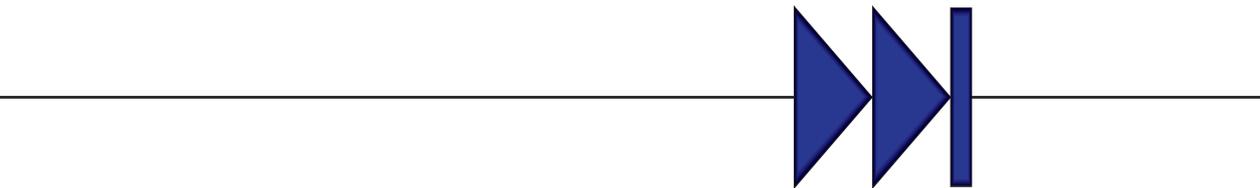
- disease in facioscapulohumeral muscular dystrophy. *Clin. Epigenetics*. 7, 37 (2015).
75. Zeng,W. et al. Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats in facioscapulohumeral dystrophy (FHD). *PLoS. Genet* 7, e1000559 (2009).
 76. Zeng,W. et al. Genetic and Epigenetic Characteristics of FHD-Associated 4q and 10q D4Z4 that are Distinct from Non-4q/10q D4Z4 Homologs. *Hum. Mutat.* 35, 998-1010 (2014).
 77. Lemmers,R.J. et al. Digenic inheritance of an SMCHD1 mutation and an FHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* 44, 1370-1374 (2012).
 78. Blewitt,M.E. et al. SmCHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* 40, 663-669 (2008).
 79. Zhao,J., Sun,B.K., Erwin,J.A., Song,J.J., & Lee,J.T. Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. *Science* 322, 750-756 (2008).
 80. Gendrel,A.V. et al. Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev. Cell* 23, 265-279 (2012).
 81. Gendrel,A.V. et al. Epigenetic functions of smchd1 repress gene clusters on the inactive x chromosome and on autosomes. *Mol. Cell Biol.* 33, 3150-3165 (2013).
 82. Nozawa,R.S. et al. Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat. Struct. Mol. Biol.* 20, 566-573 (2013).
 83. Mould,A.W. et al. Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics. Chromatin.* 6, 19 (2013).
 84. Larsen,M. et al. Diagnostic approach for FHD revisited: SMCHD1 mutations cause FHD2 and act as modifiers of disease severity in FHD1. *Eur. J. Hum. Genet.* 23, 808-816 (2015).
 85. Sacconi,S. et al. The FHD2 gene SMCHD1 is a modifier of disease severity in families affected by FHD1. *Am. J. Hum. Genet.* 93, 744-751 (2013).
 86. Cabianca,D.S. et al. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FHD muscular dystrophy. *Cell* 149, 819-831 (2012).
 87. Lim,J.W. et al. DICER/AGO-dependent epigenetic silencing of D4Z4 repeats enhanced by exogenous siRNA suggests mechanisms and therapies for FHD. *Hum. Mol. Genet.*(2015).
 88. Masny,P.S. et al. Localization of 4q35.2 to the nuclear periphery: is FHD a nuclear envelope disease? *Hum Mol. Genet* 13, 1857-1871 (2004).
 89. Petrov,A. et al. Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. *Proc. Natl. Acad. Sci. U. S. A* 103, 6982-6987 (2006).
 90. Petrov,A. et al. A nuclear matrix attachment site in the 4q35 locus has an enhancer-blocking activity in vivo: implications for the facio-scapulo-humeral dystrophy. *Genome Res.*(2007).
 91. Bodega,B. et al. Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FHD) locus and upregulation of FHD-related gene 1 (FRG1) expression during human myogenic differentiation. *BMC. Biol.* 7, 41 (2009).
 92. Pirozhkova,I. et al. 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, e3389 (2008).
 93. Rijkers,T. et al. FRG2, an FHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FHD patients. *J Med Genet* 41, 826-836 (2004).
 94. van Deutekom,J.C.T. et al. Identification of the first gene (FRG1) from the FHD region on human chromosome 4q35. *Hum Mol Genet* 5, 581-590 (1996).
 95. Gabellini,D., Green,M., & Tupler,R. Inappropriate Gene Activation in FHD. A Repressor Complex Binds a Chromosomal Repeat Deleted in Dystrophic Muscle. *Cell* 110, 339-248 (2002).
 96. Jiang,G. et al. Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum Mol. Genet* 12, 2909-2921 (2003).
 97. Klooster,R. et al. Comprehensive expression analysis of FHD candidate genes at the mRNA and protein level. *Eur. J. Hum. Genet.* 17, 1615-1624 (2009).
 98. Arashiro,P. et al. Transcriptional regulation differs in affected facioscapulohumeral muscular dystrophy patients compared to asymptomatic related carriers. *Proc. Natl. Acad. Sci. U. S. A* 106, 6220-6225 (2009).

99. Celegato,B. et al. 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 (2006).
100. Cheli,S. et al. Expression profiling of FSHD-1 and FSHD-2 cells during myogenic differentiation evidences common and distinctive gene dysregulation patterns. *PLoS. ONE*. 6, e20966 (2011).
101. Gabellini,D. et al. Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature* 439, 973-977 (2006).
102. Hanel,M.L., Wuebbles,R.D., & Jones,P.L. Muscular dystrophy candidate gene FRG1 is critical for muscle development. *Dev. Dyn*. 238, 1502-1512 (2009).
103. Liu,Q., Jones,T.I., Tang,V.W., Brieher,W.M., & Jones,P.L. Facioscapulohumeral muscular dystrophy region gene-1 (FRG-1) is an actin-bundling protein associated with muscle-attachment sites. *J. Cell Sci*. 123, 1116-1123 (2010).
104. Masny,P.S. et al. Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei. *Eur. J. Hum. Genet*. 18, 448-456 (2010).
105. Osborne,R.J., Welle,S., Venance,S.L., Thornton,C.A., & Tawil,R. Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology* 68, 569-577 (2007).
106. Pistoni,M. et al. Rbfox1 downregulation and altered calpain 3 splicing by FRG1 in a mouse model of Facioscapulohumeral muscular dystrophy (FSHD). *PLoS. Genet*. 9, e1003186 (2013).
107. Sancisi,V. et al. Altered Tnnt3 characterizes selective weakness of fast fibers in mice overexpressing FSHD region gene 1 (FRG1). *Am. J. Physiol Regul. Integr. Comp Physiol* 306, R124-R137 (2014).
108. Sun,C.Y. et al. Facioscapulohumeral muscular dystrophy region gene 1 is a dynamic RNA-associated and actin-bundling protein. *J. Mol. Biol*. 411, 397-416 (2011).
109. Tsumagari,K. et al. Gene expression during normal and FSHD myogenesis. *BMC. Med. Genomics* 4, 67 (2011).
110. Winokur,S.T. et al. Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum Mol. Genet* 12, 2895-2907 (2003).
111. Xu,X. et al. DNaseI hypersensitivity at gene-poor, FSH dystrophy-linked 4q35.2. *Nucleic Acids Res*. 37, 7381-7393 (2009).
112. Yao,Z. et al. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum. Mol. Genet.*(2014).
113. Mariot,V. et al. Correlation between low FAT1 expression and early affected muscle in facioscapulohumeral muscular dystrophy. *Ann Neurol*. 78, 387-400 (2015).
114. Puppo,F. et al. Identification of variants in the 4q35 gene FAT1 in patients with a facioscapulohumeral dystrophy-like phenotype. *Hum. Mutat*. 36, 443-453 (2015).
115. Caruso,N. et al. Deregulation of the protocadherin gene FAT1 alters muscle shapes: implications for the pathogenesis of facioscapulohumeral dystrophy. *PLoS. Genet*. 9, e1003550 (2013).
116. Snider,L. et al. RNA Transcripts, miRNA-sized Fragments, and Proteins Produced from D4Z4 Units: New Candidates for the Pathophysiology of Facioscapulohumeral Dystrophy. *Hum. Mol. Genet* 18, 2414-2430 (2009).
117. Lemmers,R.J. et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650-1653 (2010).
118. Jones,T.I. et al. Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. *Hum. Mol. Genet*. 21, 4419-4430 (2012).
119. Snider,L. et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS. Genet*. 6, e1001181 (2010).
120. Geng,L.N. et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38-51 (2012).
121. Young,J.M. et al. DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS. Genet*. 9, e1003947 (2013).
122. Ferreboeuf,M. et al. DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. *Hum. Mol. Genet*. 23, 171-181 (2014).
123. Rickard,A.M., Petek,L.M., & Miller,D.G. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum. Mol.*

- Genet. 24, 5901-5914 (2015).
124. Xu,H. et al. Dux4 induces cell cycle arrest at G1 phase through upregulation of p21 expression. *Biochem. Biophys. Res. Commun.* 446, 235-240 (2014).
 125. Dandapat,A., Hartweck,L.M., Bosnakovski,D., & Kyba,M. Expression of the human FSHD-linked DUX4 gene induces neurogenesis during differentiation of murine embryonic stem cells. *Stem Cells Dev.* 22, 2440-2448 (2013).
 126. de la Kethulle de Ryhove et al. The Role of D4Z4-Encoded Proteins in the Osteogenic Differentiation of Mesenchymal Stromal Cells Isolated from Bone Marrow. *Stem Cells Dev.* 24, 2674-2686 (2015).
 127. Feng,Q. et al. A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *Elife.* 4, (2015).
 128. Block,G.J. et al. Wnt/beta-catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. *Hum. Mol. Genet.*(2013).
 129. Himeda,C.L. et al. Myogenic enhancers regulate expression of the facioscapulohumeral muscular dystrophy-associated DUX4 gene. *Mol. Cell Biol.* 34, 1942-1955 (2014).
 130. Stadler,G. et al. Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat. Struct. Mol. Biol.* 20, 671-678 (2013).
 131. Benetti,R., Garcia-Cao,M., & Blasco,M.A. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat. Genet.* 39, 243-250 (2007).
 132. Blasco,M.A. The epigenetic regulation of mammalian telomeres. *Nat. Rev. Genet.* 8, 299-309 (2007).
 133. Hulten,M. Selective Somatic Pairing and Fragility at 1q12 in a Boy with Common Variable Immuno Deficiency. *Clinical Genetics* 14, 294 (1978).
 134. Tiepolo,L. et al. Multibranching chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. *Hum. Genet.* 51, 127-137 (1979).
 135. Hagleitner,M.M. et al. Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). *J. Med. Genet.* 45, 93-99 (2008).
 136. Weemaes,C.M. et al. Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. *Eur. J. Hum. Genet.* 21, 1219-1225 (2013).
 137. Gennery,A.R. et al. Hematopoietic stem cell transplantation corrects the immunologic abnormalities associated with immunodeficiency-centromeric instability-facial dysmorphism syndrome. *Pediatrics* 120, e1341-e1344 (2007).
 138. Jeanpierre,M. et al. An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Hum. Mol. Genet.* 2, 731-735 (1993).
 139. Miniou,P. et al. Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. *Hum Mol. Genet* 3, 2093-2102 (1994).
 140. Hansen,R.S. et al. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. U. S. A* 96, 14412-14417 (1999).
 141. Xu,G.L. et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 402, 187-191 (1999).
 142. Gowher,H. & Jeltsch,A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J. Biol. Chem.* 277, 20409-20414 (2002).
 143. Moarefi,A.H. & Chedin,F. ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. *J. Mol. Biol.* 409, 758-772 (2011).
 144. Ueda,Y. et al. Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development* 133, 1183-1192 (2006).
 145. Velasco,G. et al. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl. Acad. Sci. U. S. A* 107, 9281-9286 (2010).
 146. Youngson,N.A. et al. No evidence for cumulative effects in a Dnmt3b hypomorph across multiple generations. *Mamm. Genome* 24, 206-217 (2013).
 147. Heyn,H. et al. Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. *Epigenetics.* 7, 542-550 (2012).
 148. Jin,B. et al. DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Hum. Mol. Genet.* 17, 690-709 (2008).

149. Simo-Riudalbas, L. et al. Genome-Wide DNA Methylation Analysis Identifies Novel Hypomethylated Non-Pericentromeric Genes with Potential Clinical Implications in ICF Syndrome. *PLoS ONE*. 10, e0132517 (2015).
150. Velasco, G. et al. Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology. *Orphanet. J. Rare Dis.* 9, 56 (2014).
151. Matarazzo, M.R., Boyle, S., D'Esposito, M., & Bickmore, W.A. Chromosome territory reorganization in a human disease with altered DNA methylation. *Proc. Natl. Acad. Sci. U. S. A* 104, 16546-16551 (2007).
152. Hansen, R.S. X inactivation-specific methylation of LINE-1 elements by DNMT3B: implications for the Lyon repeat hypothesis. *Hum. Mol. Genet.* 12, 2559-2567 (2003).
153. Miniou, P., Bourc'his, D., Molina, G.D., Jeanpierre, M., & Viegas-Pequignot, E. Undermethylation of Alu sequences in ICF syndrome: molecular and in situ analysis. *Cytogenet. Cell Genet.* 77, 308-313 (1997).
154. Yehezkel, S., Segev, Y., Viegas-Pequignot, E., Skorecki, K., & Selig, S. Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. *Hum. Mol. Genet.* 17, 2776-2789 (2008).
155. Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E., & Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318, 798-801 (2007).
156. Schoeftner, S. & Blasco, M.A. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat. Cell Biol.* 10, 228-236 (2008).
157. Ehrlich, M. et al. DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. *Hum. Mol. Genet.* 10, 2917-2931 (2001).
158. Ehrlich, M. et al. ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. *Autoimmunity* 41, 253-271 (2008).
159. Jiang, Y.L. et al. DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. *Hum. Mutat.* 25, 56-63 (2005).
160. Cerbone, M. et al. Immunodeficiency, centromeric instability, facial anomalies (ICF) syndrome, due to ZBTB24 mutations, presenting with large cerebral cyst. *Am. J. Med. Genet. A* 158A, 2043-2046 (2012).
161. Chouery, E. et al. A novel deletion in ZBTB24 in a Lebanese family with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *Clin. Genet.* 82, 489-493 (2012).
162. de Greef, J.C. et al. Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *Am. J. Hum. Genet.* 88, 796-804 (2011).
163. Nitta, H. et al. Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. *J. Hum. Genet.* 58, 455-460 (2013).
164. Lee, S.U. & Maeda, T. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. *Immunol. Rev.* 247, 107-119 (2012).

Budding yeast drawing: Database Center for Life Science (DBCLS) via Wikimedia Commons
 Drosophila-drawing: I. B. Nuhanen. Licensed under CC BY-SA 3.0 via Wikimedia Commons
 Line-drawing of a human man by Mikael Häggström via Wikimedia Commons



**Correlation analysis of clinical
parameters with epigenetic
modifications in the DUX4
promoter in FSHD.**

2

Judit Balog, Peter E. Thijssen, Jessica C. de Greef, Bharati Shah, Baziël G.M. van Engelen, Kyoko Yokomori, Stephen J. Tapscott, Rabi Tawil and Silvère M. van der Maarel

Adapted from Balog et al., (2012), *Epigenetics*, 7:6, 579-584

Abstract

The aim of our study was to identify relationships between epigenetic parameters correlating with a relaxed chromatin state of the DUX4 promoter region and clinical severity as measured by a clinical severity score or muscle pathologic changes in D4Z4 contraction-dependent (FSHD1) and –independent (FSHD2) facioscapulohumeral muscular dystrophy patients. Twenty primary fibroblast (5 control, 10 FSHD1 and 5 FSHD2) and 26 primary myoblast (9 control, 12 FSHD1 and 5 FSHD2) cultures originating from patients with FSHD and controls were analyzed. Histone modification levels were determined by chromatin immunoprecipitation. We examined correlations between the chromatin compaction score (ChCS) defined by the H3K9me3:H3K4me2 ratio and an age corrected clinical severity score (CSS) or muscle pathology score (MPS). Possible relationships were investigated using linear regression analysis and significance was tested by Pearson's product-moment coefficient.

We found a significant difference of the ChCS between controls and patients with FSHD1 and between controls and patients with FSHD2. Tissue specific differences in ChCS were also observed. We also found a near-significant relationship between ChCS and the age corrected CSS in fibroblasts but not in myoblasts. Surprisingly, we found a strong correlation between the MPS of the vastus lateralis and the CSS. Our results confirm the D4Z4 chromatin relaxation previously shown to be associated with FSHD in a small number of samples. A possible relationship between clinical and epigenetic parameters could be established in patient fibroblasts, but not in myoblasts. The strong correlation between the MPS of the vastus lateralis and the CSS suggests that this muscle can be used to study for surrogate markers of overall disease severity.

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant myopathy characterized initially by progressive wasting of the facial, shoulder and upper arm muscles¹. Two distinct groups of affected individuals can be defined: > 95% of patients have a contracted D4Z4 repeat array below 10 units in the subtelomeric region of chromosome 4 (FSHD1) while < 5% of patients do not have a contracted D4Z4 array (FSHD2)². Both FSHD1 and FSHD2 patients show CpG hypomethylation and specific loss of the repressive histone modification histone 3 lysine 9 trimethylation (H3K9me3) at the D4Z4 array^{3,4}. While in patients with FSHD1 the hypomethylation is restricted to the contracted allele, both alleles on chromosome 4 and both homologous arrays on chromosome 10 are hypomethylated in FSHD2⁵. These epigenetic changes result in a relative chromatin relaxation which may facilitate the expression of full length double homeobox gene 4 (*fDUX4*) (Fig. 1)^{4,5}. However, in cultured myotubes from patients with FSHD, it is found that *fDUX4* expression occurs only on specific genetic backgrounds of chromosome 4 containing an additional *DUX4* exon immediately distal to the D4Z4 repeat which stabilizes the transcript and leads to the production of abundant amounts of DUX4 protein in sporadic myonuclei⁶⁻⁸.

The age at onset, severity and disease progression is highly variable between and within FSHD families¹. Several studies addressed the possibility of a correlation between disease severity measured by age at onset or clinical severity score (CSS) and the residual repeat size of the disease-associated D4Z4 repeat in FSHD1 patients⁹⁻¹¹. A rough and inverse correlation was reported between D4Z4 unit number and severity: individuals with 1–3 units typically exhibited the most severe phenotype, while patients with 4–10 D4Z4 units generally show a high variability in clinical presentation¹².

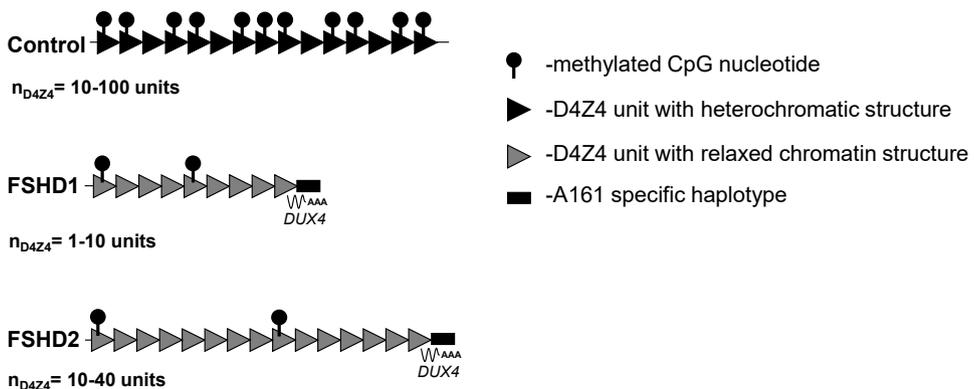


Figure 1: Schematic representation of the FSHD locus in control individuals, patients with contraction dependent (FSHD1) and patients with contraction independent (FSHD2) FSHD. D4Z4 repeat units are indicated with triangles, the additional *DUX4* exon with stabilizing polyA signal distal to the repeat is indicated with a box. Common chromatin features of D4Z4 in FSHD1 and FSHD2 are decreased CpG methylation levels and a more relaxed chromatin structure facilitating the expression of *DUX4* from the telomeric repeat unit. Overall reduction in CpG methylation of D4Z4 in FSHD is schematically indicated with reduced number of lollipop symbols.

The correlation between CpG methylation level at D4Z4 and age-corrected CSS was also studied in FSHD1 and FSHD2 patients. No linear relationship was found between the CpG methylation level of the first D4Z4 unit on the contracted allele in peripheral blood lymphocytes and the age-corrected CSS in a limited number of patients¹².

A recent study emphasizes the role of histone modifications in the pathomechanism of FSHD⁴. A combined presence of the transcriptionally permissive modification histone 3 lysine 4 dimethylation (H3K4me2) and repressive histone modification H3K9me3 was detected in the promoter region of *DUX4*. A specific loss of H3K9me3 was reported in immortalized peripheral lymphoblastoid cell lines, primary fibroblast and primary myoblast samples of FSHD patients, both FSHD1 and FSHD2. We hypothesized that the ratio of H3K9me3 and H3K4me2, herein referred to as the chromatin compaction score (ChCS), is significantly different between control and affected individuals and examined the relationship between ChCS and measures of overall disease severity as well as between ChCS and histopathological changes in muscle.

Results

Significant difference in chromatin compaction score between controls and FSHD patients

Since D4Z4 repeats in FSHD1 and FSHD2 patients show identical epigenetic changes, we tested whether the ChCS, the ratio of H3K9me3:H3K4me2, discriminates control samples from FSHD1 and FSHD2 samples. Our study includes 5 control, 10 FSHD1 and 5 FSHD2 primary fibroblast cultures in addition to 9 control, 12 FSHD1 and 5 FSHD2 myoblast cultures. Cross-linked ChIP was performed using H3K9me3 and H3K4me2 antibodies and ChCS were calculated by dividing H3K9me3 relative enrichment values by H3K4me2 relative enrichment values. Statistical analysis of fibroblast data showed significant differences between control and FSHD1 ($P < 0.01$) and control and FSHD2 samples ($P < 0.01$) (Fig. 2A, B). ChCS were decreased in patient samples indicating a D4Z4 chromatin relaxation in patient samples.

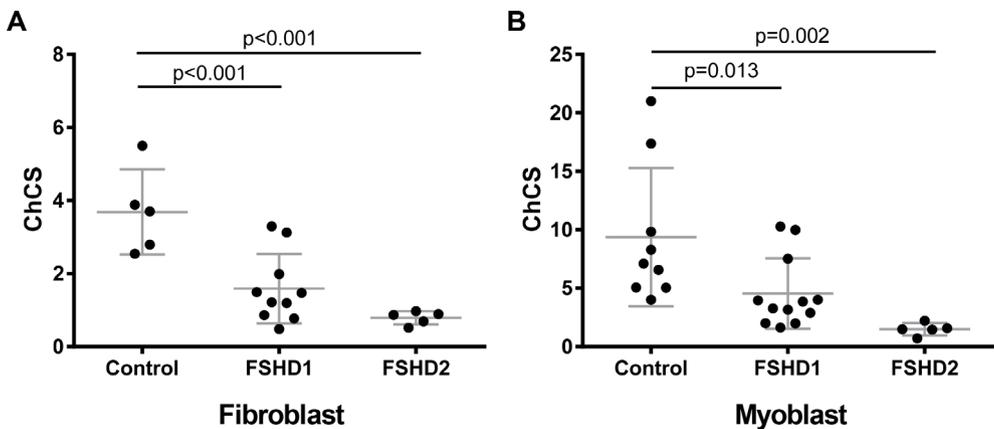


Figure 2. Scatterplot of ChCS in fibroblasts (A) or myoblasts (B) of controls, FSHD1 and FSHD2 patients. Individual samples are represented by individual data points. Mean values with 1xSD are indicated.

Since FSHD is a progressive muscle disease we hypothesized that a relationship between epigenetic marks and the clinical phenotype is more pronounced in primary myoblast cultures. To avoid possible differences caused by the replicative history of the culture, chromatin was isolated from all samples at passage 6 or 7. Similar to the fibroblast results we found that the ChCS was significantly lower in FSHD1 compared with control ($P < 0.05$) as well as in FSHD2 compared with control ($P < 0.01$) samples. We also observed that maximum ChCS values in fibroblast cultures (6 on y axis in **Fig. 2A**) are lower than maximum ChCS values in myoblast cultures (25 on y axis on **Fig. 2B**).

Correlation analysis of ChCS and age-corrected CSS in myoblast and fibroblast samples

Next we performed linear regression analysis to study the relationship between ChCS and age-correlated CSS. Analysis was done separately on fibroblast and myoblast using all samples and separating them by disease subtype (FSHD1 or FSHD2). Since gender differences have been reported for the penetrance and the severity of the disease, samples were also analyzed after separation based on gender. A summary of analyses with corresponding p values determined by Pearson product moment coefficient are shown in **table 1**.

We found significant negative correlation between ChCS and age-corrected CSS in male FSHD1 and FSHD2 fibroblast samples ($P = 0.028$) suggesting that the chromatin at D4Z4 is more relaxed in samples originating from males with a more severe phenotype. When combining male and female fibroblast samples the negative correlation reached near-significance ($P = 0.062$). There was no significant correlation found in myoblasts in any combination of samples studied. Our results of fibroblast ($P = 0.062$) and myoblast samples ($P = 0.801$) are presented in **Figure 3**.

Table 1. Summary of analyses of age corrected CSS and CRS

Analyzed samples		Fibroblast			Myoblast		
Gender	Condition	n	Pearson-r	p-value	n	Pearson-r	p-value
M	FSHD1	7	-0.612	0.144	3	0.095	0.940
F	FSHD1	3	-0.166	0.894	9	-0.12	0.758
M + F	FSHD1	10	-0.437	0.206	12	-0.186	0.562
M	FSHD2	3	-0.258	0.834	3	0.992	0.082
F	FSHD2	2	NA	NA	2	NA	NA
M + F	FSHD2	5	0.127	0.838	5	0.319	0.600
M	FSHD1 + FSHD2	10	-0.686	0.028	6	0.743	0.090
F	FSHD1 + FSHD2	5	0.014	0.982	11	-0.136	0.690
M + F	FSHD1 + FSHD2	15	-0.492	0.062	17	-0.066	0.801

Correlation analysis of ChCS and MPS in myoblast samples

Muscle biopsies were scored for MPS as described in the methods section. Correlation analysis between ChCS and MPS was done on 6 FSHD1 and 5 FSHD2 patients for whom both scores were available (**Fig. S1**). Statistical analysis showed no relationship between MPS and ChCS.

Correlation analysis of MPS and CSS

Finally we tested whether we could find a relationship between vastus lateralis MPS and CSS. Since MPS and CSS are compared in the same patient, we refrained from using the age-correction in establishing the clinical severity. As shown in **Figure 4**, in the 11 FSHD patients for which both scores were available we found a significant and positive relationship between the muscle pathology score and CSS ($r = 0.85$; $P = 0.0009$).

Discussion

FSHD presents with high and unpredictable inter- and intrafamilial clinical variability in disease onset and progression¹. To provide patients with a better prognosis it would be beneficial to identify genetic or epigenetic markers that correlate with disease severity. Earlier studies failed to identify correlations between genetic markers (D4Z4 repeat size) or epigenetic markers (D4Z4 methylation) that can be used in the clinic^{9,12}.

The focus of the present study was to investigate possible relationships between the epigenetic changes occurring in primary fibroblasts and myoblasts of FSHD1 and FSHD2 patients at the promoter region of *DUX4* and the clinical severity defined by the CSS or MPS. As a measure of chromatin changes, first we introduced the chromatin compaction score, ChCS, in which the relative enrichment of H3K9me3 at D4Z4 is divided by the relative enrichment of H3K4me2. The advantage of such a score is that it circumvents the correlation calculation problem of ChIP experiments studying repetitive sequences. Different samples contain different numbers of D4Z4 units and our experimental procedure detects all repeat units on chromosomes 4 and 10. A comparison between

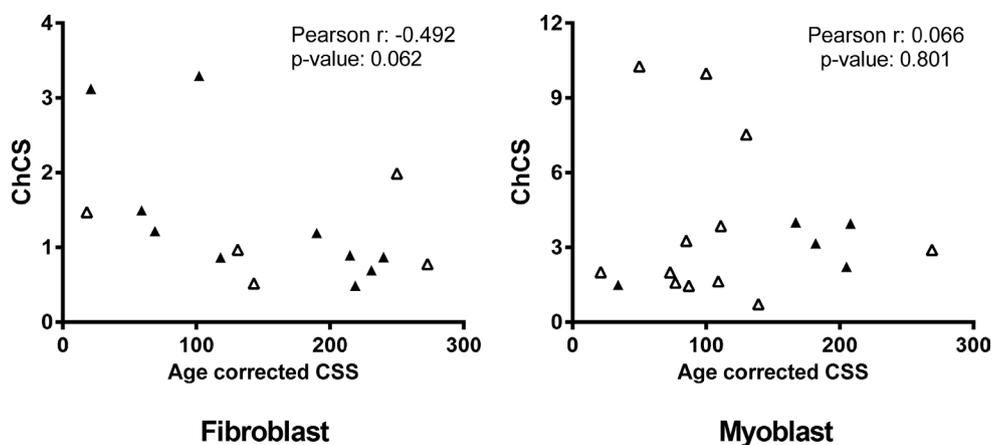


Figure 3. Scatterplot of ChCS vs. age-corrected CSS in fibroblast (A) and in myoblast samples (B). Male samples are represented by closed triangles, female samples by open triangles.

different samples is only reliable if the measured signal can be correlated with D4Z4 unit numbers. Since both H3K9me3 and H3K4me2 are present in every unit, their ratio, as calculated in the ChCS, gives a value per unit.

Statistical analysis revealed a significant difference between the ChCS of control and FSHD1 and between control and FSHD2 in fibroblast and myoblast cultures. These results, showing a significant decrease in ChCS in FSHD cell lines, confirm the previously reported epigenetic alterations at D4Z4 suggesting that the D4Z4 array in FSHD patients is less compact than in controls⁴.

Comparison of ChCS values between the disease subtypes separately in fibroblast and myoblast samples showed that independent of cell type FSHD1 ChCS values show high variation while FSHD2 ChCS values are lower and less variable. This might be explained by a much more widespread chromatin relaxation over all four repeats in FSHD2, while in FSHD1 the changes may be more restricted to the pathogenic allele. This is also observed for the loss of DNA methylation at D4Z4⁵.

We also detected tissue-specific differences in the chromatin landscape at D4Z4 arrays of fibroblast and myoblast cultures with control myoblast ChCS values being higher than control fibroblast ChCS values. Genome-wide ChIP-seq studies have shown that the chromatin landscape in a particular cell type reflects its transcriptional activity and is therefore cell type-specific¹³. Our results show that ChCS values are also cell-type specific with a more compact chromatin organization of D4Z4 in myoblasts to prevent expression of *FLUX4*. This may explain the specific tissue involvement in FSHD.

Correlation analysis of the ChCS and age-corrected CSS revealed a significant relationship in male fibroblast patient samples ($P = 0.028$). Correlation analysis of all fibroblast samples gave borderline significance ($P = 0.062$). Gender differences for disease penetrance and severity have been reported and showed that males are typically more severely affected than females¹⁴. The weak correlation may be caused by the small sample size or because of technical reasons. In our ChIP assays we measure collectively

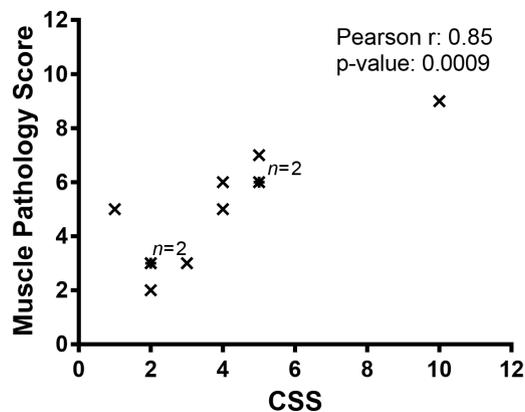


Figure 4. Relationship analysis between the CSS and MPS. There is a significant correlation between MPS and CSS ($P = 0.0009$). Note that the analysis was done without age correction.

the chromatin status of each DUX4 promoter in the entire array, rather than specifically interrogating the most telomeric unit from which fDUX4 originates. Moreover, in FSHD1 we are not only measuring the DUX4 promoters of the pathogenic allele, but instead all D4Z4 and D4Z4-like arrays on chromosomes 4 and 10. Considering that in FSHD2 the chromatin status of all four chromosomes seems to be affected, this limitation should not apply to FSHD2. Nevertheless, also in FSHD2 cases we could not detect relationship between ChCS and age corrected CSS possibly due to small sample size.

Statistical analysis of myoblast data could not detect significant relationship between ChCS and age corrected CSS in any combination of samples (**Table 1**). Myoblasts are more sensitive to culturing conditions, often less homogeneous and can quickly transform to a pre-differentiation state, factors that may all affect their epigenetic profile. An alternative explanation might be that myoblast cultures carry a 'history of disease' that may affect the chromatin status of D4Z4 while fibroblasts, being derived from a non-affected tissue are not expected to have such history.

Irrespective of the explanation, if a negative correlation between the ChCS and CSS in fibroblast samples can be confirmed in larger group of patients, it would provide the first prognostic marker that can be measured in a relatively accessible tissue. In addition, since we could observe significant ChCS changes in myoblasts and fibroblasts, it is tempting to speculate that this score can also be used for prenatal diagnosis of FSHD2, a condition for which currently no validated diagnostic test exists.

An interesting observation from our study is that there is strong correlation between the CSS and the MPS of the vastus lateralis muscle. These results indicate that the pathology observed in the vastus lateralis muscle, typically not clinically affected in the early stages of FSHD, provides a good reflection of overall disease severity. This result shows that the vastus lateralis muscle, easily and safely accessible by needle biopsy techniques, is an ideal muscle to study for surrogate markers of overall disease severity. A non-invasive modality such as MRI to score the involvement of the vastus lateralis would be even more preferable. A study of lower extremity MRI in FSHD using a simple scoring system based on T1-W sequences was recently completed¹⁵. This study showed that the combined MRI score of all lower extremity muscles was highly correlated with CSS. Whether this correlation holds true when the MRI score is based on the score of only the vastus lateralis remains to be determined.

In summary, our study of a large cohort of patient and control fibroblast and myoblast cultures confirms earlier reports of D4Z4 chromatin relaxation in FSHD and introduces a practical chromatin compaction score (ChCS). We found a near-significant correlation between the ChCS and severity in patient fibroblast samples and identified an unexpected correlation between the CSS and the MPS. We propose that factors in addition to locus intrinsic properties, influence disease severity in FSHD.

Materials and methods

Patients and control individuals

All samples from controls and FSHD patients were obtained after signing informed

consents. Samples from 22 FSHD1, 10 FSHD2 and 14 controls were studied. Samples were analyzed for array size, 4q haplotype and DNA methylation according to published methods before the study (**Tables S1 and S2**)^{3, 16}. Fibroblasts were derived from 2 mm punch skin biopsies obtained from the medial aspect of the distal forearm. Myoblasts were derived from needle muscle biopsy samples taken from the vastus lateralis muscle. An adjacent sample of muscle was processed for histopathology. Muscle and skin biopsy procedures are described here: <http://www.urmc.rochester.edu/fields-center/protocols/index.cfm>. The muscle pathologic score (MPS) is based on review of four sections stained with Hematoxylin and Eosin as well as modified Gomori Trichrome staining. The 12 point pathologic score (0: normal, 12: end stage muscle) is based on assessment of variability in fiber size, percent of central nuclei, presence of active muscle fiber necrosis and regeneration and amount of fibrofatty replacement (<http://dev.mc.rochester.edu/fields-center/protocols/index.cfm>).

Cell lines and culturing

All primary myoblast and fibroblast cell cultures were obtained from the University of Rochester's FSHD biorepository (www.FieldsCenter.org). A list of used primary cell cultures, the 4q35 genotype and methylation status of D4Z4 is shown in **Table S1** for myoblast samples and in **Table S2** for fibroblast samples. Myoblast cell lines were cultured in DMEM/F-10 (#31550 GIBCO, Grand Island, NY) supplemented with 20% heat inactivated fetal bovine serum (FBS #10270 GIBCO), 1% penicillin/streptomycin (#15140 GIBCO), 10ng/ml basic rhFGF (#G5071 Promega, Madison, Wisconsin) and 1 μ M Dexamethasone (#D2915 SIGMA, St. Louis, Missouri). Fibroblast cell lines were cultured in DMEM/F-12 media supplemented with 20% heat inactivated fetal bovine serum (FBS #10270 GIBCO), 1% penicillin/streptomycin (#15140 GIBCO), 10mM HEPES (#15630 GIBCO), 1mM Sodium Pyruvate (#11360 GIBCO).

Chromatin immunoprecipitation

Chromatin was prepared from cells fixed with 1% formaldehyde according to a published protocol¹⁷. Every sample was independently studied twice. ChIP-grade antibodies against H3K4me2 (no. 39142) and against H3K9me3 (no. 39162) used in this study were purchased from Active Motif (Carlsbad, CA, USA). Normal rabbit serum was used to measure unspecific binding of proteins to beads. Immunopurified DNA was quantified with Q-PCR primer pair⁴ and quantitative PCR measurements were done with CFX96™ real time system using iQTM SYBRR Green Supermix. Relative enrichment values were calculated by subtracting the IgG ChIP values representing background from the ChIP values with the H3K9me3 or H3K4me2 antibodies.

Statistical analysis

Differences in ChCS between controls, FSHD1 and FSHD2 samples were statistically analyzed by one-way-ANOVA, followed by LSD multiple comparisons test between individual groups. Correlations were tested by calculating the Pearson R coefficient with corresponding p-values between the age corrected CSS and ChCS or between CSS and muscle pathology score for different groups of samples.

References

1. Padberg,G.W. Facioscapulohumeral disease. 1982. Leiden University. Thesis/Dissertation.
2. de Greef,J.C. et al. Clinical features of facioscapulohumeral muscular dystrophy 2. *Neurology* 75, 1548-1554 (2010).
3. van Overveld,P.G. et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet* 35, 315-317 (2003).
4. Zeng,W. et al. Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats in facioscapulohumeral dystrophy (FSHD). *PLoS. Genet* 7, e1000559 (2009).
5. de Greef,J.C. et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mutat.* 30, 1449-1459 (2009).
6. Lemmers,R.J. et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650-1653 (2010).
7. Snider,L. et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS. Genet.* 6, e1001181 (2010).
8. Snider,L. et al. RNA Transcripts, miRNA-sized Fragments, and Proteins Produced from D4Z4 Units: New Candidates for the Pathophysiology of Facioscapulohumeral Dystrophy. *Hum. Mol. Genet* 18, 2414-2430 (2009).
9. Lunt,P.W. et al. Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). *Hum Mol Genet* 4, 951-958 (1995).
10. Ricci,E. et al. Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of KpnI repeats at the 4q35 locus and clinical phenotype. *Ann Neurol* 45, 751-757 (1999).
11. Tawil,R. et al. Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann Neurol* 39, 744-748 (1996).
12. van Overveld,P.G. et al. Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. *Ann. Neurol.* 58, 569-576 (2005).
13. Ernst,J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473, 43-49 (2011).
14. Tonini,M.M.O. et al. Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). *Neuromuscular Disorders* 14, 33-38 (2004).
15. Frisullo,G. et al. CD8(+) T cells in facioscapulohumeral muscular dystrophy patients with inflammatory features at muscle MRI. *J. Clin. Immunol.* 31, 155-166 (2011).
16. Lemmers,R.J., van der Wielen,M.J., Bakker,E., Frants,R.R., & van der Maarel,S.M. Rapid and accurate diagnosis of facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* 16, 615-617 (2006).
17. Nelson,J.D., Denisenko,O., & Bomsztyk,K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.* 1, 179-185 (2006).

Acknowledgments

We would like to thank patients and their families for supporting our research by supplying us with biological material.

Disclosure of potential conflicts of interest

S.V.D.M., S.J.T. and R.T. are co-inventors on an FSHD patent application.

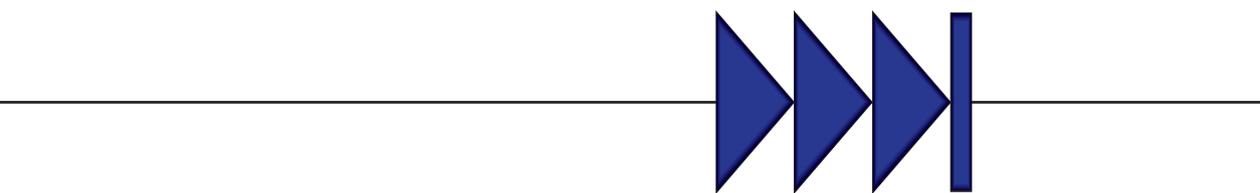
Supplemental material

Table S1: Data of fibroblast samples used in the study.

Table S2: Data of myoblast samples used in the study.

All supplemental material belonging to this chapter can be accessed through <http://goo.gl/x2rQiw> or by using the QR-code below.





**Chromatin remodeling of human
subtelomeres and TERRA promoters
upon cellular senescence:
commonalities and differences
between chromosomes.**

3

Peter E. Thijssen, Elmar W. Tobi, Judit Balog, Suzanne G. Schouten, Dennis Kremer, Fatiha El Bouazzaoui, Peter Henneman, Hein Putter, P. Eline Slagboom, Bastiaan T. Heijmans and Silvère M. van der Maarel

Thijssen et al. (2013), *Epigenetics*, 8:5, 512-521

Abstract

Subtelomeres are patchworks of evolutionary conserved sequence blocks, and harbor the transcriptional start sites for telomere repeat containing RNAs (TERRA). Recent studies suggest that the interplay between telomeres and subtelomeric chromatin is required for maintaining telomere function. To further characterize chromatin remodeling of subtelomeres in relation to telomere shortening and cellular senescence, we systematically quantified histone modifications and DNA methylation at the subtelomeres of chromosomes 7q and 11q in primary human WI-38 fibroblasts. Upon senescence, both subtelomeres were characterized by a decrease in markers of constitutive heterochromatin, suggesting relative chromatin relaxation. However, we did not find increased levels of markers of euchromatin or derepression of the 7q *VIPR2* gene. The repressed state of the subtelomeres was maintained upon senescence, which could be attributed to a rise in levels of facultative heterochromatin markers at both subtelomeres. While senescence-induced subtelomeric chromatin remodeling was similar for both chromosomes, chromatin remodeling at TERRA promoters displayed chromosome-specific patterns. At the 7q TERRA promoter, chromatin structure was co-regulated with the more proximal subtelomere. In contrast, the 11q TERRA promoter, which was previously shown to be bound by CCCTC-binding factor CTCF, displayed lower levels of markers of constitutive heterochromatin which did not change upon senescence, whereas levels of markers of facultative heterochromatin decreased upon senescence. In line with the chromatin state data, transcription of 11q TERRA but not 7q TERRA was detected. Our study provides a detailed description of human subtelomeric chromatin dynamics and shows distinct regulation of the TERRA promoters of 7q and 11q upon cellular senescence.

Introduction

The last decade has provided deeper knowledge on the stability and maintenance of the repetitive telomeric repeat structures at the ends of human chromosomes in biological processes such as cellular senescence. Non-transformed cells derived from mitotic tissues have a limited proliferative capacity, known as the Hayflick limit¹, and will eventually become senescent. Senescent cells exit the cell cycle, start secreting specific proteins that affect the microenvironment and display a number of senescence-associated markers, e.g. induction of cell cycle inhibitory proteins, senescence associated β -galactosidase (SA β -gal) and senescence associated heterochromatic foci (SAHF) formation². Cellular senescence is considered an anti-tumorigenic mechanism which in vivo may be relevant especially at higher ages, since senescent cells accumulate with age in tissues with high cell turnover. In vitro, it has been shown that progressive shortening of telomeres is a potent trigger to induce cellular senescence³⁻⁵.

Telomeres constitute the terminal structures of eukaryotic chromosomes and form a protective barrier against the incomplete replication of linear DNA. In mammals, telomeres are comprised of simple (TTAGGG)_n repeats folded into a heterochromatic loop structure, facilitated by the shelterin protein complex⁶. Numerous reports on the inverse correlation of telomere length in peripheral blood cells and donor age, age-related traits and the risk for various age-related diseases highlight the potential relevance of telomere regulation⁷⁻¹¹. It is as yet not clear however, whether telomere shortening causally contributes to these traits or represents a marker of cell division and senescence.

Although most studies on the relationship between telomere shortening and senescence focused on its potential direct consequences on chromosome stability, it may have downstream consequences, particularly on the subtelomere, that may in turn affect telomere function^{12,13}. Subtelomeres, the first non-TTAGGG sequences directly adjacent to the telomere repeat, are evolutionary conserved chromosome domains consisting of patchworks of sequence blocks with high inter- and intrachromosomal similarity^{14,15}. Subtelomeres are packed into constitutive heterochromatin, characterized by high levels of histone 3 lysine 9 trimethylation (H3K9me3) and CpG methylation¹³. A steep decrease of both markers is seen at subtelomeres upon drastic telomere shortening in mouse embryonic fibroblasts (MEFs) isolated from telomerase deficient mice, concomitant with an increase of markers for euchromatin (histone acetylation). This indicates a relative opening of the chromatin template at subtelomeres upon telomere shortening^{12,13}.

Several human studies focusing on aging, as well as age related disease like Parkinson and Alzheimer disease, showed that telomere length correlates with CpG methylation at subtelomeres and that the direction of this correlation is dependent on the disease conditions^{13,16-18}. A recent study in dyskeratosis congenita suggested an interaction between the proper maintenance of telomeres and the chromatin state of subtelomeres¹⁹. A more causal relation between dysregulation of subtelomeric chromatin and human disease is seen in the progressive muscular dystrophy facioscapulohumeral muscular

dystrophy (FSHD). In FSHD, decreased chromatin compaction by reduced levels of CpG methylation and H3K9me3 at the subtelomeric D4Z4 repeat, encoding the toxic DUX4 protein, leads to its aberrant transcription in muscle²⁰⁻²².

Chromatin alterations at subtelomeres may have other direct transcriptional consequences, as they harbor transcriptional start sites (TSS) for telomeric repeat containing RNAs (TERRA)²³. The transcription of these long non-coding RNAs is influenced by subtelomeric CpG methylation and myeloid/lymphoid or mixed lineage leukemia (MLL) mediated H3K4me3^{24, 25}. More recently, a role for CTCF and cohesin was established in controlling telomeric transcription²⁶. Telomere elongation by ectopic expression of telomerase, represses TERRA transcript levels and leads to increased levels of telomeric H3K9me3 without a concomitant change in the subtelomeric chromatin state²⁷. In yeast, recent data shows that inducing TERRA expression leads to accelerated telomere shortening, by facilitating exonuclease activity at transcribed chromosome ends^{28, 29}.

From the currently available data a feedback model is emerging in which telomere length regulates the epigenetic structure of subtelomeres, thereby possibly affecting TERRA expression levels, which in turn have an effect on telomere length, induction of senescence and chromatin state. The data supporting this model were obtained in diverse model systems, using ectopic telomere length modulation or TERRA regulation. However, chromosome-specific data on subtelomeric chromatin remodeling and TERRA promoter regulation upon cellular senescence by physiological telomere shortening in a human system is lacking. Therefore, we systematically interrogated the chromatin state of subtelomeres in senescing primary human WI-38 fibroblasts. The chromatin state was assessed by quantifying markers for euchromatin (histone acetylation) and constitutive (CpG methylation, H3K9me3) and facultative (H3K27me3, H3K36me3) heterochromatin along the subtelomere. Together with chromosome specific analysis of TERRA promoters and transcripts, a complex picture of both common and chromosome specific epigenetic changes at human subtelomeres upon cellular senescence emerges.

Results

Late passage WI-38 fibroblasts display a senescent phenotype.

To study subtelomere chromatin dynamics upon cellular senescence, we cultured WI-38 fibroblasts until they reached a senescent phenotype. Upon senescence, we observed overall telomere shortening evidenced by a shift in the smear in a TRF analysis (**Fig. 1A**). Subsequent qPCR analysis revealed a 65% reduction in telomere copy number comparing early passage cycling cells to late passage senescent cells (E vs L, **Fig. 1B**). Upon cellular senescence, expression of the cell cycle inhibitory protein p16 was induced approximately 2 fold (**Fig. 1C, D**). Moreover, we observed that late passage cells display high levels of SA β -gal activity in more than 95% of the cells, whereas in early passage only sporadic cells show staining (**Fig. S1**). Taken together, late passage WI-38 fibroblasts underwent replication induced telomere shortening and showed a senescent phenotype, in contrast to early passage cells.

Decrease in markers of constitutive heterochromatin at subtelomeres upon senescence

The subtelomeres of human chromosome arms 7q and 11q are consisting of a single copy sequence which facilitates a chromosome specific analysis of histone modification levels at multiple loci with increasing distance to the telomere (**Fig. 2A**). We assessed H3K9me3 levels at 6 loci on both subtelomeres and observed a variable 2-3 fold decrease of this modification at all loci upon cellular senescence (**Fig. 2B-C**). The decrease of H3K9me3 was not detected at the repressed, non-subtelomeric CT47 macrosatellite repeat array³⁰, arguing against a general loss at heterochromatic loci (**Fig. S2A**). At the promoter of the actively transcribed *GAPDH* gene, no change in the low level of

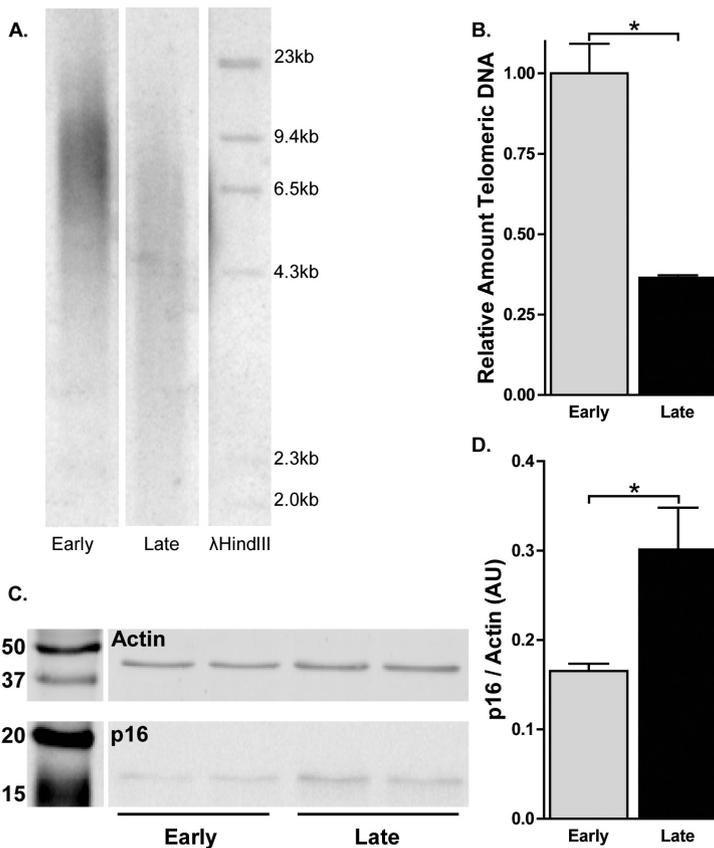


Fig. 1. Late passage WI-38 fibroblasts underwent telomere shortening and display a senescent phenotype.

A) TRF analysis on genomic DNA isolated from early (E) passage and late (L) passage cells. DNA was digested with *RsaI* and *AluI* and reduction of the average telomere length is seen by the shift in the smear of the undigested telomeric DNA. **B)** Normalized relative quantification of telomeric copy number in E and L passage cells. Telomeric copy number was quantified by qPCR, relative to the *36B4* single copy locus and normalized to E passage cells. Error bars indicate SEM of a triplicate measurement. **C)** Western blot analysis in of p16 and actin expression in E and L passage cells followed by **D)** relative quantification, showed increased expression of p16 in L passage cells. Error bars indicate stdev of two biological replicates, asterisks indicate a p-value < 0.05 based on a Student's t-test.

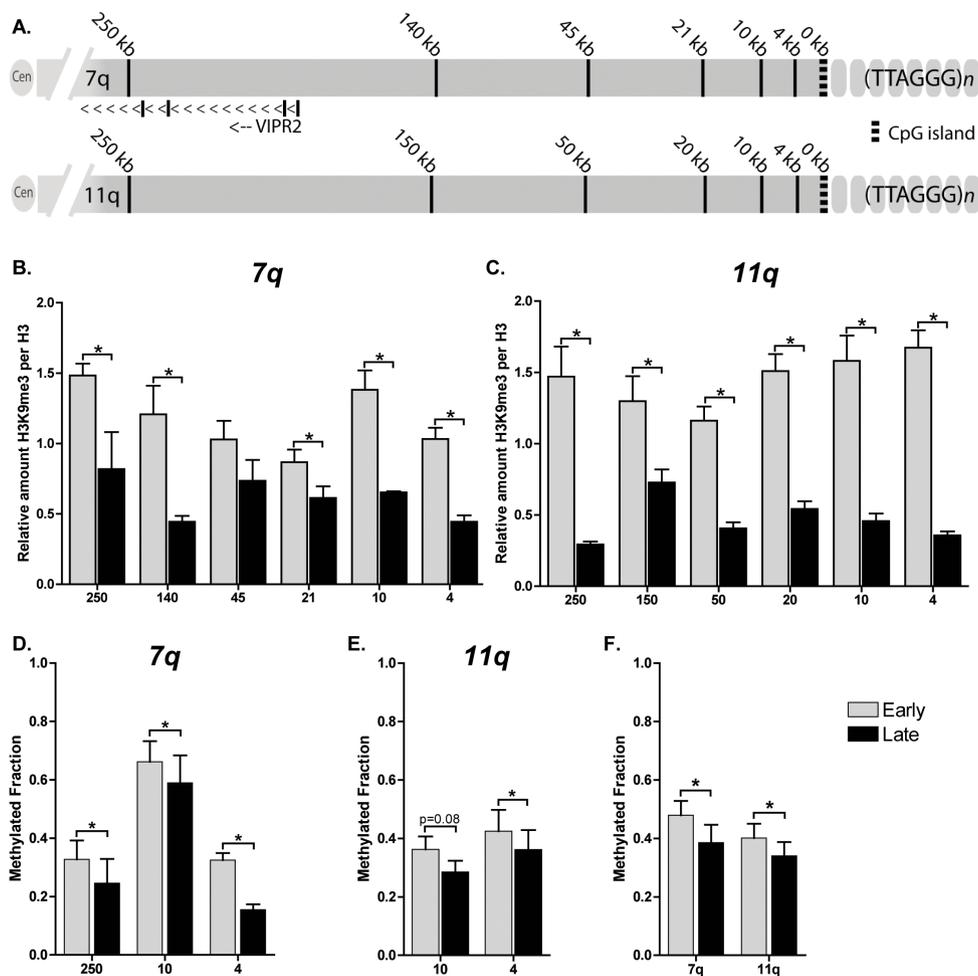


Fig. 2. Decreased levels of markers of constitutive heterochromatin at subtelomeres upon senescence.

A) Location of the generated primer pairs along the subtelomeres of chromosome 7q and 11q. **B)** Relative quantification using ChIP-qPCR of H3K9me3 along the subtelomeres of 7q and **C)** 11q shows a decrease at all loci upon senescence. H3K9me3 levels are corrected for IgG background and normalized to the relative amount of H3: $(Ct_{IgG} - Ct_{H3K9me3}) / (Ct_{IgG} - Ct_{H3})$. Error bars indicate SEM of at least duplicate qPCR measurements, asterisks indicate a p-value < 0.05 based on a Student's t-test. **D)** EpiTYPER based quantification of CpG methylation along the subtelomere of 7q and **E)** 11q showed a decreased DNA methylation at all sites measured. The mean methylation of multiple CpGs within the indicated probes is displayed. **F)** Combined mean methylation levels of all subtelomeric probes at 7q and 11q showed a decreased DNA methylation along both subtelomeres. DNA methylation was quantified in triplicate, p-values are indicated and * indicates a p-value < 0.05 based on UNIANOVA analysis.

H3K9me3 was observed upon cellular senescence (**Fig. S2B**).

Mean methylation levels of sporadic CpGs at the regions under study showed variable levels at different loci, but at 4 out of 5 tested loci a significant decrease of 8-17% upon senescence ($P < 0.05$) was observed (**Fig. 2D, E**). The 5th locus, at 10kb from the 11q telomere, showed a similar trend ($P = 0.08$). Analysis of individual CpGs supported this observation (**Fig. S3**). Analysis per chromosome, by calculating the mean methylation level over multiple amplicons, revealed a significant average decrease of ~10% at 7q and ~6% at 11q ($p < 0.05$, **Fig. 2F**). Overall, we observed a reduction in markers of constitutive heterochromatin at two human subtelomeres upon cellular senescence, indicating a relative opening of the chromatin template.

No increased levels of euchromatin markers at subtelomeres upon senescence

In *Terc*^{-/-} mice MEFs, the decrease of markers of constitutive heterochromatin with drastic telomere shortening is accompanied by increased histone acetylation levels.¹² In our human cell system, H3K9ac was virtually absent at subtelomeres both in the early passage and upon cellular senescence (**Fig. S4**). In contrast, H4K16ac was detected, but a decrease of at least 2 fold (**Fig. 3A-B**) was observed upon senescence. The latter was however not restricted to subtelomeres, as we observed a comparable decrease of this modification at the heterochromatic CT47 macrosatellite repeat array (**Fig. S2A**). In the context of the promoter of the actively transcribed *GAPDH* locus a relatively small increase of H4K16ac was observed (**Fig. S2B**). Altogether, we did not find evidence that the decreased levels of constitutive heterochromatin markers at subtelomeres upon senescence are accompanied by increased levels of markers of euchromatin.

Increased levels of H3K27me3 and H3K36me3 may compensate for loss of markers of constitutive heterochromatin

We wondered why the decrease of markers of constitutive heterochromatin marks did not go together with an increased level of euchromatin marks. Therefore, we tested whether chromatin repression was actually maintained, by involvement of different repressive mechanisms instead. As the Polycomb mediated H3K27me3 is considered a marker for both facultative and constitutive heterochromatin, we assessed its levels at the subtelomeres upon senescence. H3K27me3 was indeed detected at subtelomeres in early passage cells and we observed an up to 2 fold increase of this modification upon cellular senescence at 7 out of 12 loci on 7q and 11q (**Fig. 3C-D**). This suggests that a Polycomb-enforced chromatin repression occurs at these loci upon cellular senescence. The increase of H3K27me3 was most pronounced at 250 kb from the telomere on chromosome 7q, which is located in the fourth intron of the *VIPR2* gene. Transcripts emanating from this gene were not detected in both passages (data not shown), again indicating the repressed chromatin state is indeed preserved in senescing cells. At the heterochromatic CT47 macrosatellite repeat array and the euchromatic *GAPDH* promoter we did not see increased levels of H3K27me3 (**Fig. S2**).

Recently it was shown that H3K36me3, a marker associated with actively transcribed gene bodies, is associated with both facultative and constitutive heterochromatin regions in mouse³¹. At the subtelomeres of 7q and 11q, we detected H3K36me3 with

an increase of at least 2 fold upon senescence at both subtelomeres (**Fig. 3E-F**). All subtelomeric loci displayed higher levels of this marker, but the largest increase was again detected within *VIPR2*. Increased H3K36me3 was not restricted to subtelomeres, as similar increases were detected at the *GAPDH* promoter and CT47 macrosatellite repeat array (**Fig. S2**). Taken together, these data suggest a relative opening by reduction of constitutive heterochromatin markers is accompanied by increased levels of two

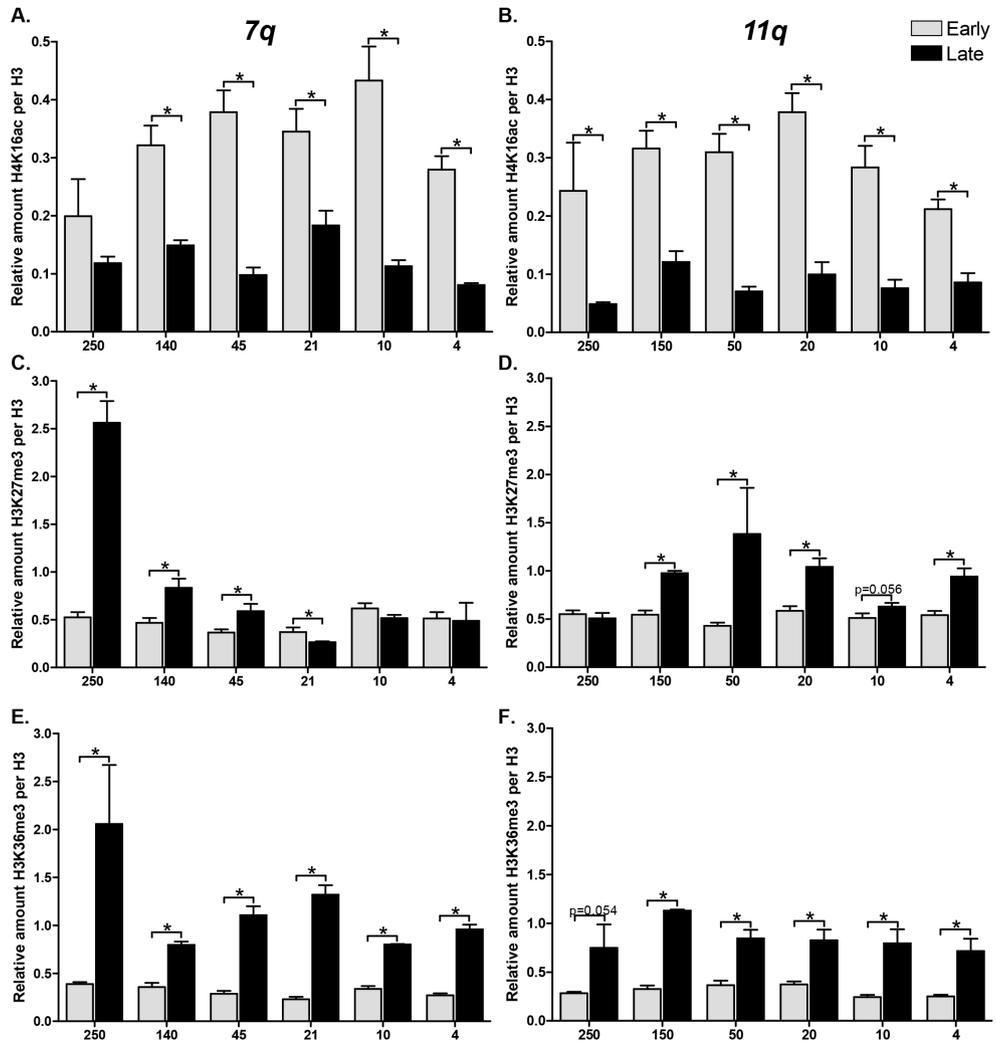


Fig. 3. Relative quantification of subtelomeric H4K16ac, H3K27me3 and H3K36me3 upon senescence Relative quantification using ChIP-qPCR of H4K16ac (**A, B**), H3K27me3 (**C, D**) and H3K36me3 (**E, F**) at increasing distance from the telomere at the subtelomeres of 7q (**A, C, E**) and 11q (**B, D, F**). H4K16ac is decreased at all loci and H3K36me3 shows an increase at both subtelomeres. H3K27me3 shows a more varied pattern; an increase is observed at 7 out of 12 loci. All values are corrected for IgG background and normalized to the relative amount of H3 ($(Ct_{\text{IgG}} - Ct_{\text{modification}})/(Ct_{\text{IgG}} - Ct_{\text{H3}})$), error bars indicate SEM of duplicate qPCR measurements, asterisks indicate a p-value < 0.05 based on a Student's t-test.

heterochromatin markers associated with different repressive mechanisms, which may maintain the repressed subtelomeric chromatin state upon senescence.

Since both model subtelomeres under study are of single copy nature, we wondered to what extent the observed effects would apply to other, more repetitive subtelomeric loci. To that end, we measured relative abundance of histone modifications at D4Z4, a subtelomeric repeat structure present at chromosomes 4q and 10q, that has been

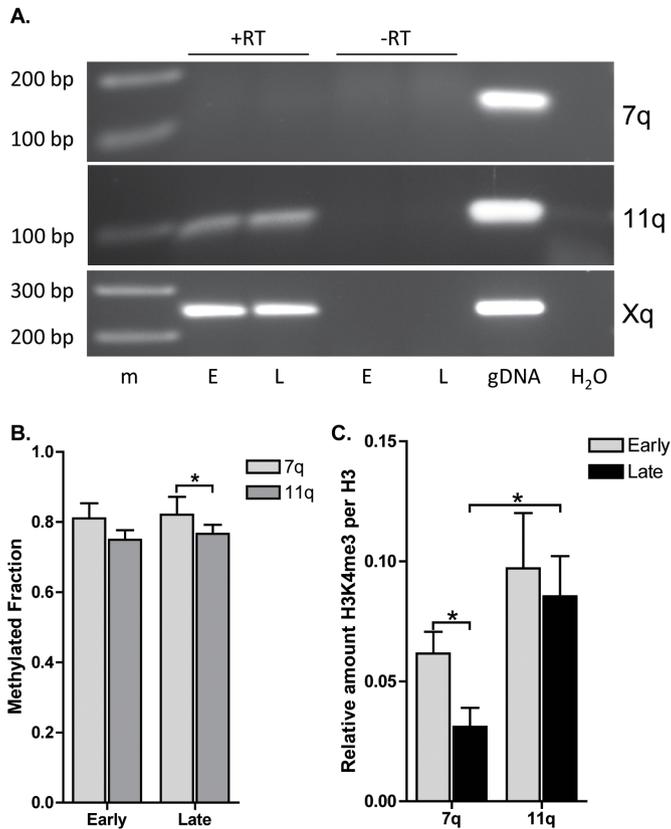


Fig. 4. TERRA promoter analysis reflects transcriptional activity at 7q and 11q irrespective of senescence

A) RT-PCR analysis of 7q, 11q and Xq specific TERRA transcripts allowed detection of 11q, but not 7q specific transcripts. Xq serves as an internal control. –RT: cDNA synthesis in the absence of reverse transcriptase to control for DNA background. gDNA: positive control for the PCR on a WI-38 genomic DNA extract. **B)** Epityper based quantification of TERRA promoter CpG methylation at 7q and 11q showed higher levels of methylation at 7q pTERRA compared to 11q pTERRA. The mean methylation of multiple CpGs within the indicated probes is displayed. Upon senescence TERRA promoter methylation at both 7q and 11q does not change. Indicated p-values ($* < 0.05$) were obtained using an UNIANOVA analysis of triplicate measurements. **C)** Relative quantification by ChIP-qPCR of H3K4me3, marking active promoters, showed higher levels at 11q pTERRA compared to 7q pTERRA, irrespective of senescence. All values are corrected for IgG background and are normalized to the relative amount of H3 ($Ct_{IgG} - Ct_{H3K4me3} / (Ct_{IgG} - Ct_{H3})$). Error bars indicate SEM of duplicate qPCR measurements, asterisks indicate a p-value < 0.05 based on a student's t-test.

extensively studied in the context of FSHD^{20, 22}. This showed that the described chromatin remodeling occurs on other subtelomeres as well, as we observed similar changes upon senescence compared to the subtelomeres of 7q and 11q (Fig. S5).

Specific chromatin regulation of 7q and 11q TERRA promoters upon senescence

The distal ends of subtelomeres harbor the TSS of TERRA transcripts. Recent data showed the involvement of the insulator protein CTCF at the TERRA promoter of 11q, but not at 7q²⁶. Indeed, querying the WI-38 CTCF ChIP-seq tracks available from the UCSC genome browser, revealed CTCF binding just proximal to the TERRA promoter site at 11q, but not at 7q. We sought evidence for transcriptional activity of 7q and 11q TERRA in our cell system. Both in early passage and senescent cells we detected 11q TERRA transcripts, however, we could not detect TERRA transcripts emanating from 7q in both cell passages (Fig. 4A). Previously identified Xq transcripts²⁴ were detected in both early and late passage cells and served as a control (Fig. 4A).

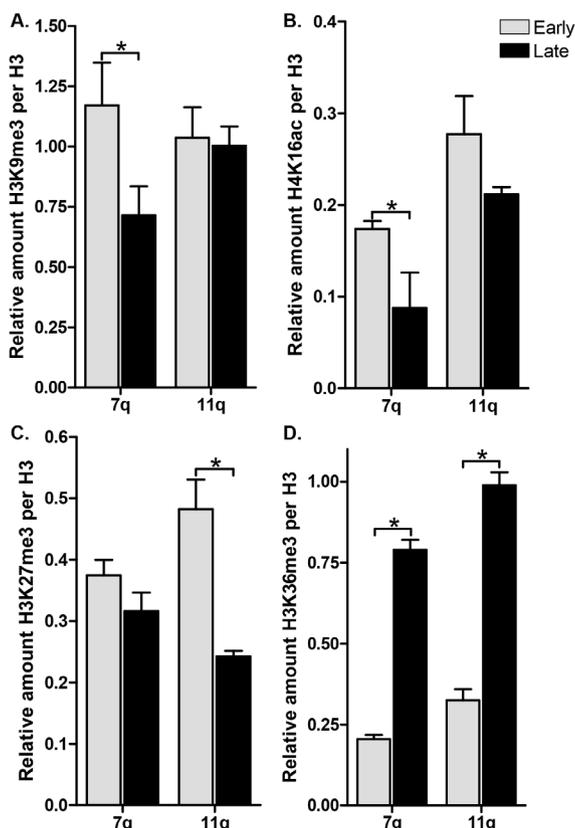


Fig. 5. Distinct chromatin remodeling at TERRA promoters of 7q and 11q upon senescence

TERRA promoter analysis showed distinct regulation of **A)** H3K9me3, **B)** H4K16ac, **C)** H3K27me3 and **D)** H3K36me3 between 7q and 11q. The changes at the 7q TERRA promoter reflect the more proximal subtelomere, whereas the 11q TERRA promoter regulation is distinct from the more proximal subtelomere. All values are corrected for IgG background and are normalized to the relative amount of H3 ($(Ct_{IgG} - Ct_{modification}) / (Ct_{IgG} - Ct_{H3})$). Error bars indicate SEM of duplicate qPCR measurements, asterisks indicate a p-value < 0.05 based on a student's t-test.

In line with this, the mean DNA methylation in senescent cells at the CpG dense TERRA promoter of 7q showed a significantly ($p < 0.05$) higher methylation level than its 11q counterpart (~81% at 7q vs ~75% at 11q). Although not significant, we observed a similar trend in early passage cells (**Fig. 4B**). Moreover, promoter-associated H3K4me3 levels, previously shown to be involved in TERRA regulation²⁵, were 40-60% lower at 7q compared to 11q (**Fig. 4C**). Global TERRA levels are reported to decrease upon senescence²⁵, which is in line with lower abundance of H3K4me3 at both 7q and 11q (**Fig. 4C**). CpG methylation levels did not change upon senescence, although some individual CpGs changed in both directions (**Fig. S6**).

To further determine the chromatin regulation of the TERRA promoters upon senescence, we quantified histone modification levels at both 7q and 11q. 7q pTERRA displayed similar chromatin dynamics upon senescence as the more proximal 7q subtelomere: reduced H3K9me3 and H4K16ac levels with an increase of H3K36me3 upon senescence. H3K27me3 showed a small decrease, however the described increase of this mark was mostly seen at greater distance to the telomere (**Fig. 5A-D**).

In contrast, at 11q, bound by CTCF, we observed differential regulation of pTERRA compared to the proximal subtelomere. In early passage cells, the relative amount of H3K9me3 at pTERRA (1.04) was lower compared to more proximal sites (~1.44 on average), but we did not observe a decrease upon senescence (**Fig. 5A**). Strikingly, H3K27me3 showed a 2 fold decrease upon senescence, contrary to the more proximal subtelomere, where a change in opposite direction was observed (**Fig. 5C**). H4K16ac and H3K36me3 levels at pTERRA are comparable to the more proximal sites, although the levels of both markers remained higher upon senescence (**Fig. 5B, D**). Altogether, we observed that upon senescence, the chromatin remodeling of the 7q TERRA promoter is similar to the proximal subtelomere. The TERRA promoter at 11q however, showed a distinct chromatin remodeling upon senescence when compared to the proximal 11q subtelomere.

Discussion

In this study, we systematically interrogated the changes in chromatin structure of human subtelomeres upon cellular senescence. Previous studies, in which ectopic telomere length modulation resulted in a large contrast in telomere length, reported opposing data on the effect of telomere length, a known trigger for cellular senescence, and subtelomeric chromatin state^{12, 27}. In our model, using primary human fibroblasts, we observed a senescence-associated reduction of constitutive heterochromatin markers, which is accompanied by a gain of facultative repressive marks. Our data indicate that the repressed state is preserved during senescence as we found no increased levels of markers of euchromatin or evidence for derepression of the only annotated gene in the studied region. Furthermore, we observed distinct regulation of two TERRA promoter sites at subtelomeres of 7q and 11q. These data complement with recent data showing the involvement of CTCF and cohesin in chromosome-specific epigenetic regulation of TERRA²⁶.

Upon cellular senescence, we observed a decrease in two markers of constitutive heterochromatin, H3K9me3 and CpG methylation, suggesting a relative relaxation of the chromatin structure. Reduced levels of these markers at subtelomeres were previously reported upon drastic telomere shortening in MEFs of a telomerase deficient mouse model¹². In contrast, telomere elongation by ectopic telomerase expression in human cell lines did not result in subtelomeric epigenetic changes²⁷. However, the contrast in telomere length by ectopic telomerase expression is small, which may better reflect human physiological conditions, compared to telomerase deficient mice. Although the exact nature of the mechanism by which telomere length controls subtelomeric heterochromatin formation remains enigmatic, our data show that physiological telomere shortening and senescence signaling result in decreased levels of constitutive heterochromatin marks at subtelomeres.

Upon reduced H3K9me3 and CpG methylation, a concomitant increase of markers of euchromatin (histone acetylation) is seen with drastic telomere shortening in *Terc*^{-/-} MEFs¹². In our model, we did not observe a similar effect upon senescence. H3K9ac levels were low in both passages and we observed decreased levels of H4K16ac upon cellular senescence. Subtelomeric H4K16ac was previously shown to be higher upon aging in yeast, as a consequence of reduced Sir2 expression³². Global H4K16 deacetylation, on the contrary, was shown to be involved in the DNA damage response and associates with cellular senescence of MEFs in a mouse model for the Hutchinson Gilford premature ageing syndrome³³. Our data show that the reduced levels of H4K16ac are not restricted to subtelomeres and are likely to be a genome-wide phenomenon, in concordance with its role in the DNA damage response and senescence.

We hypothesized that further chromatin activation upon decreased levels of markers of constitutive heterochromatin may be prevented by increased levels of the Polycomb-dependent repressive H3K27me3 modification. Both H3K9me3 and H3K27me3 were previously reported to change globally upon cellular senescence in human primary fibroblasts³⁴. O'Sullivan et al. describe decreased H3K9me3 and increased H3K27me3 levels, which is in accordance with our observations at subtelomeres. Recent data, showing that the genome wide distribution of these markers does not change with RAS induced cellular senescence in human fibroblasts, suggest a different effect of oncogene induced- compared with replicative senescence³⁵.

H3K36me3 is mostly studied in the context of actively transcribed genes, but was recently suggested to be enriched in constitutive and facultative heterochromatin in mouse cells³¹. Indeed, we find considerable levels of this marker at both subtelomeres and at the silenced CT47 locus. In yeast, H3K36me3 recruits histone deacetylase complexes (HDACs) and thereby ensures repression of intragenic cryptic promoters³⁶⁻³⁸. It may be speculated that a similar mechanism exists in which subtelomeric sites become transcriptionally derepressed upon decreased levels of constitutive heterochromatin, eventually leading to deposition of H3K36me3 and recruitment of HDACs. In this scenario, higher levels of H3K36me3 have a repressive effect on the chromatin, maintaining the silenced state upon a decrease in H3K9me3 and CpG methylation. However, the

increased H3K36me3 levels were not specific for subtelomeres and in our model we cannot disentangle the order of events to prove this concept. More understanding of the context dependent functionality of the H3K36me3 mark is therefore desired.

Subtelomere CpG methylation and histone modification levels have been subject of study in the context of several human diseases and aging. Throughout these studies, a positive correlation between telomere length and subtelomeric CpG methylation was observed, which was depended on different disease states and integrity of the telomere maintenance pathway¹⁶⁻¹⁹. Indeed, with reduced telomere length and induction of senescence we observed reduced levels of CpG methylation at subtelomere. In the progressive muscular dystrophy FSHD, the subtelomeric D4Z4 repeat array is characterized by reduced CpG methylation and H3K9me3 levels, compared to controls²⁰⁻²². In FSHD cell cultures, telomere length or senescence markers have not been studied. Our data may have implications for studying the epigenetic regulation of D4Z4, as we show that the replicative history of cell cultures can affect the chromatin structure of subtelomeres, also of the subtelomeric D4Z4 repeat array.

Since subtelomeres predominantly consist of repetitive sequences and subtelomeres of different chromosomes share highly homologous sequence blocks^{14, 15}, we selected the single copy subtelomeres of 7q and 11q as a model. This allowed us to systematically screen the subtelomere at increasing distance to the telomere to identify possible position effects. However, we did not find a clear relation between the distance to the telomere and the changes in subtelomeric chromatin structure. The observed chromatin changes at the subtelomeres of chromosomes 7q and 11q and the D4Z4 repeat array at 4q and 10q were highly similar and possibly reflect a general phenomenon of subtelomeric chromatin remodeling upon senescence. We cannot, however, rule out specific chromatin regulation at other subtelomeres, or at different locations at the subtelomeres under study. Moreover, we have only studied the effect of senescence in a single model system for cellular senescence. To assess whether our observations are in general associated with senescence, additional models should be studied.

A striking exception to the commonalities between the studied chromosomes was the distinct regulation of the TERRA promoter at the subtelomeres of 7q and 11q. TERRA expression is regulated by specific CpG island motifs which have been identified on at least 20 different chromosome arms²⁴. This specific CpG island is absent at both 7q and 11q, however CpG dense sequences are present just proximal to the telomere of both chromosomes. We demonstrated that DNA methylation at these CpG islands does not change upon senescence, in spite the fact that TERRA transcripts were shown to be downregulated upon senescence²⁵. The H3K4 methyltransferase MLL has been shown to associate with telomeres and regulate TERRA transcription²⁵. Low but detectable levels of H3K4me3 were detected at the TERRA promoters of 7q and 11q and we detected higher levels at 11q than at 7q, irrespective of senescence. Both CpG methylation and H3K4me3 levels are reflected in transcriptional activity, as we only could detect 11q transcripts. As TERRA has been shown to regulate telomere length, chromosome specific differences in TERRA regulation could offer an explanation for the observed

allelic variation of telomere length in senescent human cells³⁹.

Our data again emphasize that a chromosome arm specific analysis is needed considering the regulation of TERRA²⁶. We showed that the 11q TERRA promoter is differentially regulated compared to the more proximal subtelomere, in contrast to the TERRA promoter of 7q. This is in line with recently published data, where involvement of CTCF and cohesin in TERRA regulation was shown at 11q, but not at 7q²⁶. Next to its described role in TERRA transcription, our data suggests that the binding of CTCF proximal to the TERRA promoter results in the insulation and specific chromatin regulation of the TERRA promoter upon senescence. It may be hypothesized that TERRA promoters bound by CTCF, are protected from the effects of telomere shortening and senescence signaling on the chromatin structure of subtelomeres, which would ensure proper regulation of TERRA transcription.

In conclusion, we show that human subtelomeres undergo extensive chromatin remodeling upon cellular senescence. A decrease in markers of constitutive heterochromatin does not lead to subtelomeric derepression but is accompanied by increased levels of transcriptional repressive modifications. We observed a strong overlap between the two subtelomeres under study, however, with respect to TERRA promoters, we showed chromosome specific remodeling occurs.

Materials and Methods

Cell culture & β -galactosidase assay

WI-38 human fetal lung fibroblasts at two different population doublings (15, early and 34, late) were obtained from Corriell (Corriell Cell Repositories, Camden NJ, USA). Early passage cells were expanded for 3–4 passages in DMEM F12 (31331) supplemented with 20% heat inactivated Fetal Calf Serum, 1% pen-strep, 1% sodium pyruvate and 1% HEPES (all Invitrogen Life Technologies, Bleiswijk, The Netherlands) at 37^o C, 5% CO₂. Senescent cells were obtained by expanding PDL 34 cells until a non-proliferative state was reached. Cells were then kept in culture for an additional period of 2-3 weeks and to confirm senescence, activation of β -galactosidase was assessed as described⁴⁰. DNA, RNA, chromatin and protein were isolated in parallel. Both early passage and late passage cells were expanded, harvested, and examined in three independent experiments of which one complete set is shown.

Protein isolation, Western blot & quantification of p16

Cells were washed twice in 1 x PBS and after removal cells were lysed in RIPA buffer: 20 mM TEA, 0.14 M NaCl, 0.1% DOC, 0.1% SDS, 0.1% triton X-100. 1x Complete EDTA free protease inhibitor cocktail and 1x Phosstop phosphatase inhibitor cocktail (both Roche, Mannheim, Germany) and sheared by passing through a 29G needle. Soluble protein content was measured by standard Pierce BCA analysis (Thermo Scientific, Etten-Leur, The Netherlands) and 15 μ g protein was loaded on a standard 15% SDS PAGE. After protein transfer, membranes were blocked and incubated o/n at 4^o C with antibodies against p16 (1:500, sc-56330, Santa Cruz biotechnology, Santa Cruz CA, USA) and actin (1:1000, A2066, Sigma, Zwijndrecht, The Netherlands). Detection and relative

quantification were done using the Odyssey system (V3.0, LI-COR biosciences, Lincoln NE, USA).

Telomere length analysis by southern blot and qPCR

Genomic DNA was isolated using a standard salting out method. Southern blot based telomere length analysis was done as described before⁴¹. In brief, 3 µg genomic DNA was o/n digested with RsaI and AluI (Thermo scientific) at 37° C and size separated on a 0.9% TAE agarose gel. DNA was denatured by incubating the gel in 0.4 M NaOH, 0.6 M NaCl for 30 min, nicked and crosslinked by UV light and subsequently transferred to a Hybond XL membrane (GE Healthcare, Diegem, Belgium). The membrane was hybridized, washed, scanned with a Storm 820 phosphorimager (GE Healthcare) and analyzed using imagequant software (v2003.03. GE Healthcare). qPCR analysis of telomeric copy number was done as described before⁴² on the CFX96™ thermal cycler using iQ SYBR Green Supermix (both Bio-rad, Veenendaal, The Netherlands). Relative telomere length was calculated using the single copy *36B4* locus as a reference and subsequently normalized using the CFX software (v2.0, Bio-rad).

Chromatin Immunoprecipitation

ChIP experiments were based on the protocol described by Nelson et al. with some modifications⁴³. In brief, cells were crosslinked in 1% formaldehyde (Merck-Millipore, Amsterdam, The Netherlands) for 10 minutes and the reaction was quenched for 5 minutes in 125 mM glycine. Cross linked cells were lysed and chromatin was sheared in a sonicator bath (Bioruptor UCD-20, Diagenode, Liège, Belgium) for 4 consecutive rounds of 10 minutes at maximum output and 15 seconds on/off cycles. Shearing was analyzed by phenol-choloroform extraction of DNA and agarose gel electrophoresis. All chromatin samples had a DNA size range between 200 – 2000 bp. 3 µg (DNA amount) chromatin was precleared with blocked sepharose A beads (GE healthcare) and incubated overnight at 4° with antibodies: αH3 (ab1791, 2µl/rxn, Abcam, Cambridge, UK), αH3K9me3 (39161, 5 µl/rxn, Active Motif, La Hulpe, Belgium) αH3K27me3 (#17-622. 5 µl/rxn, Merck-Millipore) αH4K16ac (39167, 4µl/rxn, Active Motif), αH3K36me3 (2 µl/rxn, pAb-058-050, Diagenode), αH3K4me3 (#17-614, 3 µl/rxn, Merck-Millipore) and total IgG (5 µl/rxn, Merck-Millipore). IP was done with 20 µl blocked sepharose A beads/rxn for 90-120 minutes at 4°C. Beads were washed according to the online available Millipore ChIP protocol (<http://www.millipore.com/techpublications/tech1/mcproto407>). DNA was isolated using Chelex 100 resin (Bio-rad) and diluted 1:1 for Q-PCR analysis.

(q)PCR analysis of human subtelomeres

To circumvent potential PCR problems posed by the duplicated and repetitive nature of subtelomeres, we exploited the single copy nature of the subtelomeres of chromosomes 7q and 11q. To quantify histone modifications, we generated primer pairs using Primer3Plus using qPCR settings⁴⁴ on both chromosome arms ranging from as close as <1 kb from the telomere, up to 250 kb from the telomere (**Fig. 2a**). All primer pairs were tested for chromosome specificity using a monochromosomal DNA panel (Coriell). We succeeded to generate specific primers ranging from directly adjacent to

the telomere up to 250kb from the telomere on both the subtelomeres of 7q and 11q. Primer sequences are listed in **table S1** and annealing temperatures are indicated. qPCR analysis was done in duplicate using iQ SYBR Green Supermix (Bio-Rad) on the MyIQ thermal cycler or the CFX96tm real time PCR detection system (Bio-Rad) using 4 μ l 1:1 diluted ChIP DNA per reaction. Relative quantities of histone modifications were calculated by taking the fold enrichment relative to the IgG background, normalized to input or H3 enrichment: $(Ct_{IgG} - Ct_{modification}) / (Ct_{IgG} - Ct_{H3})$. Representative data of 1 of the three culture replicates is shown and error bars indicate SEM of the normalized duplicate q-PCR values.

CpG methylation analysis

CpG islands at TERRA promoters were identified using the CpGplot tool available at <http://www.ebi.ac.uk/Tools/emboss/cpgplot>. CpG methylation levels were determined using mass spectrometry based EpiTyper assays (Sequenom, San Diego CA, USA) as described before⁴⁵. In short, DNA is converted with bisulfite and PCR amplified, transcribed to RNA, cleaved by RNaseA and the resulting methylated and unmethylated fragments were quantified by mass spectrometry. All measurements were performed in triplicate. Subtelomeric probe sequences are listed in **table S1**. The DNA methylation was entered as the dependent to an UNIANOVA (SPSS 18.0), as it accounts for the correlated nature of adjacent CpG sites. Variables defining the cell passage (early and late), the CpG site, the individual triplicates and the genomic location were entered as fixed effects. Data is shown for a single culture experiment, all p-values reported are two-tailed. For **Figs. 2D-F** and **4B**, the mean methylation level of multiple CpGs within a probe and, for **Fig. 2F**, multiple probes per chromosome is displayed.

RNA isolation, TERRA cDNA synthesis and PCR detection

Cells were lysed in QIAzol lysis reagent and subsequently RNA was isolated using the RNeasy mini kit according to manufacturer's instructions (both Qiagen, Venlo, The Netherlands). RNA integrity was confirmed (RIN > 9) by RNA 6000 Nano lab on chip analysis (Agilent technologies, Waldbronn, Germany) and 2 μ g of total RNA was used for TERRA specific cDNA synthesis as described²³, using the the SuperScript[®] III First-Strand Synthesis System (Life Technologies). PCR analysis was carried out using Phusion[®] High Fidelity DNA polymerase according to manufacturers conditions. PCR conditions: initial denaturing: 95^o C, 3' followed by 40 cycles of 30" 95^o C, 30" 60^o C (11q, Xq) / 61^o C (7q), 30" 72^o C and a final extension step of 10' 72^o C. Primer sequences are indicated in **table S1**. PCR products were analyzed by standard 1.5-2% TBE agarose gel electrophoresis and visualized using the OptiGo 750 imaging system with Proxima AQ-4 software (both Isogen life science, De Meern, The Netherlands).

References

- Hayflick, L. The limited in vitro lifetime of human diploid cell strains. *Exp. Cell Res.* 37, 614-636 (1965).
- Bernardes de, J.B. & Blasco, M.A. Assessing cell and organ senescence biomarkers. *Circ. Res.* 111, 97-109 (2012).
- Harley, C.B., Futcher, A.B., & Greider, C.W. Telomeres shorten during ageing of human fibroblasts. *Nature* 345, 458-460 (1990).
- Allsopp, R.C. Models of initiation of replicative senescence by loss of telomeric DNA. *Exp. Gerontol.* 31, 235-243 (1996).
- Vaziri, H. & Benchimol, S. Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span. *Curr. Biol.* 8, 279-282 (1998).
- Palm, W. & de, L.T. How shelterin protects mammalian telomeres. *Annu. Rev. Genet.* 42, 301-334 (2008).
- Demissie, S. et al. Insulin resistance, oxidative stress, hypertension, and leukocyte telomere length in men from the Framingham Heart Study. *Aging Cell* 5, 325-330 (2006).
- Fitzpatrick, A.L. et al. Leukocyte telomere length and cardiovascular disease in the cardiovascular health study. *Am. J. Epidemiol.* 165, 14-21 (2007).
- Zee, R.Y., Castonguay, A.J., Barton, N.S., Germer, S., & Martin, M. Mean leukocyte telomere length shortening and type 2 diabetes mellitus: a case-control study. *Transl. Res.* 155, 166-169 (2010).
- Willeit, P., Willeit, J., Kloss-Brandstatter, A., Kronenberg, F., & Kiechl, S. Fifteen-year follow-up of association between telomere length and incident cancer and cancer mortality. *JAMA* 306, 42-44 (2011).
- Aubert, G. & Lansdorp, P.M. Telomeres and aging. *Physiol Rev.* 88, 557-579 (2008).
- Benetti, R., Garcia-Cao, M., & Blasco, M.A. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat. Genet.* 39, 243-250 (2007).
- Blasco, M.A. The epigenetic regulation of mammalian telomeres. *Nat. Rev. Genet.* 8, 299-309 (2007).
- Linardopoulou, E.V. et al. Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature* 437, 94-100 (2005).
- Ambrosini, A., Paul, S., Hu, S., & Riethman, H. Human subtelomeric duplicon structure and organization. *Genome Biol.* 8, R151 (2007).
- Guan, J.Z., Guan, W.P., Maeda, T., & Makino, N. The Subtelomere of Short Telomeres is Hypermethylated in Alzheimer's Disease. *Aging Dis.* 3, 164-170 (2012).
- Maeda, T., Guan, J.Z., Higuchi, Y., Oyama, J., & Makino, N. Aging-related alterations of subtelomeric methylation in sarcoidosis patients. *J. Gerontol. A Biol. Sci. Med. Sci.* 64, 752-760 (2009).
- Maeda, T., Guan, J.Z., Oyama, J., Higuchi, Y., & Makino, N. Aging-associated alteration of subtelomeric methylation in Parkinson's disease. *J. Gerontol. A Biol. Sci. Med. Sci.* 64, 949-955 (2009).
- Gadalla, S.M. et al. The relationship between DNA methylation and telomere length in dyskeratosis congenita. *Aging Cell* 11, 24-28 (2012).
- Balog, J. et al. Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. *Epigenetics.* 7, 579-584 (2012).
- van Overveld, P.G. et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* 35, 315-317 (2003).
- Zeng, W. et al. Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats in facioscapulohumeral dystrophy (FSHD). *PLoS. Genet* 7, e1000559 (2009).
- Azzalin, C.M., Reichenbach, P., Khoriavali, L., Giulotto, E., & Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318, 798-801 (2007).
- Nergadze, S.G. et al. CpG-island promoters drive transcription of human telomeres. *RNA.* 15, 2186-2194 (2009).
- Caslini, C., Connelly, J.A., Serna, A., Broccoli, D., & Hess, J.L. MLL associates with telomeres and regulates telomeric repeat-containing RNA transcription. *Mol. Cell Biol.* 29, 4519-4526 (2009).
- Deng, Z. et al. A role for CTCF and cohesin in subtelomere chromatin organization, TERRA transcription, and telomere end protection. *EMBO J.* 31, 4165-4178 (2012).
- Arnoult, N., Van, B.A., & Decottignies, A. Telomere length regulates TERRA levels through increased

- trimethylation of telomeric H3K9 and HP1alpha. *Nat. Struct. Mol. Biol.* 19, 948-956 (2012).
28. Maicher,A., Kastner,L., Dees,M., & Luke,B. Deregulated telomere transcription causes replication-dependent telomere shortening and promotes cellular senescence. *Nucleic Acids Res.* 40, 6649-6659 (2012).
 29. Pfeiffer,V. & Lingner,J. TERRA promotes telomere shortening through exonuclease 1-mediated resection of chromosome ends. *PLoS. Genet.* 8, e1002747 (2012).
 30. Balog,J. et al. Epigenetic regulation of the X-chromosomal macrosatellite repeat encoding for the cancer/testis gene CT47. *Eur. J. Hum. Genet.* 20, 185-191 (2012).
 31. Chantalat,S. et al. Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Res.* 21, 1426-1437 (2011).
 32. Dang,W. et al. Histone H4 lysine 16 acetylation regulates cellular lifespan. *Nature* 459, 802-807 (2009).
 33. Krishnan,V. et al. Histone H4 lysine 16 hypoacetylation is associated with defective DNA repair and premature senescence in Zmpste24-deficient mice. *Proc. Natl. Acad. Sci. U. S. A* 108, 12325-12330 (2011).
 34. O'Sullivan,R.J., Kubicek,S., Schreiber,S.L., & Karlseder,J. Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. *Nat. Struct. Mol. Biol.* 17, 1218-1225 (2010).
 35. Chandra,T. et al. Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Mol. Cell* 47, 203-214 (2012).
 36. Keogh,M.C. et al. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123, 593-605 (2005).
 37. Joshi,A.A. & Struhl,K. Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol. Cell* 20, 971-978 (2005).
 38. Carrozza,M.J. et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123, 581-592 (2005).
 39. Baird,D.M., Rowson,J., Wynford-Thomas,D., & Kipling,D. Extensive allelic variation and ultrashort telomeres in senescent human cells. *Nat. Genet.* 33, 203-207 (2003).
 40. Raz,V. et al. The nuclear lamina promotes telomere aggregation and centromere peripheral localization during senescence of human mesenchymal stem cells. *J. Cell Sci.* 121, 4018-4028 (2008).
 41. Luderus,M.E. et al. Structure, subnuclear distribution, and nuclear matrix association of the mammalian telomeric complex. *J. Cell Biol.* 135, 867-881 (1996).
 42. Cawthon,R.M. Telomere measurement by quantitative PCR. *Nucleic Acids Res.* 30, e47 (2002).
 43. Nelson,J.D., Denisenko,O., & Bomsztyk,K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.* 1, 179-185 (2006).
 44. Untergasser,A. et al. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35, W71-W74 (2007).
 45. Tobi,E.W. et al. Prenatal famine and genetic variation are independently and additively associated with DNA methylation at regulatory loci within IGF2/H19. *PLoS. ONE.* 7, e37933 (2012).

Acknowledgements

This study was supported by a grant from the Netherlands Consortium for Healthy Ageing [Grant 05060810] in the framework of the Netherlands Genomics Initiative/ Netherlands Organization for Scientific Research and the European Union's Seventh Framework Program IDEAL [FP7/2007-2011] under grant agreement n° 259679.

Disclosure of potential conflicts of interest

No potential conflict of interest was disclosed.

Supplemental material

Fig. S1. Activation of senescence associated β -galactosidase in late passage WI-38 fibroblasts.

Fig. S2. Relative quantification of histone modifications at the CT47 macrosatellite repeat and GAPDH promoter upon senescence.

Fig. S3. Methylation levels of individual CpGs at subtelomeric loci upon senescence.

Fig. S4. No increased levels of H3K9ac at subtelomeres upon senescence.

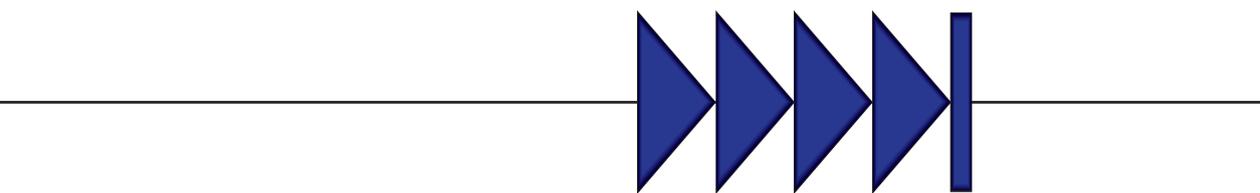
Fig. S5. Relative quantification of histone modifications at the subtelomeric D4Z4 macrosatellite repeat upon senescence.

Fig. S6. Methylation levels of individual CpGs at TERRA promoters upon senescence.

Table S1. Primer pairs used in this study.

All supplemental material belonging to this chapter can be accessed through <http://goo.gl/hNzh9w> or by using the QR-code below.





Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4.

4

Judit Balog*, Peter E. Thijssen*, Sean Shadle*, Kirsten R. Straasheijm, Patrick J. van der Vliet, Yvonne D. Krom, Marlinde L. van den Boogaard, Annika de Jong, Richard J.L.F. Lemmers, Rabi Tawil, Stephen J. Tapscott and Silvère M. van der Maarel

Adapted from Balog et al. (2015), *Epigenetics*, 10:12, 1133-1142

* Authors contributed equally to this work.

Abstract

Facioscapulohumeral muscular dystrophy is caused by incomplete epigenetic repression of the transcription factor DUX4 in skeletal muscle. A copy of DUX4 is located within each unit of the D4Z4 macrosatellite repeat array and its derepression in somatic cells is caused by either repeat array contraction (FSHD1) or by mutations in the chromatin repressor SMCHD1 (FSHD2). While DUX4 expression has thus far only been detected in FSHD muscle and muscle cell cultures, and increases with in vitro myogenic differentiation, the D4Z4 chromatin structure has only been studied in proliferating myoblasts or non-myogenic cells. We here show that SMCHD1 protein levels at D4Z4 decline during muscle cell differentiation and correlate with DUX4 derepression. In FSHD2, but not FSHD1, the loss of SMCHD1 repressor activity is partially compensated by increased Polycomb Repressive Complex 2 (PRC2)–mediated H3K27 trimethylation at D4Z4, a situation that can be mimicked by SMCHD1 knockdown in control myotubes. In contrast, moderate overexpression of SMCHD1 results in DUX4 silencing in FSHD1 and FSHD2 myotubes demonstrating that DUX4 derepression in FSHD is reversible. Together, we show that in FSHD1 and FSHD2 the decline in SMCHD1 protein levels during muscle cell differentiation renders skeletal muscle sensitive to DUX4.

Introduction

Facioscapulohumeral dystrophy (FSHD, OMIM 158900/158901) is characterized by progressive and often asymmetrical weakness of facial and upper extremity muscles typically starting in the second decade of life¹. FSHD is one of the most common myopathies with a recently reported prevalence of 12:100.000 in the Dutch population and is inherited in an autosomal dominant fashion in most families². FSHD is genetically linked to the polymorphic D4Z4 macrosatellite repeat array at 4q35, which in healthy subjects consists of 11-100 units, each of them being 3.3 kb in size and containing a copy of the DUX4 retrogene³⁻⁵.

The prevailing FSHD disease model postulates that DUX4 is normally silenced in somatic cells by repeat mediated epigenetic repression and that failure of this mechanism leads to partial decompaction of the D4Z4 chromatin structure and the expression of DUX4 in skeletal muscle of FSHD individuals⁶. Stable DUX4 expression requires a specific genetic background of chromosome 4 that contains a polymorphic DUX4 poly-adenylation signal (PAS) immediately distal to the D4Z4 repeat array⁷. Partial D4Z4 chromatin decompaction on a DUX4-PAS containing chromosome leads to the FSHD-specific pattern of DUX4 expression in myotube cultures with few myonuclei expressing relatively abundant amounts of DUX4 protein⁸. Expression of the germline double homeobox transcription factor DUX4 is harmful to skeletal muscle since it activates germline and early stem cell programs, and modulates inflammatory pathways eventually resulting in increased cell death⁹⁻¹¹.

In >95% of FSHD individuals (FSHD1) incomplete repression of DUX4 can be attributed to a contraction of the D4Z4 repeat array to a size of 1-10 units, whereas in the remainder of individuals with FSHD (FSHD2) the D4Z4 repeat arrays are in the normal size range^{5, 12, 13}. The majority of FSHD2 individuals carry heterozygous mutations in the structural maintenance of chromosomes hinge domain 1 (SMCHD1) gene on chromosome 18¹⁴. SMCHD1 is a chromatin repressor that binds to D4Z4 and germline mutations in SMCHD1 result in reduced D4Z4 chromatin compaction and DUX4 expression¹⁴. Moreover, SMCHD1 was shown to act as a modifier of disease severity in FSHD1 and the functional consequences of SMCHD1 mutations, in combination with D4Z4 repeat size, determines the disease penetrance^{15, 16}.

Several lines of evidence indicate that epigenetic derepression of the D4Z4 macrosatellite repeat array in FSHD results in DUX4 expression in skeletal muscle cells. In somatic cells originating from healthy individuals, such as myoblasts, the D4Z4 repeat array is characterized by high, but inhomogeneous levels of CpG methylation and histone 3 lysine 9 trimethylation (H3K9me3)¹⁷⁻²¹. In proliferating FSHD somatic cell cultures, a reduction of both CpG methylation and H3K9me3 at D4Z4 has been reported together with reduced cohesin and heterochromatin protein 1 gamma (HP1 γ) levels, presumably facilitating sporadic activation of DUX4²¹. In addition, members of the Polycomb Repressive Complex 2 (PRC2) and the histone modification deposited by the PRC2 complex, H3K27me3, were identified at D4Z4 in control and FSHD1 proliferating myoblasts²¹⁻²³. The mechanism of PRC2 recruitment remains however to be determined.

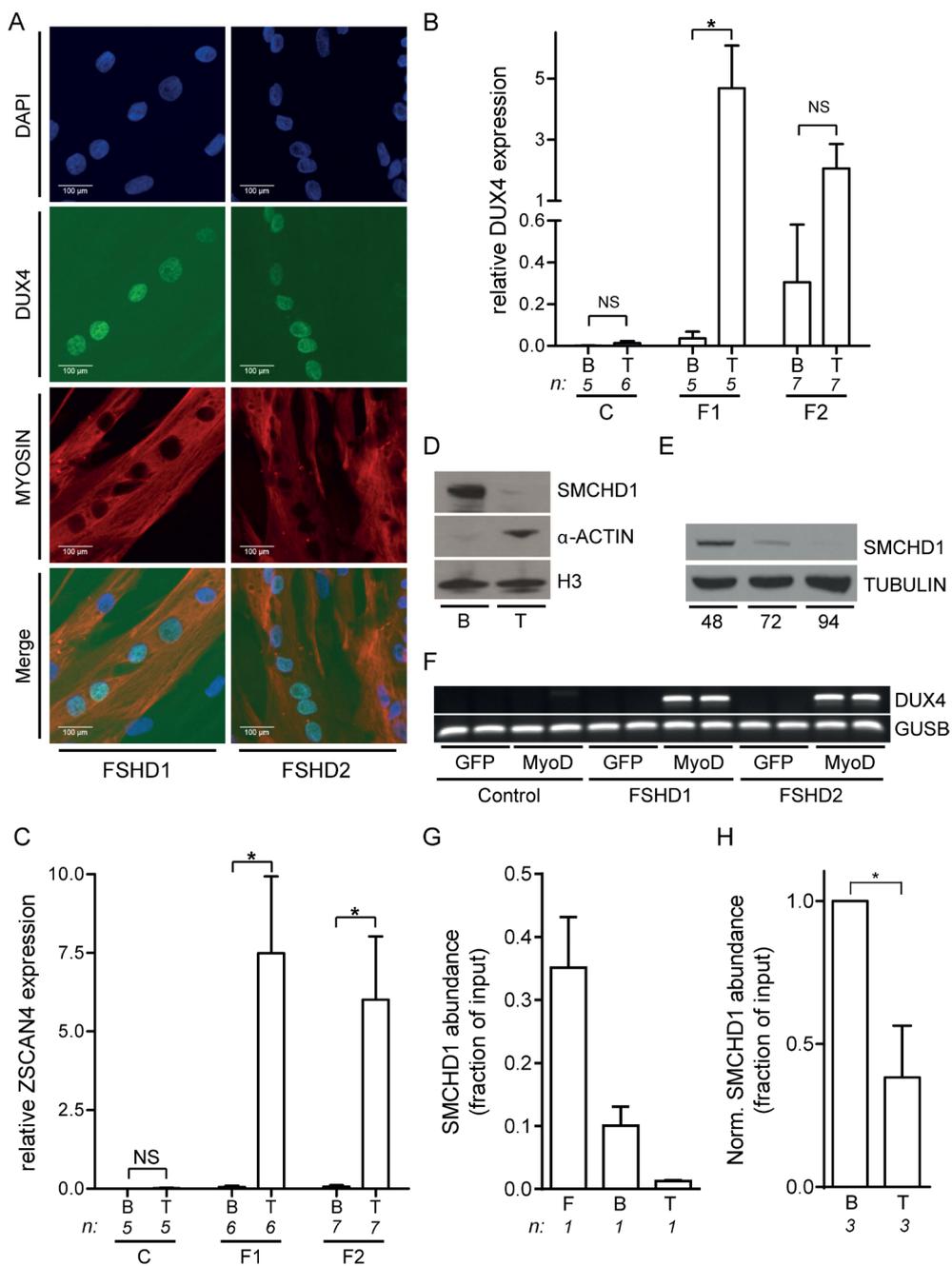
D4Z4 non-coding RNAs may have a role, since the endogenous RNAi pathway was recently demonstrated to facilitate DICER/AGO-dependent epigenetic silencing of the D4Z4 repeat arrays in muscle cells²³⁻²⁵.

Although DUX4 expression has repeatedly been shown to be increased in FSHD muscle cells upon in vitro differentiation^{8, 26-28}, data on the epigenetic regulation of the D4Z4 repeat array during this process is scarce. Nevertheless, extensive changes in the genome wide chromatin structure have been reported upon myogenic differentiation in mouse muscle cells. In particular, the Polycomb complexes were shown to be pivotal for the regulation of lineage commitment and differentiation specific genes²⁹. Moreover, multi-nucleated myotubes more closely resemble the in vivo structural composition of muscle than proliferating myoblasts. Taken together, differentiated myotubes are likely a more relevant in vitro model to study the epigenetic requirements for DUX4 expression than proliferating myoblasts.

This prompted us to examine whether we could identify changes in D4Z4 chromatin regulation specific to differentiated muscle cells that might account for the susceptibility of this tissue to express DUX4 at higher levels than in myoblasts or many other tissues. Our study highlights the importance of SMCHD1 in regulating D4Z4 chromatin structure and DUX4 expression in muscle cells of both FSHD1 and FSHD2. Whereas a decline in SMCHD1 protein levels, and specifically its reduction at D4Z4, during skeletal muscle differentiation leads to DUX4 expression, ectopic expression of SMCHD1 in either FSHD1 or FSHD2 muscle cell cultures efficiently suppresses DUX4 expression. Moreover, in FSHD2 but not in FSHD1, the loss of SMCHD1 is accompanied by a gain in PRC2-mediated H3K27 trimethylation, likely as a compensatory but incomplete mechanism to silence DUX4 in skeletal muscle.

Figure 1: DUX4 activation during myogenic differentiation coincides with SMCHD1 reduction

A) Immunofluorescence microscopy analysis confirmed the typical DUX4 protein expression pattern in myosin positive, multinucleated myotubes derived from FSHD1 and FSHD2 individuals. Images were taken using a 200x magnification, scale bars are displayed. **B)** Quantitative mRNA analysis in control (C), FSHD1 (F1) and FSHD2 (F2) primary myoblasts (B) and myotubes (T) showing increased levels of DUX4 expression upon differentiation. GAPDH and GUSB were used as reference genes, n indicates number of samples, error bars display SD and significance was calculated using a two tailed student's t-test. **C)** Quantitative mRNA analysis showed robust ZSCAN4 activation upon myogenic differentiation in FSHD myotubes. GAPDH and GUSB were used as reference genes, n indicates number of samples, error bars display SD and significance was calculated using a two tailed student's t-test. **D)** Western blot analysis showing reduced levels of SMCHD1 upon muscle cell differentiation in a primary muscle cell culture derived from a control individual. Myoblasts (B) were differentiated into myotubes (T) for 48 h. H3 serves as a control for equal protein loading, α -ACTIN serves as a control for the induction of myogenic differentiation. **E)** Western blot analysis of a control derived primary fibroblast undergoing forced myogenesis by ectopic MyoD expression for 48, 72 or 94 h showed a decrease in SMCHD1 protein expression. TUBULIN serves as a loading control. **F)** Duplicate RT-PCR analysis of DUX4 expression upon ectopic MyoD expression in a Control 1, FSHD1 1 or FSHD2 1 fibroblast (table S2) revealed DUX4 activation in patient cells exclusively. Ectopic GFP expression was used as a control and GUSB serves as an internal PCR control. **G, H)** Normalized ChIP qPCR analysis of SMCHD1 binding at D4Z4 in isogenic fibroblasts (F), myoblasts (B) and myotubes (T) derived from the same control individual (panel G) and three independent control derived myoblast – myotube pairs showed the highest SMCHD1 abundance in myoblasts with a further decrease during myogenesis. Error bars display SD in panel G and H and significance was calculated using a two tailed student's t-test. NS = not significant; * = $P < 0.05$; ** = $P < 0.005$.



Results

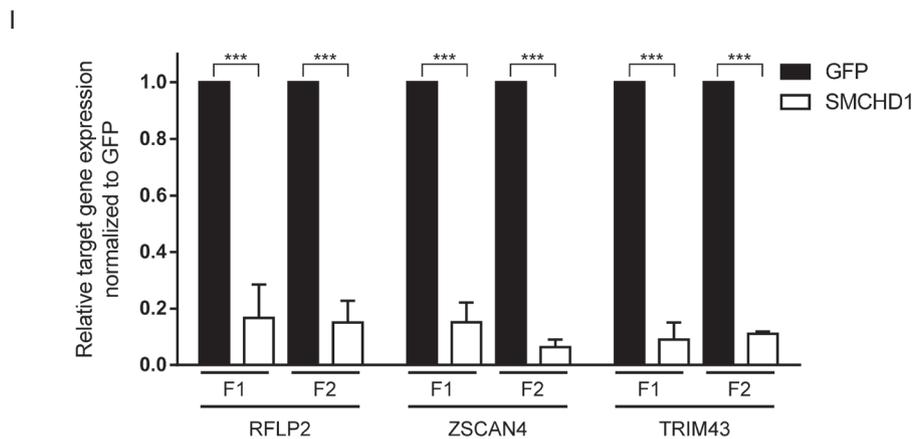
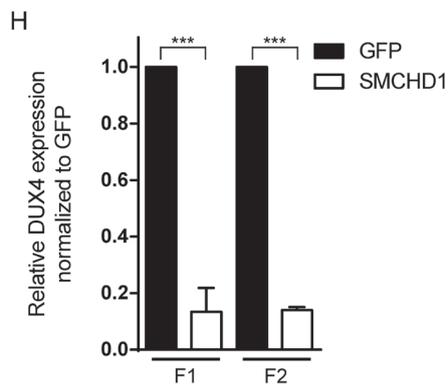
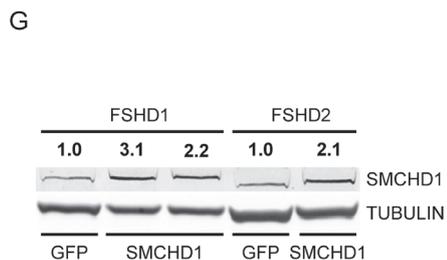
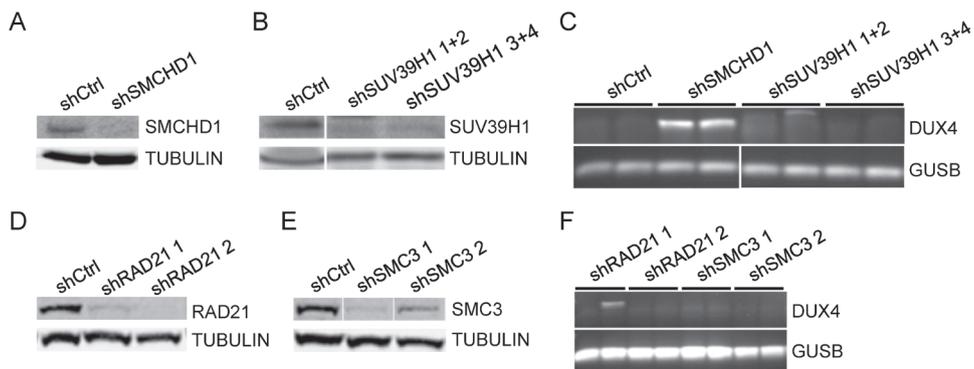
DUX4 expression is induced upon myogenic differentiation and coincides with reduced SMCHD1 levels.

Although DUX4 expression has been reported repeatedly in FSHD myotubes^{8, 26-28}, comprehensive data on the regulation of DUX4 expression in FSHD1 and in particular on FSHD2 muscle cells is scarce. Therefore, our first aim was to firmly establish that DUX4 expression increases with differentiation of FSHD1 and FSHD2 muscle cells. We cultured a large set of primary myoblasts derived from control, FSHD1 and FSHD2 individuals and harvested RNA from both proliferating myoblasts and differentiating myotubes (further referred to as myoblasts and myotubes). The differentiation status of the cultures was confirmed by myotube formation, as visualized by immunofluorescent labeling of myosin, and MYOG transcript levels (**Fig. 1A, Fig S1-S2**). Multinucleated myofibers were observed in all conditions and differentiation was confirmed by the increase in MYOG levels in control, FSHD1 and FSHD2 muscle cell cultures.

mRNA quantification by qRT-PCR confirmed the activation of DUX4 in FSHD1 and FSHD2 myotubes, whereas we could not reproducibly detect DUX4 transcripts in control samples (**Fig. 1B**). While the increase of DUX4 expression during myotube formation in FSHD1 cells was statistically significant, FSHD2 cells were more variable. In FSHD2 cells an increase of DUX4 during myogenesis can be observed, however FSHD2 muscle cell cultures already show variable levels of DUX4 transcripts before differentiation, which may explain the non-significant effect (**Fig. 1B**). Immunofluorescence microscopy on a subset of myotube cultures established the variegated pattern of DUX4 protein expression in both forms of the disease (**Fig. 1A, Fig. S2**). We also quantified expression levels of the previously established DUX4 target gene ZSCAN4 and observed FSHD-specific induction during myoblast differentiation (**Fig. 1C**)⁹.

Figure 2: SMCHD1, but not SUV39H1 and Cohesin, regulates DUX4 expression

Western blot confirmation of **A**) SMCHD1, **B**) SUV39H1, **D**) RAD21, and **E**) SMC3 knockdown in control myotubes expressing the indicated lentiviral transduced shRNAs. Tubulin was used as a loading control. Representative blots of at least duplicate experiments are shown. **C, F**) Standard gel electrophoresis analysis of DUX4 expression upon lentiviral knockdown of SMCHD1, SUV39H1, RAD21 and SMC3 in control myotubes. Only depletion of SMCHD1 resulted in reproducible activation of DUX4 transcription. Representative gel photos are shown of at least duplicate experiments. The PCR fragment visible in one RAD21 knockdown was sequenced and shown to be the product of an a-specific amplification. **G**) Western blot analysis confirms a two / threefold increase of SMCHD1 expression upon its lentiviral transduction in 2 FSHD1 and 1 FSHD2 myotube cultures. GFP transduced myotubes served as a negative control. Numbers indicate normalized relative expression levels of SMCHD1 using Tubulin as a loading control followed by setting normalized SMCHD1 levels of GFP transduced samples (only expressing endogenous SMCHD1) to 1. **H**) Expression levels of DUX4 were significantly reduced upon ectopic expression of SMCHD1 in 2 FSHD1 and 2 FSHD2 myotube cultures. Relative DUX4 expression was calculated for each sample by normalization to GUSB and GAPDH housekeeping genes. Bars show values of each samples adjusted to the expression value of GFP transduced sample as 1. **I**) Expression levels of the DUX4 target genes RFPL2, ZSCAN4, and TRIM43 showed a significant reduction upon ectopic expression of SMCHD1 in 2 FSHD1 and 2 FSHD2 myotubes, concordant with decreased DUX4 protein expression. Expression levels were normalized as described for panel Figure 1A. For panel H, I: Error bars display SD and significance was calculated using a two-tailed Student's t-test. All $P < 0.0005$ indicated by ***.



Since we previously demonstrated that SMCHD1 is a D4Z4 chromatin repressor¹⁴, we examined SMCHD1 protein levels. Western blot analysis showed a reduction in SMCHD1 protein levels after control myoblast differentiation (**Fig. 1D**), a phenomenon that could be reproduced by forced myogenic differentiation of control fibroblasts transduced with the myogenic transcription factor MyoD (**Fig. 1E**). Upon forced myogenesis by MyoD transduction in 2 FSHD1 and 2 FSHD2 fibroblast cell lines we observed induction of DUX4, whereas DUX4 remained undetectable in 2 control samples treated with MyoD (**Fig. 1F, Fig. S3**). Employing ChIP-qPCR in control fibroblast, myoblast and myotube cultures from the same donor we observed higher density of SMCHD1 at qD4Z4, located around the transcriptional start site of DUX4 (**Fig. S4**)²¹, in fibroblasts as compared to myoblasts, with a further decrease in myotubes (**Fig. 1G**). Analysis of three additional independent pairs of control myoblast and myotube cultures confirmed the significant decrease in SMCHD1 levels at D4Z4 after myoblast differentiation (**Fig. 1H**). Together, we confirm the FSHD-specific induction of DUX4 expression upon myogenic differentiation in a large set of FSHD1 and FSHD2 primary myotube cultures and upon MyoD-induced differentiation in FSHD fibroblasts. Moreover, we demonstrate that DUX4 expression negatively correlates with SMCHD1 levels at D4Z4 in myogenic cell cultures, specifically in differentiated myotubes.

SMCHD1 levels control DUX4 expression in somatic cells

To establish a causal relationship between SMCHD1 protein levels and DUX4 expression in myotubes and SMCHD1 protein levels, we conducted lentiviral shRNA-mediated depletion experiments in differentiating primary control myoblast cultures (Controls 2 and 8; **table S1**) carrying an FSHD-permissive DUX4-PAS containing 4A161 allele with D4Z4 repeat sizes varying between 20 and 32 units. Depletion of SMCHD1 resulted in robust activation of DUX4 expression (**Fig. 2A, C; Fig. S5A**), which is in agreement with our earlier observations¹⁴. This phenomenon of DUX4 activation seems to be rather specific for SMCHD1 since knocking down other known chromatin repressors of D4Z4 in parallel experiments, using the same control myoblast cultures, did not lead to DUX4 expression. Upon knockdown of SUV39H1, a histone methyltransferase previously shown to be involved in establishing the H3K9me3 modification at D4Z4²¹, we did not observe a reduction in relative abundance of H3K9me3 at D4Z4 (**Fig. 2B, Fig. S5, Fig. S6B**), nor the transcriptional activation of DUX4, in 2 independent control myotube cultures (**Fig. 2C**). Independent depletion of cohesin proteins SMC3 or RAD21 by a similar strategy did not cause a consistent activation of DUX4 either under these conditions (**Fig. 2D-F**). In support, we did not see significant changes in H3K9me3 levels or RAD21 levels between control and FSHD muscle cell cultures, in myoblasts or myotubes (**Fig. S7; Fig. S8**).

Since SMCHD1 levels decline at D4Z4 during myoblast differentiation, we next tested whether ectopic expression of SMCHD1 in FSHD myotubes can rescue the repression of DUX4. To this end, we transduced FSHD1 and FSHD2 myoblast cultures with SMCHD1 expressing lentiviruses and allowed the cells to form myotubes. In both FSHD1 and FSHD2 myotubes 2-3 fold overexpression of SMCHD1, as determined by western blot analyses, resulted in a 70-90% reduction in DUX4 mRNA levels (**Fig. 2G-H**). To confirm the silencing of DUX4, we analyzed the mRNA expression levels of three known downstream target

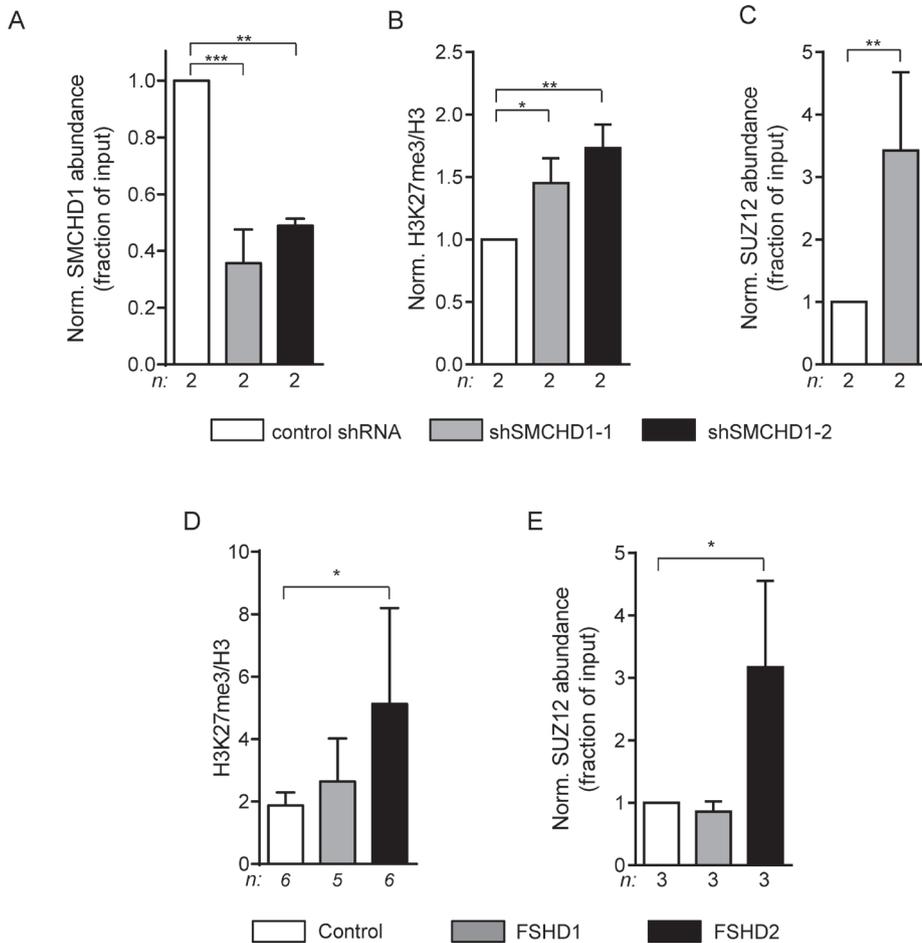


Figure 3: SMCHD1 depletion at D4Z4 leads to increased H3K27me3 and PRC2 binding.

A) ChIP-qPCR analysis showed a statistically significant reduction of SMCHD1 at qD4Z4 upon its lentiviral knockdown in control myotubes with two independent shRNA constructs in 2 independent control myotube cultures. B) ChIP-qPCR analysis of H3K27me3 at qD4Z4 upon SMCHD1 depletion showed a significant increase at qD4Z4. Enrichment values were normalized to H3 enrichment values. C) Normalized ChIP-qPCR analysis of SUZ12 upon SMCHD1 depletion showed a statistically significant increase at qD4Z4 in 2 independent experiments performed on 2 control myoblast cultures. n indicates sample size, error bars indicate SD and significance was tested with students t test. D) ChIP-qPCR analysis of H3K27me3 at qD4Z4 showed a significant increase in FSHD2 myotubes compared to controls. Enrichment values were normalized to H3 enrichment values. E) ChIP-qPCR analysis of SUZ12 abundance showed a significant increase in FSHD2 myotubes at qD4Z4. On panel A, B, D and E n indicates sample size, error bars display SD and significance was tested using a one way-ANOVA followed by Bonferroni multiple comparison test. * = $P < 0.05$ ** = $P < 0.005$ *** = $P < 0.0005$.

genes of DUX4 protein. ZSCAN4, RFPL2, and TRIM43 expression levels decreased 70-90% upon overexpression of SMCHD1 (**Fig. 2I**). To rule out that the lower DUX4 transcript levels were a consequence of reduced muscle cell differentiation, we measured MYOG transcript levels. MYOG increased similarly between the GFP expressing control and SMCHD1 expressing FSHD myotubes, with the exception of one FSHD2 myotube culture that showed substantially higher MYOG levels in the SMCHD1 expressing compared to control GFP expressing conditions (**Fig. S9**). This demonstrates that modest upregulation of SMCHD1 has no inhibiting effect on myogenic differentiation. In conclusion, DUX4 expression levels are strongly decreased upon moderate overexpression of SMCHD1, emphasizing its role as D4Z4 repressor in both genetic forms of the disease.

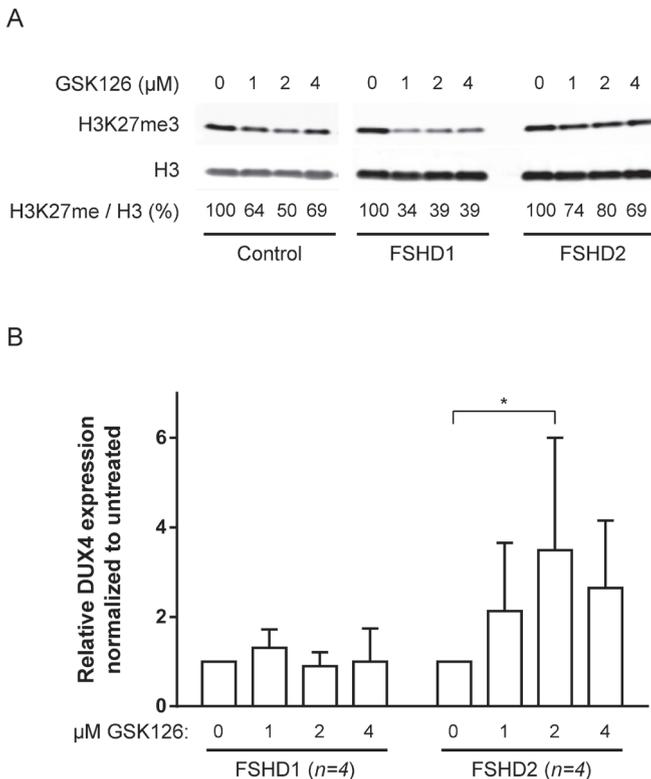


Figure 4: Treatment of FSHD2 myotube cultures with EZH2 inhibitor GSK126 increases DUX4 levels

A) Western blot analysis of control, FSHD1 and FSHD2 myotube samples after treatment with 0, 1, 2 and 4 μM GSK126. Blots were probed with H3K27me3 antibody and H3 antibody as loading control. Shown are the reduced ratios of H3K27me3:H3 signal intensities in samples treated with GSK126 compared to the untreated sample as a result of EZH2 inhibition. B) qRT-PCR analysis of DUX4 expression in FSHD1 and FSHD2 myotubes after GSK126 treatment shows that relative DUX4 transcript levels are significantly increased in FSHD2 samples treated with 2 μM GSK126 but not in FSHD1 myotubes. Graph shows the results of 4 FSHD1 and 4 FSHD2 cell lines. Results of control cell lines are not shown, DUX4 transcript was not detectable. Error bars show SD, n indicates number of independent cell lines and significance was tested by two-way ANOVA followed by Bonferroni's multiple comparison test. $*=P < 0.05$.

SMCHD1 reduction leads to PRC2 enrichment at D4Z4 in FSHD2 but not FSHD1

Since SMCHD1 was recently shown to associate with the repressive H3K27me3 modification at the inactive X chromosome³⁰, we further determined the levels of H3K27me3 and the PRC2 complex at D4Z4 upon depletion of SMCHD1 in two independent control myotube cultures. SMCHD1 knockdown resulted in a 2-fold decrease in SMCHD1 levels at D4Z4 using two different SMCHD1 shRNA vectors (**Fig. 3A**) and increased H3K27me3 levels at D4Z4 (**Fig. 3B**). Because of the increase in H3K27me3 at D4Z4 upon SMCHD1 depletion, we further analyzed the involvement of the PRC2 complex and observed a significant increase of the PRC2 protein SUZ12 at D4Z4 (**Fig. 3C**).

To validate these data derived from knock down experiments in control myotube cultures, we investigated the levels of the PRC2 complex in primary control and FSHD myotubes. Previously, H3K27me3 levels at D4Z4 were reported to be similar in control and FSHD1 proliferating myoblast cultures at the DBE locus just proximal to qD4Z4 (**Fig. S4**)²³. We measured H3K27me3 levels at qD4Z4 and observed significantly higher levels of H3K27me3 in FSHD2 myotubes (**Fig. 3D**), consistent with the increase we observed following knockdown of SMCHD1 (**Fig. 3B**). Increased H3K27me3 levels were also observed in FSHD2 myoblasts (**Fig. S10**). Next, we quantified the levels of the PRC2 protein SUZ12 at D4Z4 and detected higher levels of SUZ12 at D4Z4 in FSHD2 samples (**Fig. 3E**), all consistent with an increase in H3K27me3 through PRC2 recruitment at D4Z4.

The decreased SMCHD1 activity in FSHD2 myotubes correlates with increased PRC2 levels at D4Z4. To test the repressive role of PRC2 at D4Z4, we inhibited PRC2 activity by treating control, FSHD1 and FSHD2 muscle cell cultures with increasing concentrations of the EZH2 inhibitor GSK126³¹. After differentiation, total H3K27me3 levels were decreased in all cell cultures as shown by western blot analysis (**Fig. 4A**). DUX4 transcript levels increased significantly in FSHD2 samples treated at 2 μ M GSK126, (**Fig. 4B**) suggesting that PRC2 is involved in the repression of D4Z4 in FSHD2, but not in FSHD1.

Discussion

Despite the recent advances in our understanding of the common molecular disease mechanism in FSHD1 and FSHD2 and the consensus model of incomplete epigenetic repression of DUX4 in somatic cells, the key epigenetic changes at D4Z4 in FSHD1 and FSHD2, and the selective tissue involvement is only partly understood. FSHD is mainly a disease of skeletal muscle, suggesting that differentiated skeletal muscle might have an epigenetic state more permissive for DUX4 expression. In vitro, it was observed that DUX4 expression in FSHD muscle cells specifically increases upon myogenic differentiation. To date, however, most D4Z4 chromatin studies have focused on myoblast and non-myogenic cell cultures. The dynamics of the different epigenetic mechanisms involved D4Z4 chromatin organization during muscle cell differentiation, and their causality to DUX4 expression, have not been studied in detail. Here we show that SMCHD1 protein levels decrease during muscle cell differentiation and correlate with DUX4 expression, possibly making differentiated muscle cells particularly prone

to incomplete D4Z4 repression. Moreover we have identified the involvement of PRC2-mediated H3K27 trimethylation of D4Z4 selectively in FSHD2.

Thus far most studies have been performed in FSHD1 muscle cell cultures, but we here firmly establish that in FSHD1 and FSHD2 DUX4 expression is strongly upregulated during muscle cell differentiation, emphasizing the importance of studying D4Z4 chromatin structure in differentiated muscle cells. Our observation of the decline in SMCHD1 levels during muscle cell differentiation may provide an explanation towards the susceptibility of muscle in expressing DUX4 and prompted us to further explore the role of SMCHD1 in somatic DUX4 repression. Depletion of SMCHD1 in control myotubes resulted in robust DUX4 expression. Conversely, 1.5-3 fold increases in SMCHD1 protein levels by ectopic expression in FSHD1 and FSHD2 myotubes led to a significant decrease in DUX4 levels and that of its target genes. This demonstrates that the derepression of DUX4 in FSHD muscle cells is a reversible process that can be rescued by increasing SMCHD1 levels.

SMCHD1 was recently shown to associate with H3K27me3 at the inactive X chromosome and we observed that loss of SMCHD1 at D4Z4 leads to an increase in PRC2 and H3K27me3 levels at D4Z4 in FSHD2. This strongly suggests that there is aberrant PRC2 regulation in FSHD2, but not FSHD1, myoblasts and myotubes. Since SMCHD1 has a role in the establishment and/or maintenance of CpG methylation, this observation consistent with recent studies showing an inverse relationship between CpG methylation and PRC2-mediated H3K27 methylation³². While the overall levels of some PRC2 complex proteins decline during murine myogenic differentiation³³, our data suggest a local enrichment of PRC2 at D4Z4 in FSHD2 cells during this process. Previous studies did not observe a difference in PRC2 and H3K27me3 at D4Z4 in FSHD. However, these studies were mostly restricted to FSHD1 myoblast cultures^{22, 23}, used primers that were not specific to chromosome 4²², or analyzed a single cell line³⁴ which may explain this difference. Indeed a recent study by Zeng et al. showed that although D4Z4 units are dispersed throughout the genome, epigenetically the D4Z4 repeat arrays on chromosomes 4 and 10 are distinctly regulated from their homologs³⁴.

Our data also identifies a difference between FSHD1 and FSHD2 in the epigenetic regulation of D4Z4. Depletion of SMCHD1 in control cells, mimicking FSHD2, leads to increased abundance of PRC2. The involvement of PRC2 in FSHD2 is further substantiated by the observation that GSK126 specifically increases DUX4 transcript levels in FSHD2, but not in FSHD1. Despite this epigenetic difference, SMCHD1 depletion and overexpression respectively leads to activation and repression of DUX4 in both FSHD1 and FSHD2 derived cells. It is unlikely that a selective involvement of a single contracted D4Z4 repeat array in FSHD1 creates a technical limitation in our ChIP approach (interrogating all four D4Z4 repeat arrays simultaneously) and explains the FSHD2-specificity of the involvement of PRC2 at D4Z4. Therefore, we conclude that although in both conditions DUX4 is derepressed, there are some unique epigenetic responses to repeat contraction or SMCHD1 malfunctioning in FSHD1 and FSHD2, respectively.

Increased PRC2 recruitment and H3K27me3 levels are generally associated with transcriptional repression³⁵. However, in FSHD2 they are associated with the transcriptional derepression of the DUX4 locus. PRC2 recruitment to the derepressed D4Z4 region may reflect a rather effective failsafe mechanism to compensate for the loss of other repressive mechanisms. Indeed a compensatory repressive mechanism by PRC2 is supported by our observation of increased DUX4 transcript levels in FSHD2 myotubes in the presence of the EZH2 inhibitor GSK126. The specific nature of the trigger leading to the sporadic escape of repression typical for DUX4 remains unclear at this point and requires experimental approaches, which can dissect differences between single nuclei.

We could not confirm a role for SUV39H1 or the cohesin complex in DUX4 repression in a myogenic context: we did not observe different H3K9me3 or SMC3 levels in myotubes and knockdown of SUV39H1 and cohesin in myotubes also did not activate DUX4. Knockdown of SUV39H1 did not decrease H3K9me3 at D4Z4 in myotubes either. Previously, siRNA mediated knockdown of SUV39H1 led to reduced levels of H3K9me3 at D4Z4 in HeLa cells²¹. Similarly, reducing SUV39H1 activity in proliferating immortalized myoblasts by knock down or treatment of primary and immortalized myoblasts with Chaetocin decreased H3K9me3 levels and resulted in transcriptional derepression of DUX4³⁴. The specificity of Chaetocin was recently challenged, however, as it might also inhibit the activity of other histone lysine methyltransferases³⁶. Our data thus suggest that the regulation of this modification at D4Z4 is different between primary and immortalized myoblasts, and between proliferating and differentiating myoblasts, perhaps by involvement of additional methyltransferases such as G9a³⁴. A recent report suggested a role for the telomere in the epigenetic regulation of D4Z4³⁷. Since myoblasts are immortalized by telomerase expression, it will be important to comprehensively compare the epigenetic regulation of the D4Z4 repeat array between primary and immortalized myoblasts.

Together with earlier observations that SMCHD1 is a modifier of disease severity in FSHD1, our study suggests that D4Z4 derepression in FSHD1 and FSHD2 converge at SMCHD1. Contraction of D4Z4 in FSHD1 may lead to specific loss of SMCHD1 at the contracted allele, analogous to its interaction with telomeric repeats, which shows a non-linear positive correlation to telomere length³⁸. Altogether this emphasizes the crucial role of SMCHD1 in suppressing DUX4 activity in somatic cells and its decline during muscle cell differentiation provides a plausible explanation for the increased expression of DUX4 in skeletal muscle cells.

Materials and Methods

Culturing of human primary myoblast cell lines

Human primary myoblast cell lines were originating from the University of Rochester bio repository (<http://www.urmc.rochester.edu/fields-center/>). Muscle samples were obtained after subjects were consented under a protocol approved by the institutional review board at the University of Rochester. Myoblasts were cultured in DMEM/F-10 media (#31550 Gibco/Life Technologies, Bleiswijk, The Netherlands) supplemented with 20% heat inactivated fetal bovine serum (FBS #10270 Gibco/Life Technologies, Bleiswijk, The Netherlands), 1% penicillin/streptomycin (#15140 Gibco/Life Technologies, Bleiswijk, The Netherlands) and 10ng/ml rhFGF (#G5071 Promega, Leiden, The Netherlands) and 1 μ M dexamethasone (#D2915 Sigma-Aldrich, Zwijndrecht, The Netherlands) was added to the medium. Myoblasts were fused at 80% confluency by culturing them in DMEM/F-12 Glutamax media (#31331, Gibco/Life Technologies, Bleiswijk, The Netherlands) containing 1% penicillin and streptomycin and 2% KnockOut serum replacement formulation (#10828 Gibco/Life Technologies, Bleiswijk, The Netherlands) for 36 hours. Human control fibroblast cell lines were maintained in DMEM/F-12, supplemented with 20% FBS, 1% penicillin and streptomycin, 10 mM HEPES (#15630#, Gibco/Life Technologies, Bleiswijk, The Netherlands) and 1mM sodium pyruvate (###11360, Gibco/Life Technologies, Bleiswijk, The Netherlands). Used cell lines, D4Z4 allele information and experimental use are listed in supplemental table S1.

RNA isolation, cDNA synthesis and qRT-PCR

Myoblast and myotube samples were harvested for RNA isolation by adding QIAzol lysis reagent (#79306 Qiagen N.V., Venlo, The Netherlands). RNA was isolated by miRNeasy Mini Kit (#217004 Qiagen) including DNase treatment according to the manufacturer's instructions. cDNA was synthesized with RevertAid H Minus First strand cDNA Synthesis Kit (#K1632 Thermo Fischer Scientific Inc., Waltham, MA) using 2 μ g template RNA and poly-dT primers. Gene specific cDNA products were quantified by qPCR in duplicate using SYBR green master mix supplemented with gene specific primers (table S2) using the CFX96 system (Bio-Rad, Veenendaal, The Netherlands). qRT-PCR data were analyzed by Bio-Rad CFX manager version 3.0 (Bio-Rad) using GAPDH and GUSB as reference genes for every individual sample.

Transduction of primary cell cultures

Full length SMCHD1 was PCR amplified from cDNA and cloned into the pRRL-CMV lentiviral backbone containing a puromycin selection marker using standard cloning procedures and the obtained construct was verified by Sanger sequencing. shRNA constructs originating from the Mission shRNA library (MISSION shRNA library, TRC1 or TRC2; Sigma Aldrich) and are listed in supplemental table S2. Constructs were used to generate lentiviral particles, and myoblast cell lines were transduced at 50% confluency and 24 hours after transduction they were grown in media containing 0.5 μ g/ml puromycin for selection. Differentiation to myotubes was induced by serum reduction at 80% confluency and cells were harvested after 36 hours of differentiation. Transduction of fibroblasts to ectopically express MyoD1 and induce myogenesis was

performed as described previously³⁹.

Chromatin Immunoprecipitation

Histone ChIP studies were carried out as described before¹⁷ using antibodies against H3 (ab1791, 2 µl/rxn, Abcam, Cambridge, UK), H3K9me3 (39161, 5 µl/rxn, Active Motif, Carlsbad, USA), H3K27me3 (#17-622, 5 µl/rxn, Merck-Millipore, Amsterdam Zuid-Oost, The Netherlands), and total IgG (5 µl/rxn, Merck-Millipore, Amsterdam Zuid-Oost, The Netherlands). Non-histone ChIPs were carried out as described before¹⁴ using antibodies against SMCHD1 (ab31865, 5 µg/rxn, Abcam, Cambridge, UK) RAD21 (ab992, 5 µg/rxn, Abcam, Cambridge, UK), SMC3 (ab9263, 5 µg/rxn, Abcam, Cambridge, UK), SUZ12 (D39F6, 5 µg/rxn, Cell Signaling, Leiden, The Netherlands). All antibodies were qualified for ChIP application by the manufacturer. All ChIP-qPCR experiments were performed 2 independent times.

GSK126 treatment of primary myoblast cultures

3 control 4 FSHD1 and 4 FSHD2 human primary myoblast cultures were grown to 50-60% confluency and culturing was continued in the presence of 0 µM, 1 µM, 4 µM and 8 µM GSK126 (#1346574-57-9 MedKoo Biosciences, North Carolina, USA) in proliferation media. GSK126 was dissolved in DMSO (#D2650 Sigma-Aldrich, St Louis, USA) and the different concentration of GSK126 were added in equal volumes of DMSO to the cells. Cultures shown as 0 µM GSK126 were treated only with DMSO. After 24 hours culturing was continued in fusion media in the presence of the same concentration of GSK126 for another 24 hours. RNA and protein samples were harvested as described above.

Western blot

Cell were directly lysed in NuPAGE LDS Sample Buffer (#NP0008, ThermoFisher Scientific, Waltham, USA) and loaded on Novex 4-12% Bis-Tris Protein Gels (#WT4121A, ThermoFisher Scientific, Waltham, USA) according to manufacturer's instructions. Running of the gel and transfer to Hybond nitrocellulose membranes (#10600048, GE Healthcare, Diegem, Belgium) were carried out using the Nupage Novex SDS-page gel system (ThermoFisher Scientific, Waltham, USA). Blots were blocked in 4% milk in PBS and incubated with antibodies against SMCHD1 (1:250), SUV39H1 (0.1 µg/ml, #07-958, Merck-Millipore, Amsterdam Zuid-Oost, The Netherlands), RAD21 (1:500), SMC3 (1:1000) and Tubulin (1:2000, T6199, Sigma-Aldrich, St Louis, USA). Detection and relative quantification were done using the Odyssey system (V3.0, LI-COR Biosciences, Lincoln NE, USA), except for blots in Figure 1, which were visualized by enhanced chemiluminescence (ECL).

Statistical analysis

All statistical analysis was done using GraphPad Prism 6 software and the exact method applied at different experiments is described in the figure legends.

References

1. Tawil,R., van der Maarel,S.M., & Tapscott,S.J. Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. *Skelet. Muscle* 4, 12 (2014).
2. Deenen,J.C. et al. Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology*(2014).
3. Gabriels,J. et al. 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 (1999).
4. Wijmenga,C. et al. Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet* 336, 651-653 (1990).
5. Wijmenga,C. et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 2, 26-30 (1992).
6. van der Maarel,S.M., Tawil,R., & Tapscott,S.J. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. *Trends Mol. Med.* 17, 252-258 (2011).
7. Lemmers,R.J. et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650-1653 (2010).
8. Snider,L. et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS. Genet.* 6, e1001181 (2010).
9. Geng,L.N. et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38-51 (2012).
10. Kowaljow,V. et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord* 17, 611-623 (2007).
11. Wallace,L.M. et al. DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann. Neurol.* 69, 540-552 (2011).
12. van Deutekom,J.C. et al. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* 2, 2037-2042 (1993).
13. Gilbert,J.R. et al. Evidence for heterogeneity in facioscapulohumeral muscular dystrophy (FSHD). *Am J Hum Genet* 53, 401-408 (1993).
14. Lemmers,R.J. et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* 44, 1370-1374 (2012).
15. Lemmers,R.J. et al. Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. *Hum. Mol. Genet.* 24, 659-669 (2015).
16. Sacconi,S. et al. The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. *Am. J. Hum. Genet.* 93, 744-751 (2013).
17. Balog,J. et al. Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. *Epigenetics.* 7, 579-584 (2012).
18. de Greef,J.C. et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mutat.* 30, 1449-1459 (2009).
19. Hartweck,L.M. et al. A focal domain of extreme demethylation within D4Z4 in FSHD2. *Neurology* 80, 392-399 (2013).
20. van Overveld,P.G. et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet* 35, 315-317 (2003).
21. Zeng,W. et al. Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats in facioscapulohumeral dystrophy (FSHD). *PLoS. Genet* 7, e1000559 (2009).
22. Bodega,B. et al. 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. Biol.* 7, 41 (2009).
23. Cabianna,D.S. et al. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* 149, 819-831 (2012).
24. Lim,J.W. et al. DICER/AGO-dependent epigenetic silencing of D4Z4 repeats enhanced by exogenous siRNA suggests mechanisms and therapies for FSHD. *Hum. Mol. Genet.*(2015).
25. Snider,L. et al. RNA Transcripts, miRNA-sized Fragments, and Proteins Produced from D4Z4 Units: New Candidates for the Pathophysiology of Facioscapulohumeral Dystrophy. *Hum. Mol. Genet* 18, 2414-2430 (2009).
26. Jones,T.I. et al. Facioscapulohumeral muscular dystrophy family studies of DUX4 expression:

- evidence for disease modifiers and a quantitative model of pathogenesis. *Hum. Mol. Genet.* 21, 4419-4430 (2012).
27. Krom, Y.D. et al. Generation of isogenic D4Z4 contracted and noncontracted immortal muscle cell clones from a mosaic patient: a cellular model for FSHD. *Am. J. Pathol.* 181, 1387-1401 (2012).
 28. Tassin, A. et al. DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? *J. Cell Mol. Med.* (2012).
 29. Asp, P. et al. Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proc. Natl. Acad. Sci. U. S. A* 108, E149-E158 (2011).
 30. Nozawa, R.S. et al. Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat. Struct. Mol. Biol.* 20, 566-573 (2013).
 31. McCabe, M.T. et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* 492, 108-112 (2012).
 32. Reddington, J.P. et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* 14, R25 (2013).
 33. Juan, A.H., Kumar, R.M., Marx, J.G., Young, R.A., & Sartorelli, V. Mir-214-dependent regulation of the polycomb protein Ezh2 in skeletal muscle and embryonic stem cells. *Mol. Cell* 36, 61-74 (2009).
 34. Zeng, W. et al. Genetic and Epigenetic Characteristics of FSHD-Associated 4q and 10q D4Z4 that are Distinct from Non-4q/10q D4Z4 Homologs. *Hum. Mutat.* 35, 998-1010 (2014).
 35. Simon, J.A. & Kingston, R.E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* 10, 697-708 (2009).
 36. Cherblanc, F.L., Chapman, K.L., Brown, R., & Fuchter, M.J. Chaetocin is a nonspecific inhibitor of histone lysine methyltransferases. *Nat. Chem. Biol.* 9, 136-137 (2013).
 37. Stadler, G. et al. Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat. Struct. Mol. Biol.* 20, 671-678 (2013).
 38. Grolimund, L. et al. A quantitative telomeric chromatin isolation protocol identifies different telomeric states. *Nat. Commun.* 4, 2848 (2013).
 39. Yao, Z. et al. Comparison of endogenous and overexpressed MyoD shows enhanced binding of physiological bound sites. *Skelet. Muscle* 3, 8 (2013).

Acknowledgements

The authors would like to thank all patients and their families for participating in our study. Lentiviral cloning and shRNA experiments were supported by Dr. Rob Hoeven and Martijn Rabelink, Department of Molecular and Cellular Biology, Leiden University Medical Center. This study was supported by grants from the US National Institutes of Health (NIH) (National Institute of Neurological Disorders and Stroke (NINDS) P01NS069539, and National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) R01AR045203 and R01AR066248, the Prinses Beatrix Spierfonds (W.OR12-20 and W.OR14-04), the Muscular Dystrophy Association (MDA; 217596), the Gerald Norton and Eklund family foundation, the FSH Society, the FSHD Global Research Foundation, and Spieren voor Spieren.

Disclosure of potential conflicts of interest

All authors declare to have no potential conflicts of interest.

Supplemental material

Table S1: Overview of patient derived muscle cell cultures and their experimental use.

Sample:	D4Z4 alleles (chromosome 4):	Used for:
Control 1	27U 4B168; 38U 4A161	ChIP, expression
Control 2 #	9U 4B163; 32U 4A161	ChIP, expression, GSK126 treatment
Control 3	13U 4A161; 21U 4B168	ChIP, expression
Control 4	24U 4B168; 49U 4B163	ChIP, expression
Control 5	13U 4A161; 21U 4B168	ChIP, expression
Control 6	18U 4A161; 29U 4B163	ChIP
Control 7	13U 4A161; 21U 4B168	expression
Control 8 #	20U 4A161; 24U 4A161	GSK126 treatment
Control 9	18U 4B163; 74U 4A161	GSK126 treatment
FSHD1 1†	6U 4A161; 52U 4B168	ChIP, expression, GSK126 treatment
FSHD1 2	8U 4A161; 23U 4A163	ChIP
FSHD1 3	4U 4A161; 13U 4B163	ChIP, expression
FSHD1 4	6U 4A161; 26U 4B162	ChIP, expression, GSK126 treatment
FSHD1 5	8U 4A161; 23U 4B168	ChIP, expression
FSHD1 6†	6U 4A161; 11U 4A168	Expression
FSHD1 7	3U 4A161; 16U 4B163	GSK126 treatment
FSHD1 8	2U 4A161; 35U 4B163	GSK126 treatment
FSHD2 1	11U 4A161; 39U 4B168	ChIP, expression
FSHD2 2†	11U 4A161; 17U 4A161	ChIP, expression, GSK126 treatment
FSHD2 3*	13U 4A161; 15U 4B163	ChIP, expression
FSHD2 4†	12U 4A161; 22U 4A161	ChIP, expression, GSK126 treatment
FSHD2 5	12U 4A161; 35U 4A166H	ChIP, expression,
FSHD2 6	26U 4A161; 38U 4A166H	ChIP, expression
FSHD2 7	14U 4A161; 65U 4A161	ChIP, expression, GSK126 treatment
FSHD2 8	14U 4A161; 20U 4A161	expression
FSHD2 9	19U 4A161; 101U 4B168	expression
FSHD2 10	18U 4A161; 47U 4A161	expression
FSHD2 11	11U 4A161; 35U 4A166H	GSK126 treatment
# Control sample used for knockdown experiments		
† FSHD samples used for ectopic SMCHD1 expression		
* Sample selected on hypomethylation, no SMCHD1 mutation detected		

Table S2: Overview of fibroblast cell cultures used in the study.

Sample:	D4Z4 alleles (chromosome 4):
Control 1	20U 4A161; 24U 4A161
Control 2	12U 4B163; 23U 4A161
FSHD1 1	3U 4A161; 16U 4B163
FSHD1 2	7U 4A161; 15U 4B163
FSHD2 1	14U 4A161; 65U 4A161
FSHD2 2	14U 4A161; 17U 4A161

Table S3: Overview of primer sequences and shRNA constructs used in the study.

List of primers used in the study			
Locus	Forward	Reverse	Template
DUX4	TCCAGGAGATGTAACTCTAATCCA	CCCAGGTACCAGCAGACC	cDNA
ZSCAN4	TGGAAATCAAGTGCCAAAAA	CTGCATGTGGACGTGGAC	cDNA
RFLP2	CCCACATCAAGGAACTGGAG	TGTTGGCATCCAAGTCATA	cDNA
TRIM43	ACCCATCACTGGACTGGTGT	CACATCCTCAAAGAGCCTGA	cDNA
MYOG	GCCAGACTATCCCCTTCCTC	GGGGATGCCCTCTCCTCTAA	cDNA
GUSB	CCGAGTGAAGATCCCCTTTTAA	CTCATTGGGAATTTGCCGATT	cDNA
GAPDH	GAGTCAACGGATTGGTCGT	TTGATTTGGAGGGATCTCG	cDNA
qD4Z4	CCGCGTCCGTCCGTGAAA	TCCGTCGCCGTCTCTGTC	ChIP DNA

List of shRNA constructs (Mission shRNA library, Sigma) used in the study		
Target	#	Library identifier
control		pSHC002
SUV39H1	1	TRCN0000158337
	2	TRCN0000157251
	3	TRCN0000150662
	4	TRCN0000157285
SMCHD1	1	TRCN0000253778
	2	TRCN0000253776
SMC3	1	TRCN0000159635
	2	TRCN0000234318
RAD21	1	TRCN0000147898
	2	TRCN0000148279

Fig. S1: MYOG induction upon myogenic differentiation.

Fig. S2: Immunofluorescence analysis of sporadic DUX4 expression in FSHD2 myotubes.

Fig. S3: DUX4 transcripts are detected in FSHD1 and FSHD2 fibroblasts transduced with MyoD.

Fig. S4: Graphical overview of the D4Z4 repeat unit and the DUX4 open reading frame.

Fig. S5: SMCHD1 and SUV39H1 transcript levels are significantly lower in human primary myoblast treated with specific shRNAs than in samples treated with control shRNA.

Fig. S6: H3K9me3 levels at qD4Z4 upon depletion of SUV39H1 in control myotubes.

Fig. S7: Abundance of the repressive histone mark H3K9me3 does not show disease specific differences.

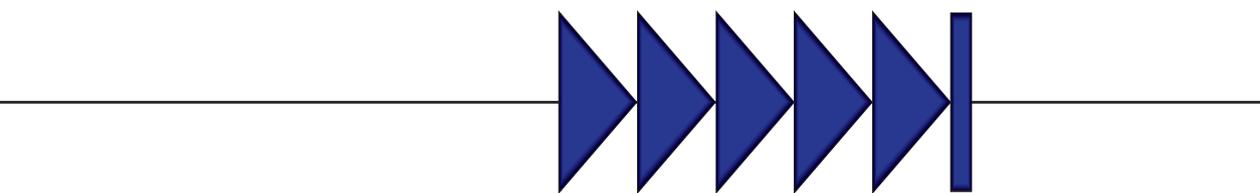
Fig. S8: Cohesin abundance at D4Z4 in patient samples is similar to levels measured in controls.

Fig. S9: MYOG expression levels upon SMCHD1 overexpression.

Fig. S10: Higher levels of H3K27me3 in FSHD2 derived myoblasts.

All supplemental figures belonging to this chapter can be accessed through <http://go.gl/oAeg57> or by using the QR-code below.





**DUX4 promotes transcription of FRG2
by directly activating its promoter
in Facioscapulohumeral muscular
dystrophy.**

5

Peter E. Thijssen, Judit Balog, Zizhen Yao, Tan Phát Pham, Rabi Tawil, Stephen J. Tapscott and Silvère M. Van der Maarel

Thijssen et al. (2014), *Skeletal Muscle*, 4:19

Abstract

Background

The most common form of facioscapulohumeral muscular dystrophy (FSHD) is caused by a genetic contraction of the polymorphic D4Z4 macrosatellite repeat array in the subtelomeric region of chromosome 4q. In some studies, genes centromeric to the D4Z4 repeat array have been reported to be over-expressed in FSHD, including *FRG1* and *FRG2*, presumably due to decreased long-distance repression by the shorter array through a mechanism similar to position-effect variegation. Differential regulation of *FRG1* in FSHD has never been unequivocally proven, however, *FRG2* has been reproducibly shown to be induced in primary FSHD-derived muscle cells when differentiated in vitro. While the molecular function of *FRG2* remains unclear and its contribution to FSHD pathology remains uncertain, recent evidence has identified the mis-expression of *DUX4*, located within the D4Z4 repeat unit, in skeletal muscle as the cause of FSHD. *DUX4* is a double homeobox transcription factor that has been shown to be toxic when expressed in muscle cells.

Results

In this study, we show that *DUX4* directly activates the expression of *FRG2*. Increased expression of *FRG2* was observed following expression of *DUX4* in myoblasts and fibroblasts derived from control individuals. Moreover, we identified *DUX4* binding sites at the *FRG2* promoter by chromatin immunoprecipitation followed by deep sequencing and confirmed the direct regulation of *DUX4* on the *FRG2* promoter by luciferase reporter assays. Activation of luciferase was dependent on both *DUX4* expression and the presence of the *DUX4* DNA binding motifs in the *FRG2* promoter.

Conclusion

We show that the FSHD-specific upregulation of *FRG2* is a direct consequence of the activity of *DUX4* protein rather than representing a regional de-repression secondary to fewer D4Z4 repeats.

Background

Facioscapulohumeral muscular dystrophy (FSHD, OMIM 158900/158901) is one of the most common myopathies with a prevalence of 1 in 8.000 according to a recent report¹. Individuals with FSHD typically suffer from progressive weakening and wasting of the facial, and upper extremity muscles with considerable inter- and intra-familial variability in disease onset and progression². The pathogenic mechanism of FSHD has been linked to the polymorphic D4Z4 macrosatellite repeat array, located on chromosome 4q35, of which each unit contains a copy of a retrogene encoding for the germline transcription factor double homeobox protein 4 (*DUX4*)³⁻⁵. *DUX4* has been shown to regulate a set of germline, early development and innate immune response related genes and leads to increased levels of apoptosis when expressed in muscle cells⁶⁻⁸. Over recent years, a combination of detailed genetic and functional analyses in FSHD families and muscle biopsies has established that sporadic expression of the *DUX4* retrogene in skeletal muscle is a feature shared by all individuals suffering from FSHD^{9, 10}.

Genetically, at least two forms of FSHD can be recognized. The majority of affected individuals (FSHD1, > 95%) are characterized by a contraction of the D4Z4 repeat array. In healthy individuals, the D4Z4 macrosatellite repeat array consists of 11-100 copies, whereas individuals with FSHD1 have at least 1 contracted allele of 1-10 repeat units^{4, 5, 11}. A second group of FSHD individuals (FSHD2, < 5%) do not show a contraction of the D4Z4 repeat array, however most often have mutations in the chromatin modifier structural maintenance of chromosomes hinge domain containing protein 1 (SMCHD1) on chromosome 18p¹². Both groups share two important (epi-)genetic features: they carry an allele permissive for stable *DUX4* transcription because of the presence of a polymorphic *DUX4* polyadenylation signal (PAS) and they display epigenetic derepression of the D4Z4 repeat array in somatic tissue^{9, 13}. More specifically, in muscle biopsies and muscle cell cultures *DUX4* expression has been correlated with decreased levels of CpG methylation and repressive histone modifications together with increased levels of transcriptional permissive chromatin markers at D4Z4^{10, 14-17}. The epigenetic changes at D4Z4 can be either attributed to repeat array contraction (FSHD1) or loss of SMCHD1 activity at D4Z4 (FSHD2)².

The time interval between the genetic association of FSHD to the D4Z4 macrosatellite repeat array and the identification of sporadic *DUX4* activation as a unifying disease mechanism encompasses almost 20 years of research into different candidate genes for FSHD^{4, 9}. In the absence of a conclusive disease mechanism, genes proximal to the D4Z4 repeat have been investigated as possible FSHD disease genes, postulating that their regulation is affected in FSHD through a position effect emanating from the D4Z4 repeat array¹⁶⁻²². Amongst those candidate genes were FSHD Region Gene 1 (*FRG1*) and FSHD Region Gene 2 (*FRG2*) (**Fig. 1A**). *FRG1* is located on chromosome 4 at 120kb proximal to the D4Z4 repeat and encodes a protein involved in actin bundle organization and mRNA biogenesis and transport²³⁻²⁶. Its overexpression leads to a dystrophic phenotype in different animal models, probably by affecting actin bundling and splicing of transcripts encoding muscle effector proteins²⁷⁻³⁰. However, most studies have failed to demonstrate *FRG1* upregulation in FSHD muscle^{17, 19, 20, 31-40}. *FRG2* is a gene at 37kb

distance from the repeat encoding a nuclear protein of unknown function⁴¹. The distal end of chromosome 4 that contains the D4Z4 macrosatellite repeat array has been duplicated to chromosome 10^{42,43}. Consequently, due to its close proximity to the D4Z4 repeat array, *FRG2* is located on both chromosomes 4 and 10 (**Fig. 1A**). Additionally, a complete copy of *FRG2* has been identified on the short arm of chromosome 3. We and others have previously reported on FSHD-specific transcriptional upregulation of *FRG2* from both the 4q and 10q copies upon in vitro myogenesis, however its overexpression did not lead to a dystrophic phenotype in a transgenic mouse model^{17,27,31,41}.

Until the discovery that mis-expression of *DUX4* is shared by all FSHD individuals, these observations led to a disease model in which the contraction of the D4Z4 repeat array would create a position effect on proximal genes, thereby leading to their transcriptional activation. Such a mechanism cannot explain the activation of the 10q copy of *FRG2*, as this would require a trans-effect of the contracted repeat array. Although the D4Z4 copies on 4q and 10q have been shown to interact in interphase nuclei⁴⁴, this seems unlikely as 3D FISH approaches have revealed that the 4q D4Z4 repeat localizes to the nuclear periphery, whereas the 10q subtelomere does not⁴⁵. More recently, it was shown that the expression of *FRG2* in FSHD cells was influenced by telomere length through telomere position effects⁴⁶, leading to the conclusion that *DUX4* and *FRG2* were independently regulated by telomere-length. In the current study we provide evidence that the activation of *FRG2* is a direct consequence of *DUX4* protein activity, providing an experimentally supported cause for its specific expression in FSHD muscle and reconfirming *DUX4* as the FSHD disease gene.

Methods

Cell culture

Human primary myoblast cell lines were obtained from the University of Rochester biorepository (<http://www.urmc.rochester.edu/fields-center/>) and were expanded and maintained in DMEM/F-10 (#31550 Gibco/Life Technologies, Bleiswijk, The Netherlands) supplemented with 20% heat inactivated fetal bovine serum (FCS #10270 Gibco), 1% penicillin/streptomycin (#15140 Gibco), 10ng/ml rhFGF (#G5071 Promega, Leiden, The Netherlands) and 1 μ M dexamethasone (#D2915 Sigma-Aldrich, Zwijndrecht, The Netherlands). Differentiation into myotubes was started at 80% confluency by serum starvation in DMEM/F-12 Glutamax (#31331, Gibco) supplemented with 2% KnockOut serum replacement formulation (#10828 Gibco) for 36 hours. All samples and their characteristics used for our study are listed in **additional table 1**. Rhabdomyosarcoma TE-671 were maintained in DMEM (#31966) supplemented with 10% FCS and 1% P/S, (all Gibco).

RNA isolation, cDNA synthesis and qRT-PCR

Cells were harvested using Qiazol lysis reagent (#79306 Qiagen N.V., Venlo, The Netherlands) and RNA was subsequently isolated using the miRNeasy Mini Kit (#217004 Qiagen) including an on column DNase treatment according to the manufacturer's instructions. 2 μ g RNA was used to synthesize poly-dT primed cDNA using the RevertAid

H Minus First strand cDNA Synthesis Kit (#K1632 Thermo Fischer Scientific Inc., Waltham, MA). Relative FRG2 expression was quantified on the CFX96 system (Bio-Rad, Venendaal, The Netherlands) using SYBR green master mix (Bio-Rad) with the following primers: FRG2_Fw: GGGAAACTGCAGGAAAA, FRG2_Rv: CTGGACAGTTCCTGCTGTGT. For relative quantification GUSB and GAPDH were used as reference genes and amplified with the following primers: GAPDH_Fw: GAGTCAACGGATTGGTCTGT, GAPDH_Rv: TTGATTTGGAGGGATCTCG, GUSB_Fw: CCGAGTGAAGATCCCCTTTTAA, GUSB_Rv: CTCATTTGGAATTTTCCGATT. All PCR reactions were carried out in duplicate and the data were analyzed using the Bio-Rad CFX manager version 3.0 (Bio-Rad).

Luciferase reporter assays

Genomic fragments containing the FRG2 promoter were amplified by regular PCR (primers pFRG2_Fw: AGGCCTTACCTTGCCTTTGT; pFRG2_Rv: TCTTGCTGGTGGATGTTGAG) using cosmids 23D11 (chromosome 10) and cY34 (chromosome 4)^{47, 48}. The obtained PCR fragments containing the promoter sites were digested with BglII and BclI and subcloned into the BglII digested pGL3 basic vector. Genomic locations of the cloned promotersites (UCSC hg19): chr4:190,948,283-190,949,163 and chr10:135,440,170-135,441,050. DUX4 binding sites were deleted by ligating PCR products obtained with internal primers (Fw_internal: ATCTGAGGGCCCTGATTCTGAGGTAGC, Rv_internal: ATCTGAGGGCCCCATTTTAAAGGTAGGAAGG) combined with RV3 and GL2 primers annealing in the pGL3 backbone. Single binding sites were destroyed using site directed mutagenesis by PCR amplifying overlapping fragments with the following primers:

Site 1Fw: CCTCAGGAATCAGGGGCTACATAGGGTAGCACTGACTCAACCT

Site 1Rv: AGGTTGAGTCAGTGCTACCCTATGTAGCCCCCTGATTCTGAGG

Site 2Fw: GGCTAATTAGTTAGCACTGACTCACCTATGCAATTCAATTTTATTGCATTGATC

Site 2Rv: GATCAAATGCAATAAAATGAATTGCATAGGGTGAGTCAGTGCTAACCTAATTAGCC

Site 3Fw: ACCTAATCAATTCAATTTTATTGCATTGCACTAAGTATCTTCCCCATTTTAAAGGTAGGAAGG

Site 3Rv: CCTTCTACCTTAAAAATGGGGAAGATACTTAGTGCAAATGCAATAAAATGAATTGATTAGGT

together with RV3 and GL2 primers. Insert sequences and correct orientation in the pGL3 vector were confirmed by Sanger sequencing. 60.000 TE671 cells were seeded in standard 24 well tissue culture plates and co-transfected with 200 ng pCS2/pCS2-DUX4 and 200 ng of the indicated pGL3 constructs, using lipofectamine 2000 according to manufacturer's instructions. 24 hours after transfection cell lysates were harvested and luciferase activity was measured using the Promega luciferase assay kit, according to manufacturer's instructions. Co-transfections with Renilla luciferase constructs for data normalization were omitted as we previously observed regulation of this construct by DUX4⁴⁹. Transfections were carried out in triplicate, error bars indicate the SEM of three independent experiments.

RNA sequencing and ChIP sequencing

RNA sequencing and ChIP sequencing data were obtained and analyzed as described before^{6, 36}. All datasets have previously been made publicly available in the Gene Expression Omnibus (accession numbers GSE56787, GSE33838). FRG2 expression in response to DUX4 overexpression is displayed for MB135, a control derived primary

myoblast, and a control derived fibroblast. The genomic snapshots of the different datasets were generated using the IGV genome viewer version 2.3.32^{50, 51}.

Results

FRG2 expression is activated in differentiating FSHD derived muscle cells

We cultured a set of primary muscle cells derived from six controls, six FSHD1 and nine FSHD2 individuals and harvested RNA to analyze FRG2 transcript levels by quantitative realtime-PCR (qRT-PCR). We confirmed the significant FSHD-specific activation of FRG2 in differentiating myotubes (**Fig. 1B**). In control samples, we observed a minor increase in FRG2 expression that was not statistically significant (**Fig. 1B**). Analysis of previously

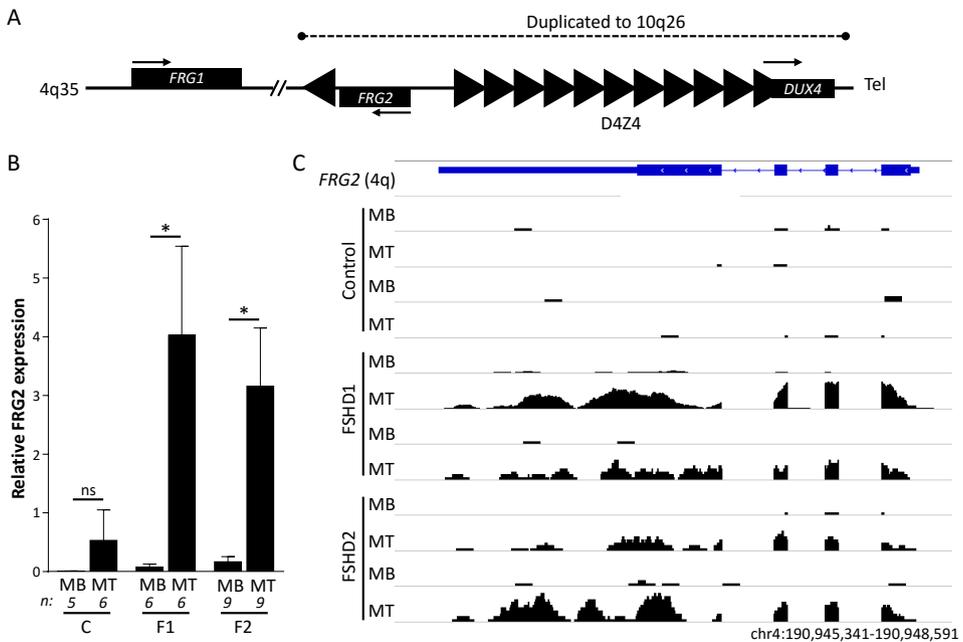


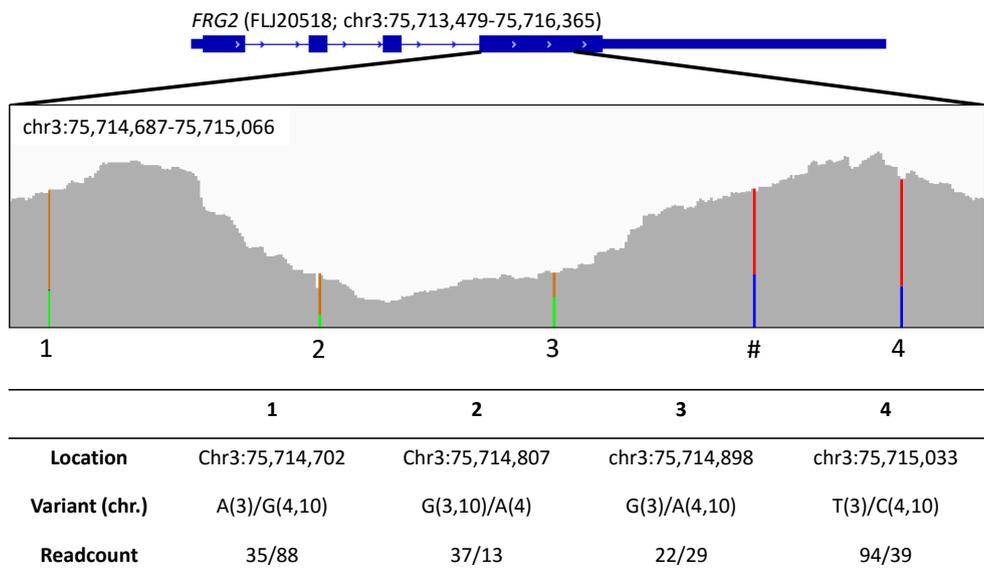
Figure 1. FRG2 activation in FSHD derived differentiating myoblasts

A) Schematic representation of the FSHD locus on chromosome 4q35. Rectangles indicate the different genes, arrows their transcriptional direction. Triangles represent D4Z4 repeat units and the single inverted repeat unit upstream of FRG2. Each unit contains the full DUX4 ORF, only the last repeat unit produces a stable transcript in FSHD patients. The dashed line indicates the duplicated region present on chromosome 10q26. **B)** qRT-PCR analysis of mean FRG2 expression levels in control (C), FSHD1 (F1) and FSHD2 (F2) derived proliferating myoblasts (MB) and differentiating myotubes (MT) shows the significant activation of FRG2 during differentiation only in F1 and F2 derived cells. Relative expression was determined using GAPDH and GUSB as reference genes. Sample numbers are indicated and error bars represent the standard error of the mean. Asterisks indicate significant differences based on a one-way ANOVA ($P = 0.0014$), followed by pairwise comparison using bonferroni correction, NS = non-significant. **C)** Genomic snapshot (location indicated at the bottom) of RNA sequencing data of two control, two FSHD1 and two FSHD2 derived proliferating myoblasts (MB) and differentiating myotubes (MT) confirms the full length expression of FRG2 in differentiating myotubes originating from FSHD individuals.

published RNA-seq data from two additional controls and a subset of FSHD samples confirmed FRG2 activation (**Fig. 1C**) and single base pair variations, known to differ between the three copies of *FRG2*⁴¹, indicated transcripts were induced from *FRG2* genes at all three genomic locations (**Fig. 2**). Therefore, we conclude that increased transcription of *FRG2* is not restricted to the copy on the 4q disease allele in FSHD1 and likely not caused by a *cis* effect of D4Z4 chromatin relaxation. The increase of FRG2 transcript levels coincided with activation of DUX4 transcription upon *in vitro* myogenesis in FSHD derived samples (**additional Fig. 1**), as was reported before by us and others^{36, 52, 53}. Altogether this confirms previously published data and highlights the robust transcriptional activation of all annotated copies of *FRG2* in differentiating FSHD myotubes.

FRG2 activation is a direct consequence of DUX4 protein activity

As the activation of *FRG2* in FSHD-derived myotubes follows the pattern of previously identified DUX4 target genes, we wondered if *FRG2* is regulated by DUX4 directly. Indeed, we previously showed that FRG2 transcription was induced at least twofold by expression array analysis in DUX4 over-expressing control myoblasts⁶. This robust increase in FRG2 transcription was confirmed by RNA-seq in both myoblasts and fibroblasts (**Fig. 3A**) that were transduced with DUX4 expressing lentiviruses, ruling



Based on UCSC genes uc003dpt.4 (FLJ20518; Chr3), uc003izv.3 (LOC100288255, Chr4) & uc010qvg.2 (FRG2B, Chr10)

#: T/C variant, reference sequences of FRG2 do not differ at this position

Figure 2. sequence analysis of RNA sequencing reads reveals activation of FRG2 from all copies

Graphical representation of RNA-seq reads mapping to the FRG2C locus at chromosome 3p. Single nucleotide polymorphisms can be identified and are indicated by colored vertical lines (A = green, C = blue, G = orange, T = red). Different reads could thereby be assigned to the three different genomic copies of FRG2. Sequence analysis was based on the reference sequences obtained from the UCSC genome browser (build 19).

out a muscle specific effect of DUX4 on *FRG2* expression. Direct targets of DUX4 were previously identified by overlaying expression data with chromatin immuno-precipitation (ChIP) sequencing data⁶. Following the same approach, we observed DUX4 binding at the promoter of *FRG2* (Fig. 3B). Sequence analysis of the 4q and 10q copies of the *FRG2* promoter revealed that both chromosomes harbor three consensus binding sites

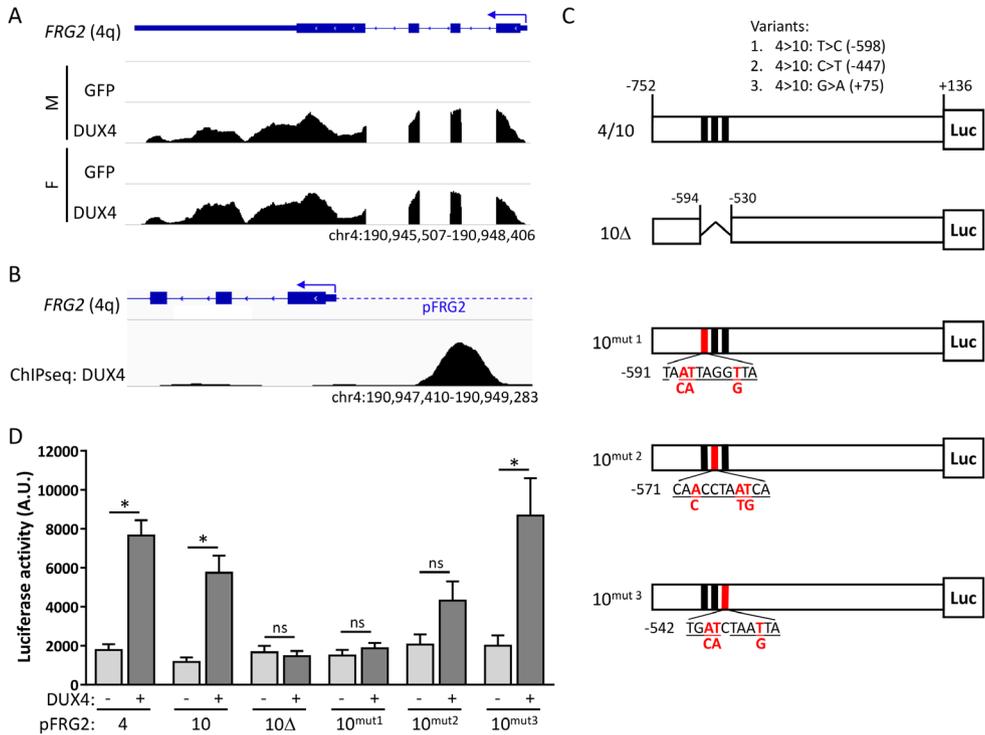


Figure 3. *FRG2* is activated as a consequence of DUX4 protein activity at its promoter

A) Genomic snapshot (location indicated at the bottom) of mapped RNA-seq reads at the chromosome 4q *FRG2* locus. Overexpression of DUX4 results in the activation of *FRG2* in myoblasts (M) and fibroblasts (F), GFP overexpression was used as a control. **B)** Graphical representation of DUX4 binding at the 4q *FRG2* promoter (genomic snapshot location indicated at the bottom) as revealed by ChIP-seq analysis. (data obtained in myoblasts). **C)** Genomic fragments obtained from chromosomes 4 and 10 were cloned upstream of the luciferase gene. Polymorphisms distinguishing both copies are indicated (Variants 1-3) and the three identified DUX4 binding sites are displayed, with nucleotides matching the previously identified core DUX4 binding sequence underlined. All numbers show the relative distance to the TSS of *FRG2*, in the 10^{mut1-3} constructs the number indicates the first displayed nucleotide. The 10Δ construct lacks all three DUX4 binding sites, whereas in 10^{mut1-3} the red nucleotides were mutated to the nucleotides indicated below them, thereby destroying the individual DUX4 binding sites. **D)** DUX4 activates the *FRG2* promoter in a luciferase reporter assay. Both the 4q and 10q copy of the *FRG2* promoter are activated by DUX4. The 10q copy lacking the DUX4 binding sites (10Δ) was not activated by DUX4. Destruction of the three individual binding sites revealed that sites 1 and 2 are mediating the DUX4 dependent activation of *FRG2*. Counts per second (CPS) are a direct measure of luciferase activity, error bars indicate the SEM of three independent experiments. Asterisks indicate significant differences based on a one-way ANOVA ($P < 0.0001$), followed by pairwise comparison using bonferroni correction, NS = non-significant.

for DUX4, which are not affected by minor sequence differences between the two loci (**Fig. 3C**). To test the functional significance of these DUX4 binding sites, we designed luciferase reporter constructs harboring the *FRG2* promoters of chromosomes 4 and 10. Co-transfection of these constructs with pCS2-DUX4 or the empty pCS2 backbone in TE-671 rhabdomyosarcoma cells confirmed that the activation of *FRG2* is mediated by DUX4 protein expression (**Fig. 3D**). To confirm that the activation of the luciferase reporter gene was indeed mediated by DUX4 binding, we generated a reporter construct with a micro-deletion of all three DUX4 binding sites in the *FRG2* promoter derived from chromosome 10 (10Δ, **Fig. 3C**). Upon co-transfection with the PCS2-DUX4 expression vector, the luciferase activation was completely ablated (**Fig. 3D**). We identified three consecutive DUX4 binding sites in the *FRG2* promoter, of which sites one and three contain a single base pair variation in the core sequence identified previously⁶. To dissect which sites were responsible for DUX4 dependent *FRG2* activation, we designed three additional constructs, in which one of the three sites was destroyed by site directed mutagenesis (**Fig. 3C**). Upon transfection of these constructs we observed that the activation of *FRG2* by DUX4 is mediated primarily through the two binding sites furthest from the *FRG2* transcriptional start site (TSS), as luciferase activity in response to DUX4 expression was no longer significantly induced. While destruction of the first binding site resulted in a complete absence of luciferase activation, the effect of DUX4 is still moderate, but non-significant, if site two is mutated (**Fig. 3D**). Destroying the third binding site did not affect the DUX4 mediated activation of the *FRG2* promoter (**Fig. 3D**). Altogether, our data shows that DUX4 binding at the *FRG2* promoter is underlying the transcriptional activation of *FRG2* in FSHD derived myogenic cultures.

Discussion

Ever since the D4Z4 macrosatellite repeat array was genetically associated with FSHD, great effort has been put into identifying the underlying disease mechanism. The initial lack of evidence for active transcription of *DUX4*, encoded in each D4Z4 repeat unit, shifted the focus to genes immediately centromeric to the array, like *FRG1* and *FRG2*. Although *FRG1* overexpression in mice leads to a dystrophic phenotype, its deregulation in muscles of FSHD patients has not been consistently demonstrated^{19, 27, 28, 31-34, 36-40}. In contrast, upregulation of *FRG2* was consistently reported in FSHD-derived differentiating muscle cells; however the mechanism of FSHD-specific upregulation of *FRG2* had not been conclusively established^{17, 31, 41}.

Activation of *FRG2* in FSHD cells was previously attributed to de-repression through a position effect mechanism secondary to the contraction of the D4Z4 repeat array^{17, 41}. Moreover, it was shown that *KLF15* regulates both *FRG2* expression and the activity of a putative enhancer in within D4Z4, thereby possibly facilitating a cis effect of the D4Z4 repeat array on the proximal *FRG2* locus⁵⁴. This model was challenged by showing that CpG methylation levels at a single CpG site centromeric to D4Z4 are unaffected in FSHD¹³, indicating that DNA methylation was not broadly altered in the region centromeric to the D4Z4 repeat array in FSHD cells. The recent establishment of a unifying disease mechanism for FSHD, that primarily centers around *DUX4*, further challenges a position

effect model that involves the deregulation of genes proximal to D4Z4 as a consequence of D4Z4 repeat array contraction. In line with this notion, we now show that *FRG2* is a direct target gene of DUX4, providing a direct explanation for its upregulation in FSHD muscle.

As sporadic *DUX4* activation is induced during *in vitro* muscle cell differentiation, the expression profile of *FRG2* fits that of previously reported DUX4 target genes. It is interesting to note that we observed increased, though not significantly, *FRG2* expression in control derived differentiating muscle cells, which might indicate a minimal activation of the locus during *in vitro* myogenesis in control cells. DUX4 expression has been reported to occur in control derived myogenic cultures, albeit at much lower frequencies, and thus *FRG2* may be activated as a consequence of that⁵⁵. Alternatively, *FRG2* expression may be sporadically induced through other mechanisms, exemplified by the reported regulation by KLF15⁵⁴.

DUX4 acts a potent transcriptional activator and induces expression of germline and early development genes through binding a specific homeobox sequence⁶. ChIP-seq analysis indeed identified DUX4 binding at the promoter of *FRG2*, which contains three consecutive DUX4 binding sites. Our experiments showed that the two sites furthest from the TSS of *FRG2* are mediating the activation of *FRG2* upon the expression of DUX4, confirming our earlier work showing that the probability of transcriptional activation by DUX4 increases with the number of consecutive DUX4 binding sites⁵⁶. The *FRG2* promoter sites at chromosomes 3, 4 and 10 are highly conserved, with identical sequences for the DUX4 sites on 4 and 10 and a limited number of single nucleotide differences between these and chromosome three that are predicted to preserve DUX4 binding, explaining the activation of all copies by DUX4⁴¹. In contrast to the complicated proposed mechanism of cis and trans effects of D4Z4 contraction at 4q^{17,41}, the protein activity of DUX4 offers a simple and experimentally supported explanation for the activation of *FRG2*.

It was previously shown that *FRG2* and *DUX4* are regulated, at least in part, by telomere position effects⁴⁶. Trans-activation of the *FRG2* promoter was ruled out by transfecting promoter reporter constructs in immortalized FSHD derived myoblasts. However, the sporadic nature of *DUX4* expression in these cells would seriously decrease the signal-to-noise ratio in this assay and a direct effect of DUX4 on *FRG2* expression would likely have been missed. Although we cannot rule out a direct telomere position effect on *FRG2* in this study, we suggest that the observed increase of *FRG2* can be attributed to the increased DUX4 levels rather than to telomere position effects on *FRG2*. DUX4 itself may indeed be partially under control of telomere length and as such its target genes would follow a similar regulation.

As of yet, the functional consequence of *FRG2* activation in FSHD remains elusive. *FRG2* localizes to the nucleus⁴¹, but its function has never been demonstrated and a possible role in FSHD disease progression is therefore unclear. The identification of FSHD individuals with proximally extended deletions, in which not only a large part of the D4Z4 repeat array, but also proximal sequences (including *FRG2*) were deleted,

again suggests that a cis-acting effect on *FRG2* is not necessary for FSHD pathology^{57,58}. However, since DUX4 activates other genomic copies of *FRG2*, it remains possible that this protein contributes to some aspect of the disease.

Conclusion

In this study we have firmly established that the long known activation of *FRG2* in FSHD derived differentiating muscle cells is a direct consequence of DUX4 activity at its promoter. This provides further evidence for DUX4 as the central player in the FSHD disease mechanism and demonstrates that the higher expression of *FRG2* in FSHD does not result from regional de-repression secondary to fewer *D4Z4* repeats.

List of abbreviations

FSHD: Facioscapulohumeral muscular dystrophy; *FRG1*: FSHD Region Gene 1; *FRG2*: FSHD Region Gene 2; *DUX4*: Double Homeobox 4; *SMCHD1*: Structural Maintenance of chromosomes flexible hinge domain containing 1; PAS: Polyadenylation signal; FISH: Fluorescence in situ hybridization; FBS: Fetal bovine serum; rhFGF: recombinant human fibroblast growth factor; SEM: Standard error of the mean; ChIP-seq: Chromatin immunoprecipitation followed by deep sequencing; qRT-PCR: quantitative realtime polymerase chain reaction; RNA-seq: RNA sequencing; KLF15: Krüppel-like factor 15; ORF: Open reading frame; MB/MT: myoblasts/myotubes; TSS: transcriptional start site; ANOVA: analysis of variance; CPS: counts per second.

References

1. Deenen, J.C. et al. Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* (2014).
2. Tawil, R., van der Maarel, S.M., & Tapscott, S.J. Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. *Skelet. Muscle* 4, 12 (2014).
3. Gabriels, J. et al. 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 (1999).
4. van Deutekom, J.C. et al. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* 2, 2037-2042 (1993).
5. Wijmenga, C. et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 2, 26-30 (1992).
6. Geng, L.N. et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38-51 (2012).
7. Kowaljow, V. et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord* 17, 611-623 (2007).
8. Wallace, L.M. et al. DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann. Neurol.* 69, 540-552 (2011).
9. Lemmers, R.J. et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650-1653 (2010).
10. Snider, L. et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS. Genet.* 6, e1001181 (2010).
11. Gilbert, J.R. et al. Evidence for heterogeneity in facioscapulohumeral muscular dystrophy (FSHD). *Am J Hum Genet* 53, 401-408 (1993).
12. Lemmers, R.J. et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* 44, 1370-1374 (2012).
13. de Greef, J.C. et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mutat.* 30, 1449-1459 (2009).
14. Balog, J. et al. Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. *Epigenetics.* 7, 579-584 (2012).
15. Zeng, W. et al. Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats in facioscapulohumeral dystrophy (FSHD). *PLoS. Genet* 7, e1000559 (2009).
16. Cabianca, D.S. et al. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* 149, 819-831 (2012).
17. Gabellini, D., Green, M., & Tupler, R. Inappropriate Gene Activation in FSHD. A Repressor Complex Binds a Chromosomal Repeat Deleted in Dystrophic Muscle. *Cell* 110, 339-248 (2002).
18. Winokur, S.T. et al. 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 Res* 2, 225-234 (1994).
19. Jiang, G. et al. Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum Mol. Genet* 12, 2909-2921 (2003).
20. Bodega, B. et al. 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. Biol.* 7, 41 (2009).
21. Petrov, A. et al. Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. *Proc. Natl. Acad. Sci. U. S. A* 103, 6982-6987 (2006).
22. Petrov, A. et al. A nuclear matrix attachment site in the 4q35 locus has an enhancer-blocking activity in vivo: implications for the facio-scapulo-humeral dystrophy. *Genome Res.* (2007).
23. van Deutekom, J.C.T. et al. Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. *Hum Mol Genet* 5, 581-590 (1996).
24. Liu, Q., Jones, T.I., Tang, V.W., Brieher, W.M., & Jones, P.L. Facioscapulohumeral muscular dystrophy region gene-1 (FRG-1) is an actin-bundling protein associated with muscle-attachment sites. *J. Cell Sci.* 123, 1116-1123 (2010).

25. Sun,C.Y. et al. Facioscapulohumeral muscular dystrophy region gene 1 is a dynamic RNA-associated and actin-bundling protein. *J. Mol. Biol.* 411, 397-416 (2011).
26. Hanel,M.L. et al. Facioscapulohumeral muscular dystrophy (FSHD) region gene 1 (FRG1) is a dynamic nuclear and sarcomeric protein. *Differentiation* 81, 107-118 (2011).
27. Gabellini,D. et al. Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature* 439, 973-977 (2006).
28. Hanel,M.L., Wuebbles,R.D., & Jones,P.L. Muscular dystrophy candidate gene FRG1 is critical for muscle development. *Dev. Dyn.* 238, 1502-1512 (2009).
29. Sancisi,V. et al. Altered Tnt3 characterizes selective weakness of fast fibers in mice overexpressing FSHD region gene 1 (FRG1). *Am. J. Physiol Regul. Integr. Comp Physiol* 306, R124-R137 (2014).
30. Pistoni,M. et al. Rbfox1 downregulation and altered calpain 3 splicing by FRG1 in a mouse model of Facioscapulohumeral muscular dystrophy (FSHD). *PLoS. Genet.* 9, e1003186 (2013).
31. Klooster,R. et al. Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. *Eur. J. Hum. Genet.* 17, 1615-1624 (2009).
32. Osborne,R.J., Welle,S., Venance,S.L., Thornton,C.A., & Tawil,R. Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology* 68, 569-577 (2007).
33. Cheli,S. et al. Expression profiling of FSHD-1 and FSHD-2 cells during myogenic differentiation evidences common and distinctive gene dysregulation patterns. *PLoS. ONE.* 6, e20966 (2011).
34. Winokur,S.T. et al. Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum Mol. Genet* 12, 2895-2907 (2003).
35. Masny,P.S. et al. Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei. *Eur. J. Hum. Genet.* 18, 448-456 (2010).
36. Yao,Z. et al. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum. Mol. Genet.*(2014).
37. Xu,X. et al. DNaseI hypersensitivity at gene-poor, FSH dystrophy-linked 4q35.2. *Nucleic Acids Res.* 37, 7381-7393 (2009).
38. Arashiro,P. et al. Transcriptional regulation differs in affected facioscapulohumeral muscular dystrophy patients compared to asymptomatic related carriers. *Proc. Natl. Acad. Sci. U. S. A* 106, 6220-6225 (2009).
39. Celegato,B. et al. 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 (2006).
40. Tsumagari,K. et al. Gene expression during normal and FSHD myogenesis. *BMC. Med. Genomics* 4, 67 (2011).
41. Rijkers,T. et al. FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. *J Med Genet* 41, 826-836 (2004).
42. Bakker,E. et al. The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve* 2, 39-44 (1995).
43. Deidda,G. et al. Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the Facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur J Hum Genet* 3, 155-167 (1995).
44. Stout,K. et al. Somatic pairing between subtelomeric regions: implications for human genetic disease? *Chrom Res* 7, 323-329 (1999).
45. Masny,P.S. et al. Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum Mol. Genet* 13, 1857-1871 (2004).
46. Stadler,G. et al. Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat. Struct. Mol. Biol.* 20, 671-678 (2013).
47. van Deutekom,J.C.T. et al. Search for the FSHD gene using cDNA selection in a region spanning 100 kb on chromosome 4q35. *Muscle Nerve Supplement*, S19-S26 (1995).
48. van Geel,M. et al. Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. *Genomics* 79, 210-7 (2002).
49. Krom,Y.D. et al. Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD. *PLoS. Genet.* 9, e1003415 (2013).
50. Thorvaldsdottir,H., Robinson,J.T., & Mesirov,J.P. Integrative Genomics Viewer (IGV): high-

- performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178-192 (2013).
51. Robinson, J.T. et al. Integrative genomics viewer. *Nat. Biotechnol.* 29, 24-26 (2011).
 52. Block, G.J. et al. Wnt/beta-catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. *Hum. Mol. Genet.* (2013).
 53. Krom, Y.D. et al. Generation of isogenic D4Z4 contracted and noncontracted immortal muscle cell clones from a mosaic patient: a cellular model for FSHD. *Am. J. Pathol.* 181, 1387-1401 (2012).
 54. Dmitriev, P. et al. The Kruppel-like factor 15 as a molecular link between myogenic factors and a chromosome 4q transcriptional enhancer implicated in facioscapulohumeral dystrophy. *J. Biol. Chem.* 286, 44620-44631 (2011).
 55. Jones, T.I. et al. Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. *Hum. Mol. Genet.* 21, 4419-4430 (2012).
 56. Young, J.M. et al. DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS. Genet.* 9, e1003947 (2013).
 57. Deak, K.L. et al. Genotype-phenotype study in an FSHD family with a proximal deletion encompassing p13E-11 and D4Z4. *Neurology* 68, 578-582 (2007).
 58. Lemmers, R.J.L.F. et al. Inter- and intrachromosomal subtelomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum Mol Genet* 7, 1207-1214 (1998).

Acknowledgements

The authors would like to express their gratitude to all the patients and their families who participated in this study. This study was financially supported by grants from the US National Institutes of Health (NIH) (National Institute of Neurological Disorders and Stroke (NINDS) P01NS069539, and National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) R01AR045203), the Gerald Norton and Eklund family foundation, the FSH Society, Spieren voor Spieren, and The Friends of FSH Research.

Competing interests

The authors declare that they have no competing interests.

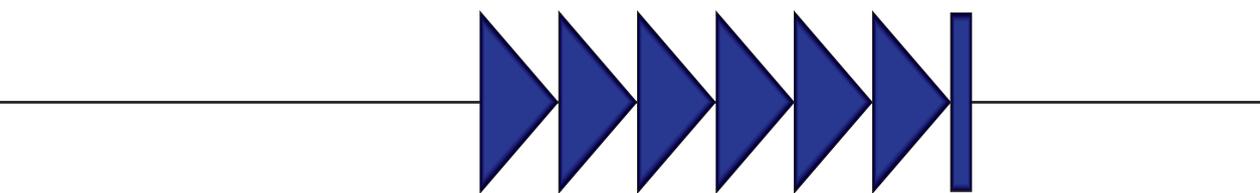
Additional files

Additional Fig. 1. RNA-sequencing reads mapping to D4Z4 showed DUX4 activation in differentiating FSHD derived muscle cells.

Additional table 1: Characteristics of samples used for qRT-PCR and RNA-seq analysis.

All additional files belonging to this chapter can be accessed through <http://goo.gl/Zlu1Bj> or by using the QR-code below.





**Intrinsic epigenetic regulation of
the D4Z4 macrosatellite repeat in a
transgenic mouse model for FSHD.**

6

Yvonne D. Krom*, Peter E. Thijssen*, Janet M. Young, Bianca den Hamer, Judit Balog, Zizhen Yao, Lisa Maves, Lauren Snider, Paul Knopp, Peter S. Zammit, Tonnie Rijkers, Baziel G.M. van Engelen, George W. Padberg, Rune R. Frants, Rabi Tawil, Stephen J. Tapscott and Silvère M. van der Maarel

Krom et al. (2013), PLoS Genetics, 9(4): e1003415.

* Authors contributed equally to this work.

Abstract

Facioscapulohumeral dystrophy (FSHD) is a progressive muscular dystrophy caused by decreased epigenetic repression of the D4Z4 macrosatellite repeats and ectopic expression of *DUX4*, a retrogene encoding a germline transcription factor encoded in each repeat. Unaffected individuals generally have more than 10 repeats arrayed in the subtelomeric region of chromosome 4, whereas the most common form of FSHD (FSHD1) is caused by a contraction of the array to fewer than 10 repeats, associated with decreased epigenetic repression and variegated expression of *DUX4* in skeletal muscle. We have generated transgenic mice carrying D4Z4 arrays from an FSHD1 allele and from a control allele. These mice recapitulate important epigenetic and *DUX4* expression attributes seen in patients and controls, respectively, including high *DUX4* expression levels in the germline, (incomplete) epigenetic repression in somatic tissue, and FSHD-specific variegated *DUX4* expression in sporadic muscle nuclei associated with D4Z4 chromatin relaxation. In addition we show that *DUX4* is able to activate similar functional gene groups in mouse muscle cells as it does in human muscle cells. These transgenic mice therefore represent a valuable animal model for FSHD and will be a useful resource to study the molecular mechanisms underlying FSHD and to test new therapeutic intervention strategies.

Author Summary

Facioscapulohumeral dystrophy (FSHD) is a progressive muscle disorder which is associated with contraction and chromatin relaxation of the D4Z4 macrosatellite repeat on chromosome 4q. Each unit of the repeat contains a copy of the primate-specific *DUX4* retrogene, encoding a germline transcription factor which is repressed in somatic tissue. In FSHD, somatic repression of the *DUX4* gene is compromised, leading to a variegated expression pattern of *DUX4* in muscle cells. The complex (epi)genetic etiology of FSHD has long hampered the generation of a faithful animal model and thus far the role of FSHD candidate genes has only been studied in model organisms by overexpression approaches. Here we present two transgenic mouse models containing either patient- or control-sized D4Z4 repeats. In our mice, the regulation of the FSHD locus is preserved in both lines and only in the disease model somatic derepression and variegated expression of *DUX4* is observed. These mice thus reflect many aspects of the complex regulation of *DUX4* expression in humans. These models may therefore become valuable tools in understanding the *in vivo* regulation and function of *DUX4*, its role in FSHD and the evaluation of therapeutic strategies.

Introduction

Each unit of the D4Z4 macrosatellite repeat contains a copy of the *DUX4* retrogene that encodes a double homeobox transcription factor¹⁻⁴. *DUX4* is highly expressed in the germline and epigenetically repressed in most somatic tissues, including skeletal muscle^{5, 6}. Recently, we and others showed that facioscapulohumeral dystrophy (FSHD), a muscular dystrophy predominantly affecting facial and upper extremity muscles⁷, is caused by D4Z4 repeat contraction-dependent (FSHD1) or –independent (FSHD2) chromatin relaxation in somatic tissues and low levels of *DUX4* mRNA expression in skeletal muscle^{5, 8-10}. On normal chromosomes 4, the D4Z4 repeat array varies between 11-100 units, while in FSHD1 one of the chromosomes 4 has an array of 1-10 units associated with a less repressive D4Z4 chromatin structure¹¹⁻¹³. In FSHD2, the D4Z4 repeats are not contracted and D4Z4 chromatin relaxation can be observed on all arrays⁸.

The low abundance of *DUX4* mRNA in FSHD muscle tissue represents a variegated pattern of expression with abundant *DUX4* protein expressed in a small number of nuclei^{6, 14}, presumably due to an occasional escape from the inefficient epigenetic repression. The polyadenylation (pA) site for *DUX4* mRNA is in the DNA sequence immediately telomeric to the last D4Z4 repeat unit and chromosome 4 haplotypes non-permissive for FSHD contain inactivating polymorphisms at the pA site, explaining the haplotype-specificity of this disease^{2, 15, 16}.

When expressed in skeletal muscle, the *DUX4* transcription factor activates genes normally expressed in the germline, essentially inducing a stem cell program in the postmitotic muscle cell. In addition, *DUX4* binds and transcriptionally activates endogenous retrotransposons and simultaneously blocks the innate immune response, at least in part through the transcriptional activation of a beta-defensin⁵.

The parental gene to primate *DUX4* was necessarily expressed in the germline, since germline retro-transposition was necessary for it to enter the primate lineage. As a retrogene, however, *DUX4* was dissociated from its evolved enhancers, promoters, and pA site, suggesting that the *DUX4* retrogene adopted independently evolved mechanisms to regulate its developmental expression¹⁷. One hypothesis is that the repression of *DUX4* transcription in most somatic tissues relies on an independently evolved mechanism of repeat-mediated silencing. If true, then it is to be expected that somatic *DUX4* silencing is an evolutionary conserved mechanism that can be recapitulated in other species such as mouse.

No studies have yet addressed whether the D4Z4 repeat array and its flanking sequence is sufficient to accurately reproduce the developmental pattern of *DUX4* expression, nor whether the FSHD mutation can recapitulate the decreased epigenetic repression and variegated *DUX4* expression in a mouse model. The latter question is particularly relevant as both primates and rodents have lost the parental copy of the *DUX4* retrogene¹⁸, but only primates have *DUX4* integrated in the context of a D4Z4 macrosatellite repeat array and it is not known whether integration of the human array in mice can – at least in part – recapitulate the molecular characteristics of FSHD.

Here we report the generation and molecular characterization of two transgenic mouse lines: one carrying a D4Z4 genomic region from a contracted pathogenic FSHD1 allele and one carrying a normal sized, non-pathogenic allele. Our data suggest that somatic epigenetic silencing of *DUX4* indeed is an evolutionary conserved mechanism and that contracted D4Z4 repeat arrays are silenced less efficiently, leading to a variegated expression pattern of DUX4 protein in skeletal muscle nuclei.

Results

Generation of transgenic mouse models

To determine whether the D4Z4 repeat with the *DUX4* retrogene contains the regulatory elements necessary for germline expression and copy-number dependent somatic epigenetic repression, we generated two transgenic mouse lines. One line carries an EcoRI fragment derived from the lambda-42 (L42) clone of an FSHD1 allele (**Fig. 1A**). This allele is of the FSHD-permissive 4A161 background, containing the *DUX4* pA signal, but lacking the more downstream exons 6 and 7⁶, and contains two-and-a-half copies of the D4Z4 unit and flanking sequences (herein referred to as D4Z4-2.5 mice)¹¹. Integration of the L42 clone at mouse chromosome 17 was confirmed by conventional and COBRA-FISH analyses and in total 4 copies of the EcoRI fragment were integrated as evidenced by MLPA analysis (**Fig. 1B-D**). The second transgenic mouse line was generated using two overlapping PAC clones, containing the upstream *FRG1* and *FRG2* genes, an array of 12.5 D4Z4 repeat units and flanking sequences, also harboring the FSHD-permissive 4A161 haplotype and lacking exons 6 and 7 (herein referred to as D4Z4-12.5 mice) (**Fig. 1E**)⁶. Recombination and single integration of the two PAC clones at chromosome 2 was confirmed by fiber-FISH, COBRA-FISH and MLPA analyses (**Fig. 1D, E, F**). Thus, D4Z4-2.5 mice have D4Z4 repeat lengths that cause FSHD in humans, whereas D4Z4-12.5 mice have an array length sufficient to maintain epigenetic silencing of *DUX4* in somatic tissue in humans.

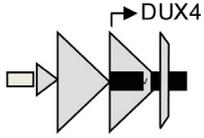
DUX4 expression in transgenic mice

We have previously reported that human *DUX4* is expressed in stem cells of the male germline⁶. In both D4Z4-2.5 and D4Z4-12.5 mice abundant levels of *DUX4* mRNA were observed in germ line tissues, most notably in testis (**Fig. 2**). In D4Z4-2.5 mice,

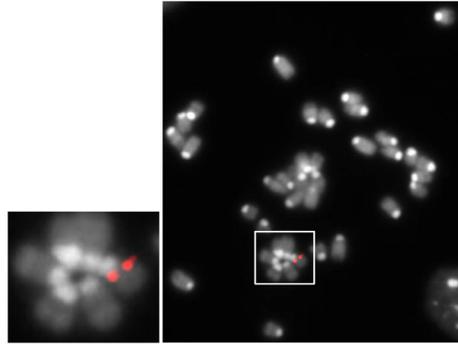
Figure 1. Integration site and copy number of D4Z4-2.5 and D4Z4-12.5 constructs in the mouse genome.

A) Schematic draw of the L42 EcoRI fragment used to generate the D4Z4-2.5 mouse line. **B)** Metaphase spread of D4Z4-2.5 fibroblasts co-stained with dapi and the CY3 labeled L42 probe shows integration at a single pair of chromosomes. **C)** COBRA-FISH analysis on D4Z4-2.5 fibroblast metaphase spreads probed with biotinylated-L42 fragments shows integration of L42 on chr17. **D)** Detection of copy number of the integrated fragments in both mouse models by MLPA analysis. The probe mix contained three probes specific for wild type alleles, one probe designed against the human p13E-11 region and one probe against D4Z4. **E)** Schematic draw of PAC clones used to generate the D4Z4-12.5 mouse. **F)** COBRA-FISH analysis on D4Z4-12.5 fibroblast metaphase spreads probed with a biotinylated PAC clone shows integration of the PAC clone on chr2. **G)** Fiber-FISH analysis of D4Z4-12.5 fibroblasts. Both PAC clones, labeled and hybridized to DNA fibers, were shown to be recombined during integration into the mouse genome.

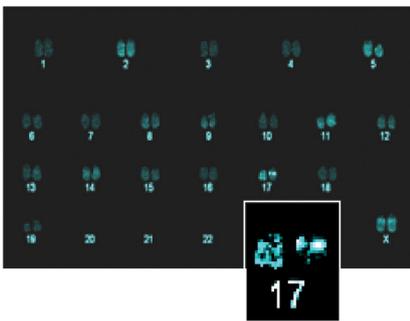
A



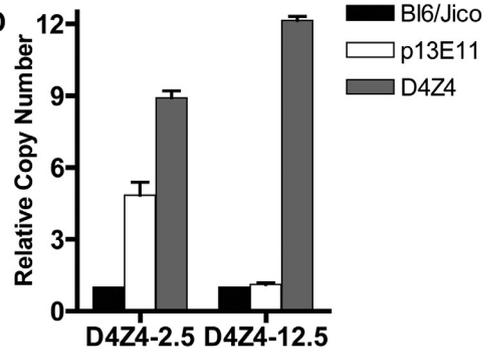
B



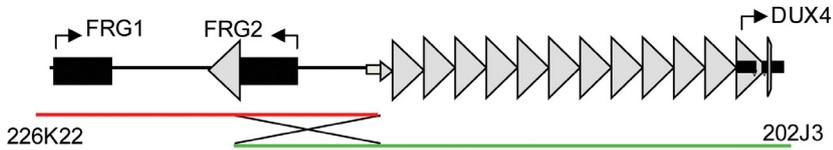
C



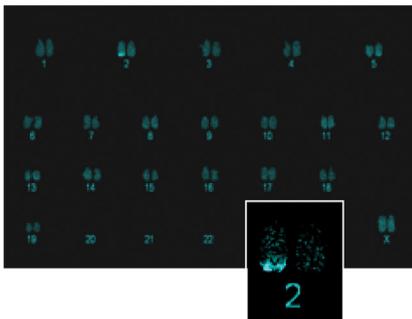
D



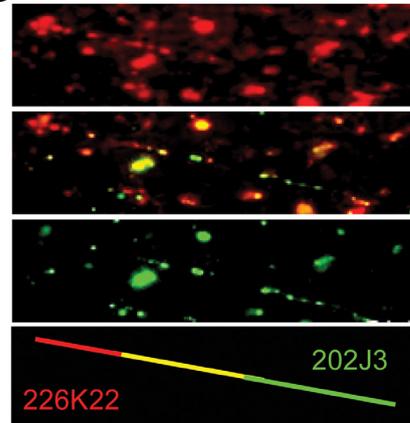
E



F



G



in situ hybridizations with specific 5' and 3' DUX4 probes revealed DUX4 transcripts in cells near the periphery of the seminiferous tubules (**Fig. S1**) consistent with spermatogonia and primary spermatocytes, as has been reported for normal human testis⁶. In D4Z4-2.5 mice, abundant DUX4 mRNA levels were also detected in ES cells and early developmental stages, which were chosen based on timing of key myogenic developmental waves, showing a gradual decline during development (**Fig. 2**)¹⁹⁻²². We also detected DUX4 mRNA in D4Z4-12.5 embryos, albeit at lower abundance, and in ES cells (**Fig. 2; Fig. S3**).

We selected a panel of somatic tissues, including several skeletal muscles which are typically affected in FSHD^{7, 23}. Reproducible levels of DUX4 mRNA were detected in all analyzed skeletal muscles of adult D4Z4-2.5 mice, including affected muscles of the limbs, trunk and head (**Fig. 3A; Fig. S2**). These levels were low and varied considerably between gender-matched littermates, as judged from semiquantitative analysis. Expression of DUX4 in non-muscle tissue could be expected as the candidate orthologue *Dux* is found to be expressed in cerebellum tissue¹. Indeed, DUX4 transcripts were detected in non-muscle tissues, including cerebellum but with the exception of liver, where only in one mouse DUX4 could be detected once (**Fig. 3B; Fig. S3**). In D4Z4-12.5 mice, DUX4 transcripts could only be reproducibly detected in the tibialis anterior and pectoralis muscles, whereas all other somatic tissues did not show reproducible DUX4 expression (**Fig. 2A-B; Fig S2; Fig. S3**). This suggests that, as seen in humans^{5, 9}, DUX4 is expressed variably in skeletal muscle of our transgenic mice and that decreased D4Z4 copy number contributes to inefficient DUX4 repression in somatic tissue, leading to a higher probability of expression.

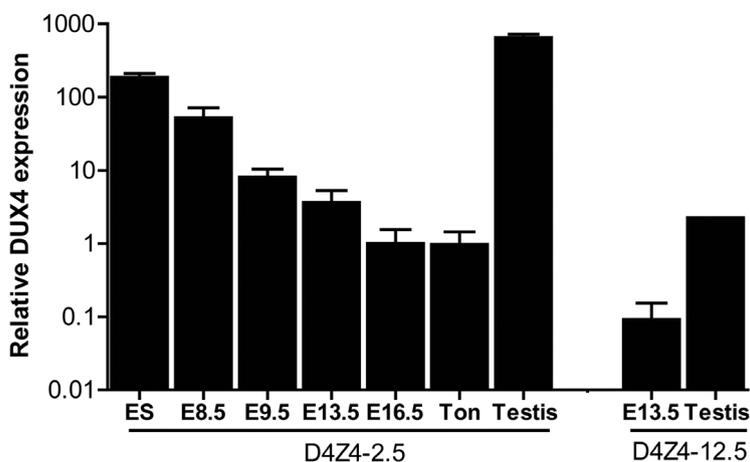


Figure 2. Quantitative expression analysis of DUX4 transcripts from the telomeric D4Z4 unit in D4Z4-12.5 and D4Z4-2.5 mice.

Quantitative RT-PCR data of DUX4 in D4Z4-2.5 embryonic stem cells (ES), in complete embryos of day E8.5, 9.5, 13.5 and 16.5, representing key myogenic developmental stages, and in adult ton=tongue, testis and in complete embryo day 13.5 and testis tissue of D4Z4-12.5 mice. Expression is normalized to the mouse reference gene *Hprt* and plotted in log₁₀ scale. Error bars indicate SEM of the mean (n=2-5).

In human FSHD muscle cell cultures, inefficient *DUX4* repression results in occasional nuclei expressing abundant amounts of *DUX4*. Therefore, we tested the expression of human *DUX4* in D4Z4-2.5 and D4Z4-12.5 satellite-cell-derived myoblasts, both by (quantitative) RT-PCR (**Fig. 3C-D**) and by immunofluorescent labeling (**Fig. 4**). In satellite-cell-derived myoblasts and differentiated myotube cultures obtained from

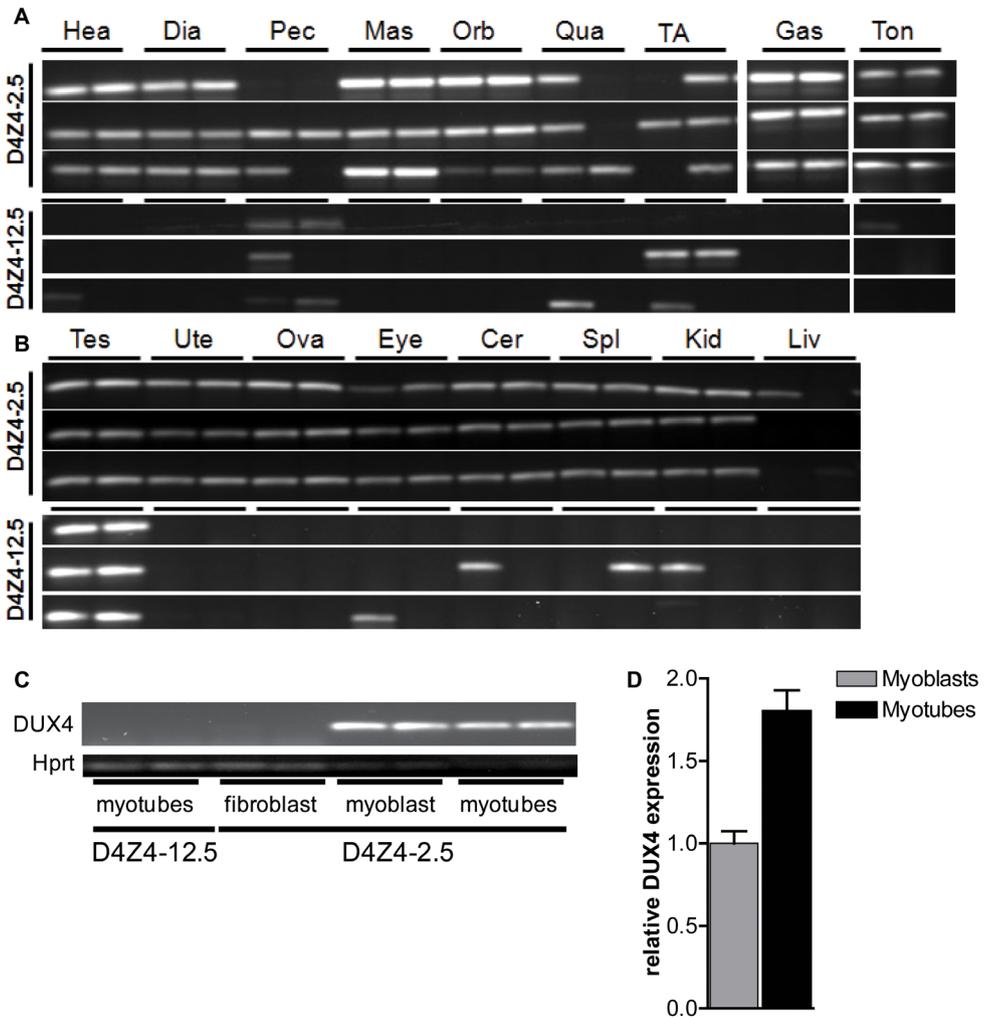
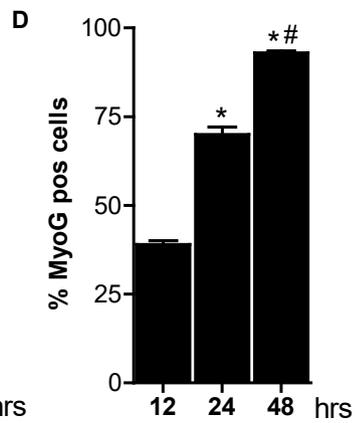
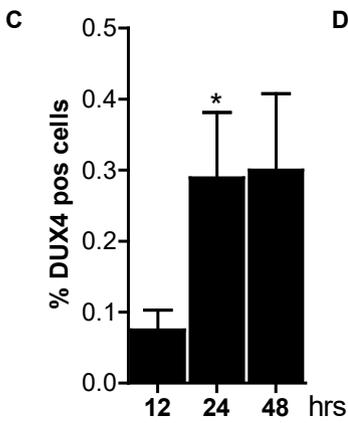
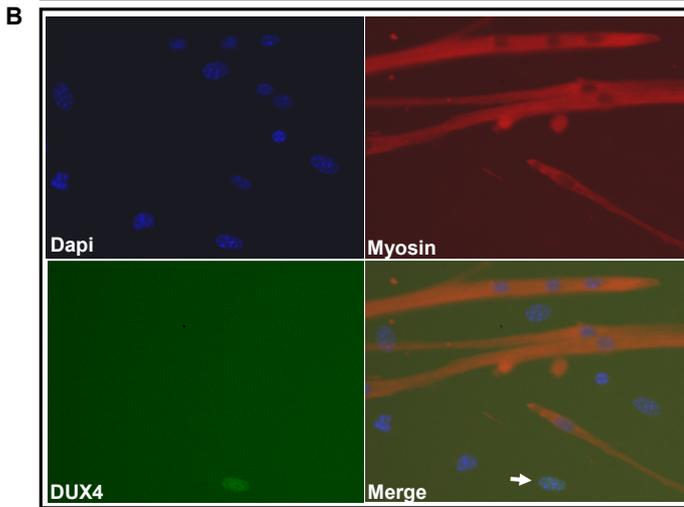
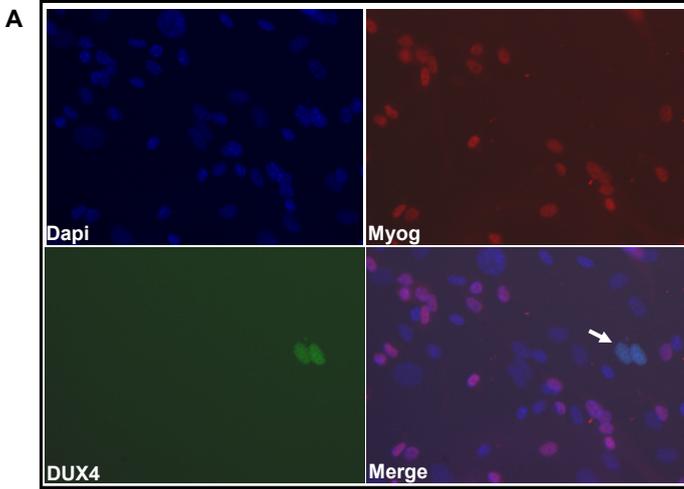


Figure 3: Analysis of transcriptional activity of *DUX4* in a panel of tissues of D4Z4-2.5 and D4Z4-12.5 mice

DUX4 transcripts measured in 7 weeks old D4Z4-2.5 and D4Z4-12.5 mice (n=3) **A**) muscle tissue: Hea = Heart, Dia = Diaphragm, Pec = Pectoralis Mas = Masseter, Orb = Orbicularis oris, Qua = Quadriceps, TA = Tibialis anterior, Gas = Gastrocnemius, Ton = Tongue and **B**) somatic non-muscle and germline tissue: Tes = Testis, Ute = Uterus, Ova = Ovarium, Eye, Cer = Cerebellum, Spl = Spleen, Kid = Kidney, Liv = Liver. **C**) *DUX4* transcripts measured in satellite-cell-derived myoblasts, myotubes and interstitial fibroblast extracted from EDL muscle of D4Z4-12.5 and D4Z4-2.5 transgenic mice. **D**) Quantitative RT-PCR data of *DUX4* expression in D4Z4-2.5 myoblasts (n=2) and myotubes (n=2) 48 hours after induction of differentiation. Errors indicate SEM of the plotted mean.



D4Z4-12.5 mice, DUX4 transcripts and DUX4 protein could not be detected (**Fig. 3C** and data not shown). In contrast, in satellite-cell-derived myoblasts of D4Z4-2.5 mice, DUX4 transcripts and sporadic DUX4-positive nuclei could be observed. Remarkably, all sporadic DUX4 positive nuclei were Myog negative and did not fuse into myotubes, as evidenced by co-staining with Myosin heavy chain (**Fig. 4A-B**). These data suggest DUX4-mediated inhibition of myoblast differentiation, as has been shown previously in zebrafish²⁴. Total expression levels and frequency of DUX4-positive nuclei increased by 2-4 fold upon differentiation into myotubes (**Fig. 3D and 4D**). Satellite-cell-derived myoblasts obtained from fast-twitch (EDL) and slow-twitch (soleus) fibers from D4Z4-2.5 mice showed the same DUX4 protein expression pattern (data not shown). Of interest, interstitial fibroblasts obtained from collagenase digested D4Z4-2.5 EDL and soleus muscle did not express DUX4 transcripts (**Fig. 3C**), indicating that the expression of DUX4 in the D4Z4-2.5 EDL and soleus is muscle cell specific.

Altogether, these data show that both the RNA and protein expression pattern of DUX4 in our D4Z4-2.5 mouse model recapitulates several features of FSHD. *DUX4* is more efficiently silenced with increasing D4Z4 copy number in our two mouse models, thereby forming a paradigm for the difference between human FSHD1 patients and healthy controls.

Chromatin structure of D4Z4 in transgenic mice

Next, we studied the chromatin structure of the integrated D4Z4 repeats in order to determine whether the observed DUX4 expression patterns correlate with epigenetic differences in the FSHD and control transgenic loci. At control alleles in humans, D4Z4 is characterized by high CpG methylation levels and the co-occurrence of histone 3 lysine 9 trimethylation (H3K9me3) and histone 3 lysine 4 dimethylation (H3K4me2). FSHD alleles are epigenetically characterized by reduced D4Z4 CpG methylation and a reduced H3K9me3:H3K4me2 ratio, referred to as the chromatin compaction score (ChCS)^{12, 25-28}. We assessed DNA methylation and the ChCS in the two mouse lines at indicated sites within the transgenic loci (**Fig. 5A**). DNA methylation analysis using methylation-sensitive restriction enzymes followed by Southern blotting (representative blot shown in **Fig. 5B**) showed that both the proximal and internal D4Z4 units of the array were highly methylated (60-90%) in gastrocnemius muscle of D4Z4-12.5 mice (**Fig. 5C**), similar to unaffected individuals. In gastrocnemius muscle of D4Z4-2.5 mice, the D4Z4 units were

Figure 4. Bursts of DUX4 protein expression in differentiating D4Z4-2.5 muscle cells.

Satellite-cell-derived myoblasts extracted from single EDL fibers of D4Z4-2.5 mice were differentiated for 12, 24 and 48 hrs and co-stained for DUX4 and Myog or for DUX4 and Myosin heavy chain. **A)** Representative DUX4 and Myog IF staining images of D4Z4-2.5 myotubes, 24hrs after induction of differentiation, indicate absence of Myog in DUX4 expressing cells. **B)** Representative DUX4 and Myosin HC IF staining images of D4Z4-2.5 myotubes, 24hrs after induction of differentiation, indicate exclusion of DUX4 positive cells from newly formed myotubes. Both DUX4 (panel C) and Myog (panel D) positive nuclei in relation to total amount of nuclei (DAPI staining) were counted during the differentiation process. **C)** Approximately 2:1000 nuclei showed nuclear DUX4 staining. **D)** The percentage of Myog positive nuclei revealed an increase in differentiation committed cells with time. After 48 hours of differentiation almost all myoblasts are committed to differentiation. Error bars indicate stdev of the plotted mean (n=7); **P* < 0.05 compared to t=12 hrs; #*P* < 0.05 compared to t=24 hrs.

relatively hypomethylated (10-20%; $P < 3.10^{-5}$) (Fig. 5C), similar to FSHD patients and in concordance with the observed difference in DUX4 expression between the two mouse lines (Fig. 3A). Similar results were obtained for quadriceps, heart, brain and liver of both transgenic lines (data not shown). Bisulphite DNA methylation analysis of different regions within D4Z4 in embryonic and adult tissues showed that D4Z4 hypomethylation in D4Z4-2.5 mice is indeed stable and uniform between tissues (Fig. S4). Similar analysis in 10 month old mice showed that this epigenetic signature is stable with age (data not shown). Chromatin immunoprecipitation (ChIP) analyses in mouse fibroblasts and

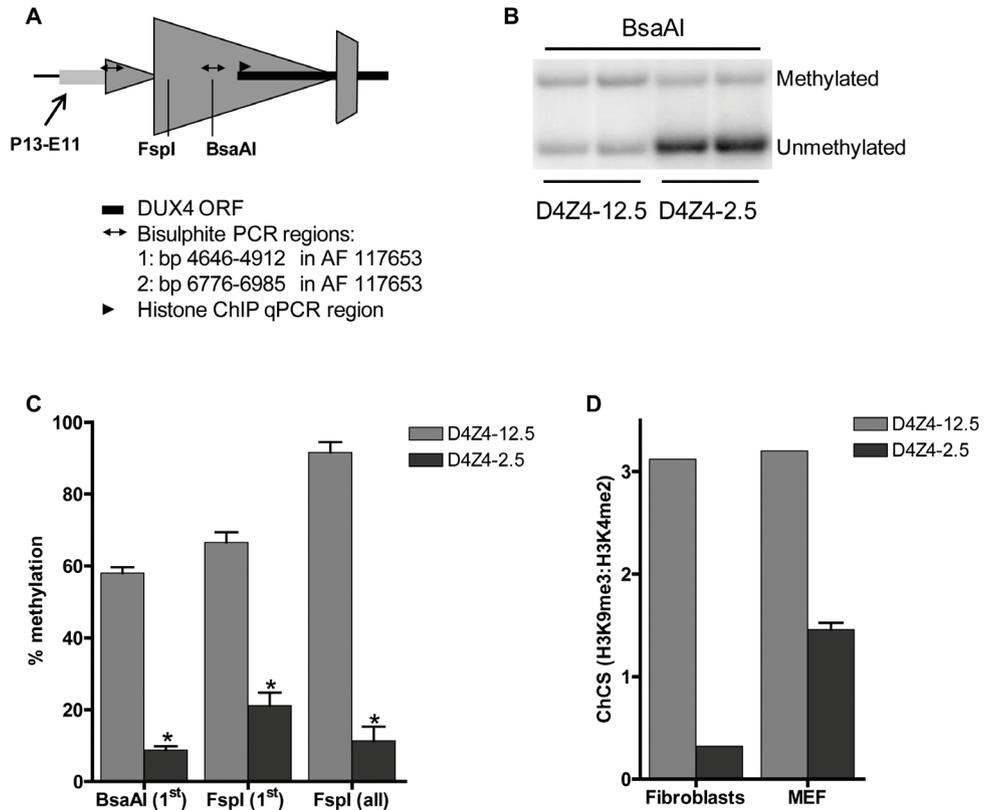


Figure 5. Epigenetic structure of D4Z4 in D4Z4-2.5 and D4Z4-12.5 mice.

A) Schematic draw of the regions within D4Z4 where CpG and histone methylation were interrogated. **B)** Representative figure of a methylation sensitive Southern blot assay to quantify DNA methylation levels. Upon BsaAI digestion, gel separation and blotting, two distinct bands representing the unmethylated and methylated fragment are visualized and quantified. **C)** Southern Blot analysis was done using two different methylation sensitive restriction enzymes, BsaAI and FspI, in adult gastrocnemius muscle tissue of D4Z4-12.5 and D4Z4-2.5 mice. Both probes p13E-11 and D4Z4 were used to measure CpG methylation levels in the most proximal unit and all units, respectively. The methylation percentages of the two different CpG sites are plotted. Error bars indicate stdev of the plotted mean ($n=4$ D4Z4-12.5 vs $n=5$ D4Z4-2.5, $*P < 0.001$). **D)** Histone methylation levels of D4Z4 in transgenic D4Z4-12.5 and D4Z4-2.5 embryonic (MEFs) and adult fibroblasts. Chromatin was precipitated with H3K4me2, H3K9me3 and control IgG antibodies. Precipitated DNA was amplified with qPCR primers amplifying the transcription start site of DUX4. Levels of H3K9me3 in relation to H3K4me2 have been plotted as the chromatin compaction score (ChCS).

myoblasts showed a relative decrease in ChCS in D4Z4-2.5 compared to D4Z4-12.5 mice (**Fig. 5D** and data not shown), similar to what was seen in patient derived human skin fibroblasts and primary myoblasts when compared to control subjects²⁸. Taken together, CHIP and DNA methylation analyses indicate a relative chromatin relaxation of the D4Z4 repeats in D4Z4-2.5 mice compared to D4Z4-12.5 mice, concordant with the observed DUX4 expression pattern, thereby accurately modeling the difference between FSHD patients and control individuals.

Genic consequence of DUX4 expression in mouse muscle cells

In human muscle cells, DUX4 activates germline and early stem cell programs while suppressing several genes involved in the innate immune response⁵. To assess the

Table 1: deregulated genes in response to DUX4 expression in C2C12, linked to functional groups shown to be deregulated in human myoblasts upon DUX4 expression.

Germline & early development	logFc	FDR
Gm397 (Zscan4c)	1.74	1.07E-12
Trim36	1.75	1.44E-13
Pvrl3	1.15	2.01E-10
Psme4	1.08	1.65E-06
Lmo4	1.96	5.88E-12
Id1	-1.21	1.08E-09
Id3	-1.17	2.33E-09
Immune response	logFc	FDR
Wfdc3	4.44	1.00E-19
Ankrd1	-2.06	1.53E-11
Ccl7	-1.86	1.61E-11
Sema7a	-1.61	5.11E-10
Sema4b	-1.22	1.03E-08
Cxcl1	-1.56	3.55E-13
Irf1	-1.47	1.67E-10
Socs3	-1.14	3.37E-08
PRAMEF orthologues	logFc	FDR
Gm13040 *	2.33	1.54E-14
Gm13043 *	2.31	3.55E-13
BC080695 *	2.03	5.87E-13
Gm16367	1.12	3.06E-09
Gm13119	2.21	7.84E-17

* Genes in cluster, possible duplications: >97% sequence homology.

consequence of ectopic DUX4 expression on global gene expression in mouse muscle cells, proliferating C2C12 cells were transiently transfected with a DUX4 expression vector (pCS2-DUX4) or, as a control, the empty pCS2 control vector⁵. After 24 hours, DUX4 expressing cells were enriched by FACS sorting and global gene expression changes were mapped by performing array analysis. We identified 183 significantly deregulated genes (2-fold change and FDR<0.05), of which 142 genes showed up-regulation and 41 genes showed down-regulation (**table S1**). GO pathway analysis was hampered by the small number of deregulated genes, but manual inspection of the gene list revealed a considerable overlap with the deregulated genes in human myoblasts expressing DUX4⁵. Out of the 183 genes regulated by DUX4 in mouse C2C12 cells, 43 (23%) were previously determined to be regulated by DUX4 in human muscle cells, of which 39 changed in the same direction (**table S2**).

As in human cells, DUX4 regulated a number of germline-specific, early stem cell and innate immune response genes in the mouse C2C12 cells (**table 1**). Mouse orthologs of human genes regulated by DUX4, such as *Zscan4c* and at least three orthologs of the *PRAMEF* gene family (although poorly annotated in mice) showed transcriptional activation in the presence of DUX4 (**table 1; table S1; Fig. S5**). Immune modulation by DUX4 in human muscle involves a number of innate immunity related genes. Also in C2C12 cells, DUX4 regulates at least eight genes related to the innate immune response, for example *Wfdc3*, encoding a secreted peptide proposed to have antimicrobial activity^{5, 29}. Using quantitative RT-PCR, we validated expression levels of DUX4 and a selection DUX4 regulated genes in DUX4 transfected C2C12 muscle cells (**Fig. 6A-D**). We confirmed deregulation of genes which were switched on by DUX4 (panel A), genes which are deregulated in both human and mouse (panel B) and genes implicated in germ cell biology and early development (panel C). Deregulated genes linked to innate immunity (panel D) showed activation upon transfection, which was dampened in the presence of DUX4, as was shown in similar experiments with human myoblasts⁵. Altogether, ectopic expression of DUX4 in C2C12 cells results in deregulation of a gene set which shows overlap with DUX4 responsive genes and gene sets in human myoblasts.

DUX4 can act as a transcriptional activator in C2C12

To identify genes in the mouse genome that are directly regulated by DUX4, C2C12 myoblasts were transiently transfected with pCS2-DUX4 and subjected to ChIP-seq analysis. We identified a total of 2784 peaks (p-value < 10⁻⁵, FDR 0.02) and identified a DUX4 consensus binding sequence, highly similar to what was found in human muscle cells (**Fig. S6A** and previously described⁵), with a strong conservation of the core TAAYYAATCA double homeobox binding motif (**Fig. S6A**). The genomic distribution of the identified peaks showed a slight bias for promoter sequences, as is seen for transcription factors, but not for DUX4 in human myoblasts (**Fig. S7**)⁵. Next we identified DUX4 regulated genes (log₂FC > [.58], FDR < 0.05) that have a DUX4 binding site within a CTCF insulator domain surrounding their transcriptional start site (TSS). In this way, 91 potential direct DUX4 target genes were identified, of which 10 genes showed DUX4 enrichment within a -2 to +2 kb window from their TSS (**table S3**). To validate the direct effect of DUX4, we selected 4 of these 10 direct target genes and validated

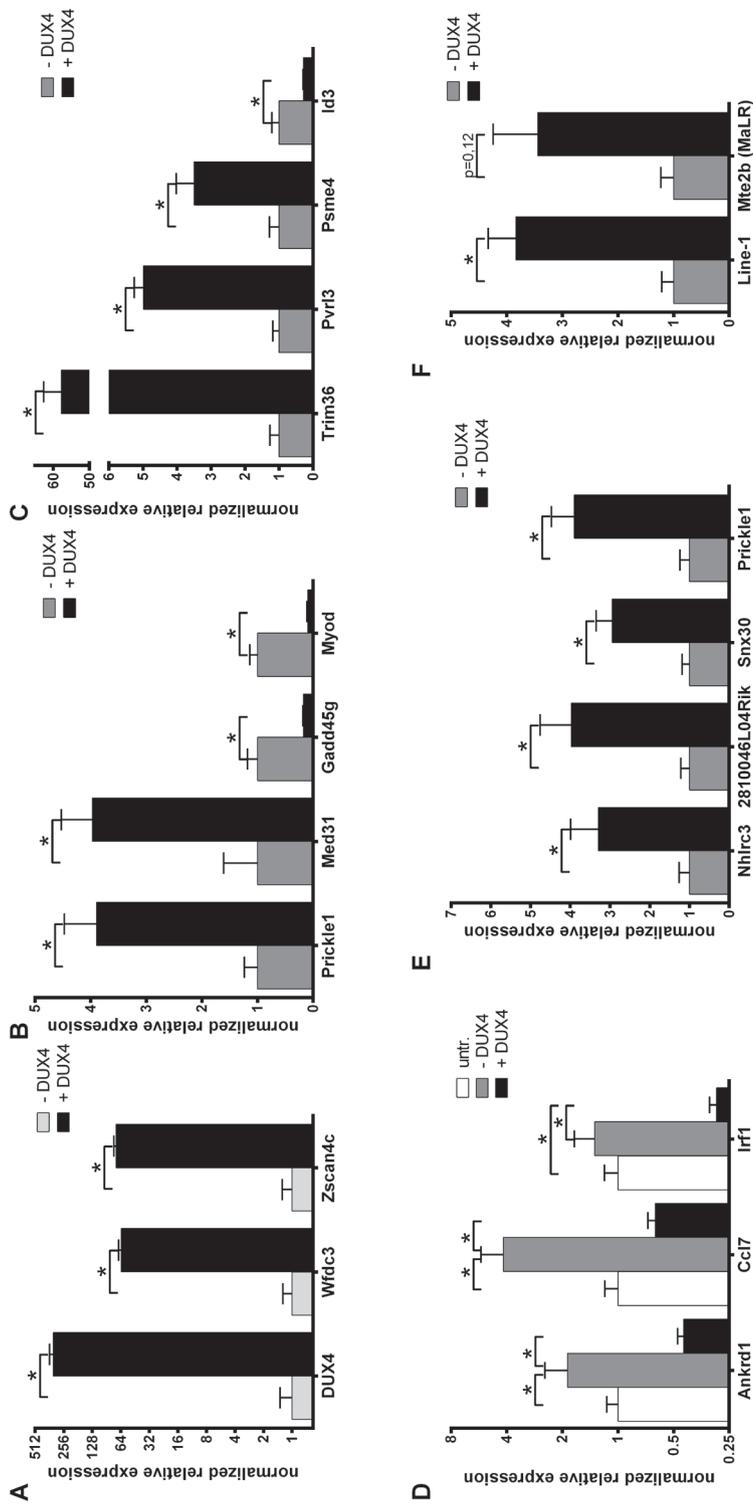


Figure 6. Validation of expression levels of DUX4 deregulated genes in C2C12 myoblasts.

A set of deregulated genes obtained from expression array analysis was confirmed by qRT-PCR. Expression analysis of **A**) DUX4 and genes that are switched on by DUX4 in C2C12 cells, **B**) genes that respond to DUX4 in humans and mice, **C**) germ line and early development associated genes, **D**) innate immunity genes, **E**) genes directly regulated by DUX4 expression, **F**) genes directly regulated by DUX4 which were identified by ChIP-seq and **F**) activated L1 and MaLR retrotransposons. For panel A and Mte2b in panel F, DUX4- values refer to the DUX4 depleted FACS sorted fraction, enabling proper normalization of genes switched on upon DUX4 expression. In all other panels DUX4- refers to pCS2 transfected cells. All expression levels are relative to Cyclophilin-B and normalized to DUX4- or wt conditions. Error bars indicate SEM of at least triplicate measurements, asterisks indicate P -values < 0.05 based on a student's t-test (panels A, B, C, E & F) or one way ANOVA (panel D) analysis.

their expression levels by qRT-PCR (**Fig. 6E**). *Nhlrc3* and *2810046L04Rik* are adjacent and transcribed in opposite direction, indicating DUX4 might enhance expression from both promoters. To confirm that DUX4 indeed functions as a transcriptional activator at these sites, the DUX4 consensus binding site found at *2810046L04Rik* and *Nhlrc3* was cloned in both orientations upstream of a luciferase reporter gene and it significantly induced expression of the reporter construct when co-transfected with DUX4 (**Fig. S8**). Taken together, we find evidence that DUX4 can act as a transcriptional activator in mouse muscle cells as it does in human muscle cells. In addition, we identify direct DUX4 targets that might serve as suitable biomarkers in our D4Z4-2.5 mouse model of FSHD.

Non-genic consequence of DUX4 expression in mouse muscle cells

In human cells, DUX4 has been shown to bind and activate retrotransposons, mainly of the MaLR type⁵. Our ChIP-seq analysis showed that DUX4 binds to several different types of retrotransposons also in the mouse genome. Both uniquely mappable (**table S4**) and non-unique sequence reads (**table S5**) show enrichment for DUX4 binding at LTR, LINE and, to a lesser extent, SINE retroelements. Quantitative RT-PCR analysis supports an upregulation of transcripts emanating from LINE-L1 and MaLR (*Mte2b*) retrotransposons in the presence of DUX4 (**Fig. 6F**). Like in humans, the DUX4 core TAAYYYAATCA binding motif is present in each repetitive element, although flanking nucleotides show repeat specific differences (**Fig. S6B-C**)⁵. We conclude that under these experimental conditions DUX4 binds and transcriptionally activates repetitive elements in the murine genome, similar to its activity in the human genome.

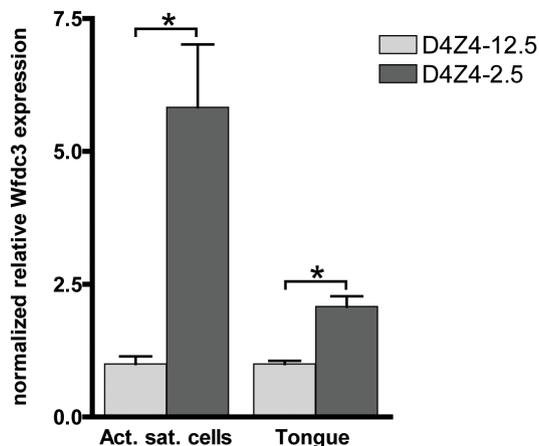


Figure 7. Expression of the DUX4 induced *Wfdc3* gene in myoblasts and tongue muscle of D4Z4-2.5 mice.

Relative expression of *Wfdc3* in satellite-cell-derived myoblast cultures from single EDL fibers (D4Z4-2.5: n=6 and D4Z4-12.5: n=5) and tongue tissue isolated from 7-8 weeks old mice (D4Z4-2.5: n=6 and D4Z4-12.5: n=6). Expression levels are relative to Cyclophilin-B and *Hprt* and normalized to levels in D4Z4-12.5 mice, plotted as the mean \pm SEM. Asterisks indicate $P < 0.05$ according to an independent two-tailed student's t-test.

DUX4 responsive genes in D4Z4-2.5 derived muscle cell cultures, adult skeletal muscle and embryos

Since DUX4 expression levels in adult D4Z4-2.5 skeletal muscle are low and only a subset of myonuclei show DUX4 protein immunoreactivity in the D4Z4-2.5 satellite-cell-derived myoblast cultures, we examined whether we could observe changes in the expression levels of DUX4 responsive genes identified by ectopic expression of DUX4 in C2C12 cells. In a set of five D4Z4-12.5 and six D4Z4-2.5 muscle cell cultures we found significant deregulation of *Wfdc3* (**Fig. 7**). In tongue muscle of D4Z4-2.5 mice, where we observed relative robust DUX4 transcript levels, again *Wfdc3* levels were found to be significantly increased in D4Z4-2.5 adult mice as compared to D4Z4-12.5 adult mice (**Fig. 7**). Since we observed relatively robust levels of DUX4 during embryogenesis, we also investigated whether we could observe the activation of *Wfdc3* during this stage. In D4Z4-2.5 embryos at day E9.5 we observed robust expression of DUX4 (**Fig. 3**) and concordantly, *Wfdc3* mRNA levels increased by 2-fold compared to wildtype (WT) controls ($P = 0.002$) (**Fig. S9**). Together these data showed that despite the low and variable levels of DUX4 itself, *Wfdc3* showed reproducible upregulation in D4Z4-2.5 cells and tissues, which makes it a suitable candidate for serving as a biomarker of DUX4 activity.

Phenotype of D4Z4-2.5 mice

The phenotype of FSHD patients shows high inter- and intrafamilial variation. Next to the progressive muscle weakening, hearing loss and retinopathy are frequently observed extramuscular features³⁰. The most obvious phenotype in our D4Z4-2.5 mice is the development of eye abnormalities in approximately 54% of the D4Z4-2.5 mice with an onset of around 8-12 weeks of age (**Fig. S10**). Although variable, the mice develop a progressive keratitis of unknown etiology, possibly reflecting incomplete eyelid closure or another yet to be determined cause.

The overall morphology and histology of the limb and some head muscles appeared normal. Inducing mild muscle damage by down-hill running (eccentric activation) did not induce measurable muscle weakness and damage in D4Z4-2.5 mice. Grip strength, creatine kinase levels, Evan's blue dye uptake and proportion of central nuclei were all similar between WT and D4Z4-2.5 mice (**table S6**). In addition, expression of myogenic and immunogenic markers shown to be deregulated in many mouse models for muscular dystrophies³¹, were not changed in muscle of the D4Z4-2.5 mice (**table S6**). Inducing muscle regeneration by cardiotoxin injection, also did not cause significant differences with respect to regeneration capacity between WT and D4Z4-2.5 mice. At 10 and 28 days after treatment, we examined the percentage of central nuclei, distribution of fiber sizes and fibrotic tissue formation. Only in the formation of fibrotic tissue, a trend towards delayed regeneration was observed, as in D4Z4-2.5 a slight increase was observed at day 28 (**table S6** and data not shown).

Discussion

DUX4 has been implicated in FSHD pathology based on its inappropriate expression in skeletal muscle of patients with FSHD^{6, 16, 17}. *DUX4* encodes a double homeobox transcription factor which activates germline, early stem cell and other programs in FSHD muscle, eventually leading to cell death^{5, 32, 33}. FSHD has a complex etiology: insufficient epigenetic silencing caused by D4Z4 contraction in FSHD1 or heterozygous mutations in the chromatin modifier SMCHD1 in FSHD2 patients results in the inappropriate expression of the retrotransposed *DUX4* gene in skeletal muscle^{8, 9, 14, 25, 27, 34}.

In this study, we have established mouse models that recapitulate several important aspects of FSHD and control individuals with respect to the aforementioned genetic and epigenetic features of the D4Z4 macrosatellite repeat array, encoding the *DUX4* gene. Like in FSHD and control individuals, *DUX4* is expressed in the germline of D4Z4-2.5 and D4Z4-12.5 mice and, like in FSHD, D4Z4-2.5 mice show low and variable *DUX4* expression levels in somatic tissue. In FSHD patients, there is little information about somatic expression of *DUX4* in non-muscle tissue, but our studies in mice suggest that derepression of *DUX4* is not limited to skeletal muscle, consistent with the observed body-wide hypomethylation of D4Z4. In our D4Z4-12.5 mouse model we observe more efficient somatic repression of the *DUX4* locus, where *DUX4* can only be reproducibly detected in pectoralis and tibialis anterior muscle, typically affected in FSHD. Excitingly, the D4Z4-2.5 mouse model also reproduces the characteristic variegated expression pattern of *DUX4* protein in FSHD muscle cell cultures: only a small sub-population of myonuclei express abundant levels of the *DUX4* protein.

Thus far, several animal models over-expressing FSHD candidate genes have been produced. Some of them focused on the proximally located FSHD candidate gene *FRG1*³⁵. Transgenic mice and *Xenopus laevis*, over-expressing *FRG1* either muscle-specifically or systemically, both demonstrated an abnormal musculature³⁶⁻³⁸. Nevertheless high over-expression of *FRG1* does not reflect the human FSHD expression profile and *FRG1* upregulation in FSHD muscle remains controversial³⁹⁻⁴². *DUX4* over-expressing models have revealed the robust toxicity of *DUX4* in somatic tissue, as demonstrated by massive cellular loss and abnormal development^{32, 33, 37, 43-45}. In muscle cells, *DUX4* has been shown to cause cell death and renders myoblasts hypersensitive to oxidative stress^{32, 33, 43}. While these models provide insight into the potential harmful effect of *DUX4* in muscle or organismal development, none of them take into account the specific endogenous expression pattern of *DUX4* and its related effect.

The mouse models presented here, both carry human genomic constructs with the FSHD permissive subtelomeric region necessary for somatic *DUX4* expression. Our data strongly suggests that the transcriptional profile of the *DUX4* retrogene seems to be maintained and follows the pattern observed in patients and controls. While we can find reproducible evidence of somatic derepression of *DUX4* in adult D4Z4-2.5 mice, including the variegated pattern of *DUX4* positive myonuclei in cell culture, D4Z4-12.5 mice show more efficient repression of *DUX4* in their somatic tissues. This pattern of *DUX4* expression suggests a locus-intrinsic property of the D4Z4 repeat array, of which

the regulation is conserved between mouse and human muscle. However, we cannot rule out that some differences between the two transgenic mouse lines in the expression of *DUX4* are the consequence of different chromatin regulation at the sites of integration. We should exert some extra caution, since we only were able to generate one founder line for D4Z4-2.5 after over 450 attempts. Nonetheless, the D4Z4-2.5 mouse model is the first organismal model that can provide more insights in *DUX4* regulation *in vivo* during development; e.g., why, how and when are sudden bursts of *DUX4* expression in skeletal muscle regulated. This is particularly relevant since primates have lost the parental gene that retrotransposed to create the *DUX4* retrogene¹⁸. As it was recently demonstrated that the detrimental effects of *DUX4* expression in mouse muscle can be reversed by RNA interference⁴⁶, our model may also serve well therapeutic intervention studies targeting *DUX4* expression in skeletal muscle.

The FSHD specific somatic derepression of *DUX4* has been associated with changes in D4Z4 chromatin structure, characterized by decreased DNA methylation and a lower ChCS^{12, 27, 28}. D4Z4-12.5 mice, containing a normal sized D4Z4 repeat array of 12.5 units, show a heterochromatic D4Z4 repeat structure comparable to human control subjects. D4Z4-2.5 mice, containing an FSHD1 sized D4Z4 array of two-and-a-half units, show a more open D4Z4 chromatin structure similar to FSHD patients. Although we cannot rule out integration-site specific effects, the CpG methylation and histone marker data of the D4Z4 arrays in both mouse lines support the model that the chromatin status of this locus is determined by a repeat length-dependent mechanism and that somatic *DUX4* repression is enforced through an evolutionary conserved mechanism of repeat-mediated silencing. The most proximal D4Z4 unit displayed lower CpG methylation than the internal units (**Figs. 2B, D, E vs. 2F, G**), a phenomenon also observed for the D4Z4 array in humans^{25, 47}. In addition, we clearly demonstrated differences in DNA methylation between different CpG dinucleotides within D4Z4. Thus, even though D4Z4 is integrated at different sites within a mouse genomic background, its “human” epigenetic profile seems to be preserved. It will be of interest to study the conservation of the recently described involvement of the Polycomb/Trithorax complex in the regulation of D4Z4 and the transcripts emanating from it⁴⁸. Our mouse model could serve as a suitable model to further study the specific epigenetic regulation as analyses are not hampered by the presence of other homologous repeat arrays or degenerate copies as is the case in human samples.

DUX4 can act as a transcription factor causing the deregulation of specific gene programs when ectopically expressed in human cells^{2, 5, 14}. To better understand if the *DUX4* protein can exert similar functions in the context of the mouse genome, we analyzed its effects in C2C12 cells by combining ChIP-seq with transcriptome data sets. Our ChIP-seq data showed that the core motif bound by *DUX4* in human was highly conserved in the mouse, although the relative abundance of binding sites was relatively enriched for promoters in the mouse dataset⁵. In addition to regulating specific gene sets, *DUX4* also binds and activates several classes of retrotransposons in the mouse genome, including Line-1 and Mte-2b, the latter one closely related to human MaLR LTRs which are specific *DUX4* targets in the human genome⁵. This indicates that the

primate specific retrogene *DUX4* can – at least in part – elicit a similar transcriptional response at retrotransposons in mice and raises the question whether DUX-related transcription factors in the germline are involved in retrogene biology.

Our data also showed that ectopic *DUX4* expression has some similar genic consequences in mouse cells as it has in human cells. However, in contrast to humans, where *DUX4* alters the expression of a large number of genes, in mice *DUX4* only affects the expression levels of 183 genes, possibly reflecting the primate specificity of *DUX4*. This limited number of genes precluded pathway analysis, but we observed that ~25% of genes deregulated in the mouse genome are also deregulated by *DUX4* in humans, including genes involved in early development, germline biology and innate immunity. In addition, we identified a number of genes which are not shared between the mouse and human dataset, but fall in one of the aforementioned categories. Over 40% of the genes we identified to be regulated by *DUX4* were also identified in the inducible *DUX4* overexpression study done by Bosnakovski et al.³². Combining our ChIP-seq analysis with the transcriptome analysis allowed direct *DUX4* target identification and revealed that *DUX4* can act as a transcriptional activator in mouse as it does in human cells.

We observed that some genes, including the early development and germline genes, become activated in mouse C2C12 muscle cells where normally these genes are not expressed. These genes may serve as good biomarkers in our mouse models for future studies tailored towards therapeutic effects of *DUX4* downregulation. When studied in tissues and cell cultures isolated from our mouse models, indeed *Wfdc3* was shown to be significantly increased in D4Z4-2.5 mice compared to the D4Z4-12.5 mice. Although only a limited number of nuclei show expression of *DUX4*, we see robust upregulation of *Wfdc3* transcription, which is also seen for the activation of target genes in human FSHD samples. The nature of this observation remains elusive, however the structure of the *DUX4* transcript makes it a likely target for nonsense mediated decay, whereas its target genes are generally not.

Although we were not able to document an obvious skeletal muscle phenotype in D4Z4-2.5 mice expressing low and variable levels of *DUX4* in their muscles, we did notice a trend towards muscle weakness. The EDL muscle consists of somewhat smaller fibers and after induction of severe muscle damage by cardiotoxin injection, D4Z4-2.5 mice show a small delay in muscle regeneration. To improve therapeutic readout in our D4Z4-2.5 mouse, it will be interesting to assess *DUX4* expression and muscle regeneration after multiple rounds of muscle damage, for example by crossbreeding D4Z4-2.5 with *mdx* mice.

Interestingly, over time more than half of the D4Z4-2.5 mice develop an abnormal eye phenotype eventually leading to blindness. Weakness of the eyelid muscles (orbicularis oculis) and thereby the difficulty in closing eyes is characteristic for FSHD. Moreover, 60% of FSHD patients also develop retinal telangiectasis, which can even lead to retinal detachment, known as Coat's syndrome⁴⁹. Therefore it will be imperative to assess the eye pathology in our D4Z4-2.5 mouse in more detail.

In conclusion, we here report on the first transgenic mouse models which accurately model the epigenetic regulation of normal-sized and FSHD-sized D4Z4 macrosatellite repeats. While D4Z4-2.5 mice show strong overlap with the molecular phenotype of FSHD, D4Z4-12.5 mice more reflect D4Z4 regulation observed in control individuals. These mouse models will facilitate studies focusing on the *in vivo* regulation of *DUX4* and the consequences of somatic derepression of this germline transcription factor. These mouse models can also be utilized to evaluate and optimize future therapeutic strategies for FSHD.

Material & Methods

Ethics statement

All animal experiments were approved by the local animal experimental committee of the Leiden University Medical Center and by the Commission Biotechnology in Animals of the Dutch Ministry of Agriculture.

Generation & maintaining Transgenic mice

D4Z4-2.5 mice were generated by microinjection of the λ 42 (L42) phage into pronuclei of fertilized oocytes of B6CBAF1/J mice (Charles River Laboratories, Wilmington MA, USA). The λ 42 phage contains a 13.5 kb EcoRI fragment encompassing the partially deleted D4Z4 locus of a patient with FSHD⁵⁰. D4Z4-12.5 mice were generated by co-injection of the 226K22 (accession number AF146191) and 202J3 PAC clones into pronuclei of fertilized oocytes of B6CBAF1/J mice (Charles River Laboratories). PAC clone 226K22 was derived from chromosome 4A161 and isolated as described before⁵¹. The 202J3 PAC contained a genomic fragment extending from the D4S2463 into the D4Z4 repeat, including the permissive 4A161 poly-adenylation site, but not including exon 6 and 7, and has been isolated from the RPCI-6 library (detailed on Roswell park cancer institute, Buffalo NY, USA). The transgenic mice were genotyped by PCR analysis on tail DNA. Presence of permissive haplotype containing intact polyA site distal to last D4Z4 units has been assessed by PCR followed by Sanger sequencing of the pLAM region (LGTC, Leiden, Netherlands). Primers are listed in **table S7**. The transgenic mice used in these experiments were all back-crossed with C57bl6Jico mice for at least 20 generations and were bred at the animal facility of the LUMC. Mice were housed in individually ventilated cages with 12-h light–dark cycles. Standard mouse chow and water was given *ad libitum*.

COBRA & Fiber FISH

Metaphase spread FISH analysis was done as described before¹⁶. COBRA-FISH was performed on D4Z4-2.5 and D4Z4-12.5 fibroblasts essentially according to the method of Szuhai et al.⁵². In short, whole chromosomal painting probes (Cytocell, Adderbury, Banbury, UK) for COBRA were labeled with diethylaminocoumarin (DEAC)-, Cy3- and Cy5-ULS [reagents included in the Universal Linkage System (Kreatech Biotechnology, Amsterdam, The Netherlands), used as ratio-fluorochromes], and dGreen-ULS (used as binary fluorochrome) and combined with either biotin-16-dUTP labeled D4Z4-2.5 or 226K22 or digoxigenin-11-dUTP (Dig) labeled 202J3 (Roche Applied Science, Mannheim, Germany) by nick translation. Streptavidin-LaserPro and mouse-anti

digoxigenin were applied to detect biotin and digoxigenin labeled probes, respectively. Recorded images were further processed and analyzed with an in house developed software tool (ColourProc)⁵².

Fiber-FISH was done on both cultured spleen and fibroblast cells isolated from the D4Z4-2.5 and D4Z4-12.5 transgenic mice. The cells were first attached to aminosilane-coated microscope slides and then lysed to produce linear DNA fibers, which were fixed to the slide with methanol-acetic acid (3:1). The PAC clones 226K22 and 202J3 were labeled with biotin and digoxigenin respectively. Both were hybridized simultaneously to the fiber preparations and immunocytochemically detected using Alexa Fluor 594 and fluorescein iso-thiocyanate (FITC), respectively. The hybridizations were analyzed using a fluorescence microscope. At least 40 fibers were assessed to obtain the order and the number of the red and green signals.

MLPA

Probes were designed against the transgene and wild type alleles (**table S7**), containing all the criteria as described in White et al⁵³. Reagents for the MLPA reaction and subsequent PCR amplification were purchased from MRC-Holland (Amsterdam, The Netherlands). The MLPA reactions were performed essentially as described in Schouten et al.⁵⁴. Briefly, 200 ng of genomic DNA (concentration determined using a UV spectrophotometer) in a final volume of 5 μ l was heated at 98°C for 5 minutes. After cooling to room temperature, 1.5 μ l probe mix and 1.5 μ l SALSA hybridization buffer was added to each sample. Next, the samples were denatured at 95°C for 1 minute and hybridized for 18 hrs at 60°C. Ligation was performed at 54°C for 15 minutes by adding 25 μ l water, 3 μ l buffer A, 3 μ l buffer B and 1 μ l ligase. The reaction was stopped by heat inactivation at 98°C for 5 minutes. PCR amplification was carried out for 30–33 cycles in a final volume of 25 μ l. The MLPA primers were labeled with FAM and added with a final concentration of 200nM. From each PCR reaction, 1.5 μ l of product was mixed with 10 μ l (Hi Di) formamide and 0.05 μ l ROX500 size standard in a 96 well plate. Product separation was performed using capillary electrophoresis on the ABI 3700 (Applied Biosystems/Life technologies, Bleiswijk, The Netherlands). To obtain a ratio for each product, the peak height was divided by the sum of the peak heights of the wild type probes.

Isolation single muscle fibers and culturing of the mouse myoblasts

Mice were sacrificed by cervical dislocation and the EDL and soleus muscles were carefully dissected from tendon to tendon and digested in 0.2% collagenase (Sigma C0130)/DMEM (31966 Gibco/Life technologies) supplemented with 1% penstrep (P0781, Sigma, Zwijndrecht, The Netherlands) at 37°C for 1¼ hrs and 2 hrs respectively. The individual myofibres were dissociated by gently passing them through Pasteur pipettes with different sized apertures and then abundantly washed, as described in detail elsewhere⁵⁵. To extract and expand the satellite cell pool, muscle fibers were cultured on matrigel (354230, BD biosciences, Breda, The Netherlands) coated 6-wells plate in DMEM 31966 supplemented with 30% FBS, 10% HS, 1% CEE and 2.5ng/ml FGF, 1% pen-strep (all Gibco/Life technologies), 150 fibers per well. After 3 days, the fibers

were detached and removed. The attached myoblasts were trypsinized, counted and plated for further analysis. To induce differentiation into myotubes, serum-rich medium was replaced with serum-poor medium (DMEM 31966), supplemented with 2% HS and 1% pen-strep, 48 hrs after plating. After 48 hours of differentiation cells were either fixed for immunofluorescence or lysed to isolate RNA.

Generation MEFs and adult skin fibroblasts

Mouse embryonic fibroblasts were generated from at embryonic stage E13.5. First, embryos were dissected out of the uterine horns, rinsed with 70% EtOH and washed in PBS (14190-169 Gibco/Life technologies). Next, the embryos were separated from the placenta and surrounding membranes. Tails of embryos were cut and used for genotyping. Next, the dark red organs were removed and embryos were finely minced and suspended in 1.5ml of trypsin-EDTA (25300-096, Gibco/Life technologies) for 15 minutes at 37°C with gentle shaking. Trypsin was inactivated by the addition of 2 ml MEF medium: DMEM high glucose (41966-052, Gibco/Life technologies) supplemented with 10% FCS, 1% L-glutamine (25030-024, Gibco/Life technologies) and 1% pen/strep. Upon centrifugation (5 minutes, 1200 rpm), the minced and trypsinized cell pellet was suspended in MEF medium and plated on gelatin coated culture dishes.

Adult skin fibroblasts were generated by dissecting skin tissue from the belly of 5 months old mice. The skin of each mice was dissociated overnight at RT in 2 ml dispase/collagenase mix, containing 2 mg dispase (17105-041, Invitrogen/Life Technologies), 2 mg collagenase (C-9891, sigma), 0.04 ml pen/strep, 0.04 ml glutamine, 0.25 mg fungizone and 0.2 mg gentamicine (all Gibco/Life technologies). Next day, cells were centrifuged (5 minutes, 1200rpm) and resuspended in 6 ml MEF medium and plated in T25 culture flasks.

Immunofluorescence

For co-IF staining of DUX4, myogenin (Myog) and myosin heavy chain, cells were fixed in 2% paraformaldehyde for 7 minutes at room temperature and then washed twice with PBS. Cells were permeabilized with 1% Triton X-100 (Sigma) in PBS for 10 minutes at room temperature with gentle rocking. Primary rabbit-DUX4 antibody directed against the C-terminal region of DUX4 (E5-5; 1:100)⁵⁶, Myog (1:100, Dako North America, Carpinteria CA, USA) and MF-20 anti myosin heavy chain (1:100, Developmental studies hybridoma bank, Univ. of Iowa, Iowa city, IA USA) were diluted in PBS and cells were incubated overnight at 4°C with the first antibody. After washing three times in PBS, followed an incubation with diluted Alexa 488 conjugated donkey anti-rabbit and Alexa 594 conjugated donkey anti-mouse (A21206, A21206, 1:500, Invitrogen/Life technologies) for one hour, gently rocking in the dark. Next, cells were washed three times with PBS-0.025% TWEEN before they were mounted on microscope slides using Aqua Poly/Mount (PolySciences, Warrington PA, U.S.A.) containing 500 ng/ml DAPI. Stained cells were analyzed on a Leica DMRA2 microscope (Leica microsystems, Wetzlar, Germany).

gDNA, RNA isolation & cDNA synthesis

gDNA isolation from different tissues of both D4Z4-2.5 and D4Z4-12.5 mice was carried

out using the Genomic DNA from tissue kit (740952, Machery-Nagel, Düren, Germany) following manufacturers instruction. Total RNA was isolated from tissue using miRNeasy kit (217004, Qiagen, Venlo, The Netherlands), including a DNase treatment, according to the instructions of the manufacturer. For C2C12 expression analysis, RNA was isolated using the RNeasy microkit (74004, Qiagen). Both DNA and RNA concentrations were determined using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington DE, USA). The quality of the RNA was assessed by using a RNA 6000 nanochip on an Agilent 2100 BioAnalyzer (Agilent Technologies Netherlands BV, Amstelveen, The Netherlands) and RIN scores > 9 were obtained. cDNA was synthesized using 1-3 µg total RNA using the Revert Aid H Minus first strand cDNA synthesis kit using oligo-dT primers (Fermentas/Thermo scientific, St. Leon-Rot, Germany) according to the manufacturer's instructions. The cDNA was subsequently treated with 0.5 U RNaseH for 20 minutes at 37°C and total cDNA was diluted in 50 µl water. For all C2C12 expression studies, oligo-dT primed cDNA was synthesized using 300 ng of RNA using the Omniscript RT kit (205111, Qiagen) according to manufacturer's instructions. RT was done at 50°C and cDNA was diluted to a total volume of 100 µl.

DNA methylation analysis by Southern blotting

Methylation levels of individual CpGs was determined by Southern blot analysis using the methylation-sensitive restriction enzymes BsaAI and FspI as described before¹². In short, 5 µg of genomic DNA was digested with the restriction enzymes EcoRI and BglII and either BsaAI (NEB, Ipswich MA, USA) or FspI (NEB). All digestions were performed according to the manufacturer's instructions. After digestion, DNA was separated by standard linear gel electrophoresis (0.8%) followed by Southern blotting of the DNA on a Hybond-XL membrane (GE Healthcare) and hybridization with the radioactive labeled probe p13E-11 (D4F104S1) or D4Z4 to determine the methylation in the proximal D4Z4 repeat unit and all internal units respectively. Hybridizations were performed for a minimum of 16 hours at 65°C in 0.125 M Na₂HPO₄ (pH 7.2), 0.25 M NaCl, 1 mM EDTA, 7% SDS and 100 µg/ml denatured fish sperm DNA (Roche). After hybridization, membranes were consecutively washed with 2xSSC/0.1%SDS, 1xSSC/ 0.1%SDS and 0.3xSSC/0.1%SDS. Finally, the membranes were exposed to a phospho-imager screen and signal intensities were quantified with ImageQuant software (Amersham/GE healthcare).

Quantitative methylation analysis by bisulphate converted DNA

Bisulphite treatment was performed with the EZ DNA Methylation kit (Zymo Research, Irvine CA, USA) and bisulphite primers were designed with MethPrimer software⁵⁷. Primer sequences are listed in **table S7**. PCR was performed in a final volume of 25 µl containing 250 µM dNTPs, 1X Supertaq PCR Buffer (HT Biotechnology Ltd, Cambridge UK), 10 pM of each primer and 1 U Silverstar DNA polymerase (Eurogentec, Maastricht, The Netherlands). Cycling conditions: 94 °C for 15 min followed by 40 cycles of 94°C for 40 s, 58°C for 40 s, and 72°C for 40 s, and a final extension step of 15 min at 72°C. PCR products were purified directly or by gel extraction using the NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel) and subjected to Sanger sequencing (LGTC). Quantitative methylation ratios were calculated from sequence traces using the ESME software,

including appropriate sequence quality control, normalization of signals, correction for incomplete bisulphite conversion, and mapping of positions in the trace file to CpGs in reference sequences⁵⁸.

Histone Chromatin immunoprecipitation

ChIP experiments were based on the protocol described by Nelson et al. with some modifications⁵⁹. In brief, subconfluent cell cultures were crosslinked in 1% formaldehyde for 10 minutes and the reaction was quenched for 5 minutes with glycine at a final concentration of 125 mM. Crosslinked cells were lysed and chromatin was sheared in a sonicator bath (Bioruptor UCD-20, Diagenode, Liège, Belgium) for 4-6 consecutive rounds of 10 minutes at maximum output and 15 seconds on/off cycles. Shearing was analyzed by phenol-chloroform extraction of DNA and agarose gel electrophoresis. All chromatin samples had a DNA size range between 200 – 2000 bp with a peak around 200 bp. Per reaction 3 µg (DNA content) of chromatin was precleared with blocked sepharose A beads (GE healthcare, Diegem, Belgium) and incubated overnight at 40 with antibodies: α H3K9me3 (17-625, Millipore, Billerica MA, USA; 5 µl/rxn) α H3K4me2 (07-030, Millipore; 5 ul/reaction) and total IgG (Millipore; 5 µl/rxn). IP was done with 20 µl sepharose A beads/reaction and washing was according to the online available Millipore ChIP protocol. DNA was isolated using Chelex resin and diluted 1:1 for qPCR analysis⁵⁹. Relative abundance, IgG corrected and input normalized, at D4Z4 was determined with previously published primers²⁷.

Quantitative PCR

All quantitative RT-PCR analysis were performed in duplicate using SYBR green mastermix on the MyIQ or CFX96 system (Bio-Rad, Veenendaal, The Netherlands) using 0.5-0.75 pM of each primer in a final volume of 10-15 µl per reaction. For gene expression analysis 2-5 µl of diluted cDNA and for ChIP analysis 5 µl of 1:1 diluted ChIP DNA was used per reaction. Cycling conditions: initial denaturation step at 95°C for 3 min, followed by 35-40 cycles of 10-15 s at 95°C and 45 s at primer Tm (**table S7**). Specificity of all reactions was monitored by standard gel electrophoresis and/or melting curve analysis: initial denaturation step at 95°C, followed by 1 min incubation at 65°C and sequential temperature increments of 0.5°C every 10 s up to 95°C. All primer sets were designed using Primer3 software and, for cDNA analysis, spanned at least one intron. Results were analyzed using iQ5 / Bio-Rad CFX manager version 2.0 (Bio-Rad). For cDNA, relative expression was calculated, using Cyclophilin, Hprt or Gapdh as a reference gene (indicated) for cDNA input, using the CFX manager software. For ChIP, relative quantification was done by background subtraction based on the signal in the normal IgG ChIP. Normalization was done using the relative abundance of the product in DNA isolated from ChIP input samples.

C2C12 culture, transfection and cell sorting

For C2C12 cultures, plates and dishes were coated with collagen (Purecol, Advanced Biomatrix, San Diego CA, USA) 1:30 diluted in MiliQ, for 1 hr at 37°C and then dried for at least 30 minutes. Before plating the cells, plates and dishes were washed with 1xPBS. C2C12 cells were maintained subconfluent in DMEM (11880, Gibco/Life technologies)

supplemented with 20% FCS, 1% pen/strep, 1% glucose and 1% glutamax (all Gibco/Life technologies). Transfections were performed using lipofectamine reagent, combined with plus reagent (both Invitrogen/Life Technologies) according to manufacturer's instructions in 12/6 well plates, 9 or 20 cm dishes with 0.6/1.5, 5 or 12 μ g total plasmid DNA, respectively. To enrich for DUX4 expressing cells used for gene expression analysis, pCS2-DUX4⁵ or the empty pCS2 backbone were equimolarly co-transfected with pEGFP-C1 (Life technologies) and then sorted on a FACS aria cell sorter II (BD biosciences). In short, the living single cells were first gated based on forward and side scatter and then sorted by gating for GFP. Both the GFP enriched and depleted fractions were collected in PBS supplemented with 1% FCS, spun down and stored at -80^o C.

Expression array analysis

Global gene expression changes upon DUX4 expression were obtained with illumina MouseWG-6 v2.0 expression arrays (Illumina, San Diego CA, USA). RNA from FACS sorted C2C12 transfected with pCS2-DUX4 were compared with pCS2 backbone transfected cells in triplicate. Labeling of RNA, hybridization of the arrays and primary data analysis were carried out by ServiceXS (Leiden, The Netherlands) according to manufacturer's instructions. Probe intensities were corrected and normalized using the lumiExpresso function from the lumi Bioconductor package⁶⁰ with default options. If a probe was not present in any of the 6 arrays according to lumi's "detectionCall" function, we removed it from further consideration. Differentially expressed probes were then identified using the limma Bioconductor package⁶¹ and p-values were adjusted to account for multiple testing using Benjamini and Hochberg's method⁶².

DUX4 ChIP-seq analysis

Crosslinking, generation of chromatin and ChIP were performed as indicated above. In brief, chromatin of wt and DUX4 transfected C2C12s containing 12 μ g DNA was precleared with IP beads and incubated o/n at 4^o C with the rabbit polyclonal MO489 antibody directed against the C-terminus of DUX4⁵. After washing 6 times in ChIP buffer, immunoprecipitated ChIP DNA was isolated and purified with a standard phenol extraction. Sequencing and sample preparation were done with the Illumina genome analyzer following manufacturer's instructions. Data analysis was performed as previously described⁵. Briefly, sequences were extracted by GApipeline-0.3.0. Reads mapping to the X and Y-chromosomes were excluded from our analysis. Duplicate sequences were discarded to minimize effects of PCR amplification and each read was extended in the sequencing orientation to a total of 200 bases to infer the coverage at each genomic position. Peak calling was performed by an in-house developed R package, which models background reads by a negative binomial distribution as previously described⁶³. To identify the DUX4 consensus binding sequence, we applied an in-house developed Bioconductor package motifRG for discriminative de novo motif discovery as previously described^{64, 65}.

References

1. Clapp,J. et al. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* 81, 264-279 (2007).
2. Dixit,M. et al. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci. U. S. A* 104, 18157-18162 (2007).
3. Gabriels,J. et al. 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 (1999).
4. Lyle,R., Wright,T.J., Clark,L.N., & Hewitt,J.E. 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 (1995).
5. Geng,L.N. et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38-51 (2012).
6. Snider,L. et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS. Genet.* 6, e1001181 (2010).
7. Tawil,R. & Van Der Maarel,S.r.M. Facioscapulohumeral muscular dystrophy. *Muscle & Nerve* 34, 1-15 (2006).
8. de Greef,J.C. et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mutat.* 30, 1449-1459 (2009).
9. Jones,T.I. et al. Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. *Hum. Mol. Genet.* 21, 4419-4430 (2012).
10. Lemmers,R.J. et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* 44, 1370-1374 (2012).
11. van Deutekom,J.C. et al. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* 2, 2037-2042 (1993).
12. van Overveld,P.G. et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet* 35, 315-317 (2003).
13. Wijmenga,C. et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 2, 26-30 (1992).
14. Tassin,A. et al. DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? *J. Cell Mol. Med.*(2012).
15. Lemmers,R.J. et al. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am. J Hum. Genet* 81, 884-894 (2007).
16. Lemmers,R.J. et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650-1653 (2010).
17. van der Maarel,S.M., Tawil,R., & Tapscott,S.J. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. *Trends Mol. Med.* 17, 252-258 (2011).
18. Leidenroth,A. & Hewitt,J.E. A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC. Evol. Biol.* 10, 364 (2010).
19. Furst,D.O., Osborn,M., & Weber,K. Myogenesis in the mouse embryo: differential onset of expression of myogenic proteins and the involvement of titin in myofibril assembly. *J. Cell Biol.* 109, 517-527 (1989).
20. Cusella-De Angelis,M.G. et al. MyoD, myogenin independent differentiation of primordial myoblasts in mouse somites. *J. Cell Biol.* 116, 1243-1255 (1992).
21. Messina,G. & Cossu,G. The origin of embryonic and fetal myoblasts: a role of Pax3 and Pax7. *Genes Dev.* 23, 902-905 (2009).
22. Buckingham,M. et al. The formation of skeletal muscle: from somite to limb. *J. Anat.* 202, 59-68 (2003).
23. Yamanaka,G. et al. Tongue atrophy in facioscapulohumeral muscular dystrophy. *Neurology* 57, 733-5 (2001).
24. Snider,L. et al. RNA Transcripts, miRNA-sized Fragments, and Proteins Produced from D4Z4 Units: New Candidates for the Pathophysiology of Facioscapulohumeral Dystrophy. *Hum. Mol. Genet* 18, 2414-2430 (2009).
25. de Greef,J.C., Frants,R.R., & van der Maarel,S.M. Epigenetic mechanisms of facioscapulohumeral

- muscular dystrophy. *Mutat. Res.* 647, 94-102 (2008).
26. van Overveld, P.G. et al. Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. *Ann. Neurol.* 58, 569-576 (2005).
 27. Zeng, W. et al. Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats in facioscapulohumeral dystrophy (FSHD). *PLoS. Genet* 7, e1000559 (2009).
 28. Balog, J. et al. Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. *Epigenetics.* 7, 579-584 (2012).
 29. Hagiwara, K. et al. Mouse SWAM1 and SWAM2 are antibacterial proteins composed of a single whey acidic protein motif. *J. Immunol.* 170, 1973-1979 (2003).
 30. Statland, J.M. & Tawil, R. Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions. *Curr. Opin. Neurol.* 24, 423-428 (2011).
 31. Turk, R. et al. Common pathological mechanisms in mouse models for muscular dystrophies. *FASEB J.* 20, 127-129 (2006).
 32. Bosnakovski, D. et al. An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *Embo J* 27, 2766-2779 (2008).
 33. Kowaljow, V. et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord* 17, 611-623 (2007).
 34. Gabellini, D., Green, M., & Tupler, R. Inappropriate Gene Activation in FSHD. A Repressor Complex Binds a Chromosomal Repeat Deleted in Dystrophic Muscle. *Cell* 110, 339-248 (2002).
 35. van Deutekom, J.C.T. et al. Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. *Hum Mol Genet* 5, 581-590 (1996).
 36. Gabellini, D. et al. Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature*(2005).
 37. Wuebbles, R.D., Long, S.W., Hanel, M.L., & Jones, P.L. Testing the effects of FSHD candidate gene expression in vertebrate muscle development. *Int. J. Clin. Exp. Pathol.* 3, 386-400 (2010).
 38. Wuebbles, R.D., Hanel, M.L., & Jones, P.L. FSHD region gene 1 (FRG1) is crucial for angiogenesis linking FRG1 to facioscapulohumeral muscular dystrophy-associated vasculopathy. *Dis. Model. Mech.* 2, 267-274 (2009).
 39. Jiang, G. et al. Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum Mol. Genet* 12, 2909-2921 (2003).
 40. Klooster, R. et al. Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. *Eur. J. Hum. Genet.* 17, 1615-1624 (2009).
 41. Osborne, R.J., Welle, S., Venance, S.L., Thornton, C.A., & Tawil, R. Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology* 68, 569-577 (2007).
 42. Homma, S. et al. A unique library of myogenic cells from facioscapulohumeral muscular dystrophy subjects and unaffected relatives: family, disease and cell function. *Eur. J. Hum. Genet.* 20, 404-410 (2012).
 43. Bosnakovski, D., Daughters, R.S., Xu, Z., Slack, J.M., & Kyba, M. Biphasic myopathic phenotype of mouse DUX, an ORF within conserved FSHD-related repeats. *PLoS. ONE.* 4, e7003 (2009).
 44. Wallace, L.M. et al. DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy *in vivo*. *Ann. Neurol.* 69, 540-552 (2011).
 45. Mitsuhashi, H., Mitsuhashi, S., Lynn-Jones, T., Kawahara, G., & Kunkel, L.M. Expression of DUX4 in zebrafish development recapitulates facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* (2012).
 46. Wallace, L.M. et al. RNA Interference Inhibits DUX4-induced Muscle Toxicity *In Vivo*: Implications for a Targeted FSHD Therapy. *Mol. Ther.* 20, 1417-1423 (2012).
 47. de Greef, J.C. et al. Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. *Neurology* 69, 1018-1026 (2007).
 48. Cabianca, D.S. et al. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* 149, 819-831 (2012).
 49. Fitzsimons, R.B. Retinal vascular disease and the pathogenesis of facioscapulohumeral muscular dystrophy. A signalling message from Wnt? *Neuromuscul. Disord.* 21, 263-271 (2011).
 50. van Deutekom, J.C. et al. FSHD associated DNA rearrangements are due to deletions of integral

- copies of a 3.2 kb tandemly repeated unit. *Hum Mol Genet* 2, 2037-2042 (1993).
51. van Geel, M. et al. The FSHD region on human chromosome 4q35 contains potential coding regions among pseudogenes and a high density of repeat elements. *Genomics* 61, 55-65 (1999).
 52. Szuhai, K. & Tanke, H.J. COBRA: combined binary ratio labeling of nucleic-acid probes for multi-color fluorescence in situ hybridization karyotyping. *Nat. Protoc.* 1, 264-275 (2006).
 53. White, S.J., Breuning, M.H., & den Dunnen, J.T. Detecting copy number changes in genomic DNA: MAPH and MLPA. *Methods Cell Biol.* 75, 751-768 (2004).
 54. Schouten, J.P. et al. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* 30, e57 (2002).
 55. Collins, C.A. & Zammit, P.S. Isolation and grafting of single muscle fibres. *Methods Mol. Biol.* 482, 319-330 (2009).
 56. Geng, L.N., Tyler, A.E., & Tapscott, S.J. Immunodetection of human double homeobox 4. *Hybridoma (Larchmt.)* 30, 125-130 (2011).
 57. Li, L.C. & Dahiya, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics.* 18, 1427-1431 (2002).
 58. Lewin, J., Schmitt, A.O., Adorjan, P., Hildmann, T., & Piepenbrock, C. Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplicates. *Bioinformatics.* 20, 3005-3012 (2004).
 59. Nelson, J.D., Denisenko, O., & Bomsztyk, K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.* 1, 179-185 (2006).
 60. Du, P., Kibbe, W.A., & Lin, S.M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics.* 24, 1547-1548 (2008).
 61. Wettenhall, J.M. & Smyth, G.K. limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics.* 20, 3705-3706 (2004).
 62. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 289-300 (1995).
 63. Fong, A.P. et al. Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev. Cell* 22, 721-735 (2012).
 64. Cao, Y. et al. Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev. Cell* 18, 662-674 (2010).
 65. Pali, C.G. et al. Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. *EMBO J.* 30, 494-509 (2011).

Acknowledgements

We thank Corry van Teijlingen and Weike Wang for technical assistance. We thank the employees of the Transgene Facility Leiden for generating the transgenic mice. The MF-20 developed by D.A. Fischman et al., was obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by The University of Iowa, Department of Biology, Iowa City, IA 52242.

Competing interests

The authors have declared that no competing interests exist.

Supporting information

Fig. S1: In situ hybridization to detect DUX4 mRNA in D4Z4-2.5 mouse testis

Fig. S2: DUX4 Expression analysis in muscles of D4Z4-2.5 and D4Z4-12.5 mice

Fig. S3: DUX4 Expression analysis in non-muscle tissue of D4Z4-2.5 and D4Z4-12.5 mice

Fig. S4: CpG methylation analysis of D4Z4 in D4Z4-2.5 and D4Z4-12.5 mice

Fig. S5: Intensity plot of array probes of C2C12+pCS2-DUX4 versus C2C12+pCS2

Fig. S6: Consensus binding site sequence sequences of DUX4 at different genetic contexts

Fig. S7: Distribution of DUX4 binding sites in the mouse genome

Fig. S8: Luciferase reporter assays using a direct DUX4 target site

Fig. S9: Expression analysis of the DUX4 target *Wfdc3* in E9.5 embryos

Fig. S10: Keratitis in D4Z4-2.5 mice at different ages

Table S1: Deregulated genes in response to ectopic DUX4 expression in C2C12 myoblasts

Table S2: DUX4 induced deregulated genes overlapping between C2C12 and human myoblasts

Table S3: Identified direct targets of DUX4 in C2C12 myoblasts

Table S4: Unique DUX4 binding sites in transposable elements in C2C12 myoblasts

Table S5: DUX4 binding sites in transposable elements in C2C12 myoblasts

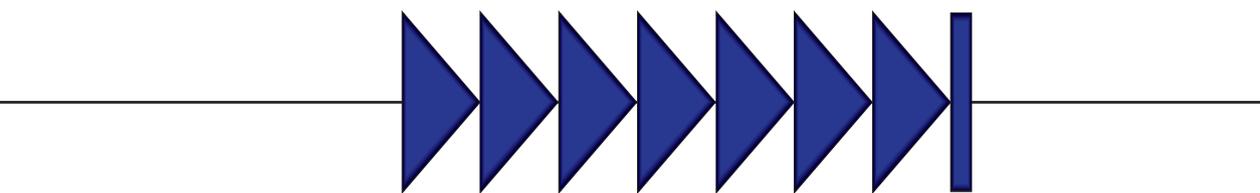
Table S6: Overview of the histological and functional tests performed to investigate muscle integrity and performance in D4Z4-2.5 mice

Table S7: List of primers and corresponding sequences

Text S1: Material & Methods and References

All supporting information belonging to this chapter can be accessed through <http://goo.gl/qV8XZa> or by using the QR-code below.





Mutations in *CDCA7* and *HELLS* cause immunodeficiency, centromeric instability and facial anomalies syndrome.

7

Peter E. Thijssen, Yuya Ito, Giacomo Grillo, Jun Wang, Guillaume Velasco, Hirohisa Nitta, Motoko Unoki, Minako Yoshihara, Mikita Suyama, Yu Sun, Richard J.L.F. Lemmers, Jessica C. de Greef, Andrew Gennery, Paolo Picco, Barbara Kloeckener-Gruissem, Tayfun Güngör, Ismail Reisli, Capucine Picard, Kamila Kebaili, Bertrand Roquelaure, Tsuyako Iwai, Ikuko Kondo, Takeo Kubota, Monique M. van Ostaijen-Ten Dam, Maarten J.D. van Tol, Corry Weemaes, Claire Francastel, Silvère M. van der Maarel and Hiroyuki Sasaki
Thijssen et al. (2015), *Nature Communications*, 6:7870

Abstract

The life threatening immunodeficiency, centromeric instability, and facial anomalies syndrome is a genetically heterogeneous autosomal recessive disorder. Twenty percent of all patients cannot be explained by mutations in the ICF genes DNA methyltransferase 3B or zinc finger and BTB domain containing 24. Here we report mutations in the cell division cycle associated 7 and the helicase, lymphoid-specific genes in ten unexplained ICF cases. Our data highlights the genetic heterogeneity of ICF syndrome and provides evidence that all genes act in common or converging pathways leading to the ICF phenotype.

Introduction

ICF syndrome is characterized by recurrent and often fatal respiratory and gastrointestinal infections as a consequence of hypo- or a-gammaglobulinemia in the presence of B-cells^{1, 2}. Nearly all patients also present with distinct facial anomalies, including hypertelorism, flat nasal bridge and epicanthus^{1, 2}. Centromeric instability is the cytogenetic hallmark of ICF syndrome. It involves the juxtacentromeric heterochromatin repeats on chromosomes 1, 9, and 16 and is comparable to what is observed after treatment of cells with demethylating agents^{3, 4}. Therefore, CpG hypomethylation of juxtacentromeric satellites type II and III is diagnostic for ICF syndrome, with additional hypomethylation of centromeric α -satellite repeats in *DNMT3B* mutation-negative patients^{5, 6}.

Mutations in the DNA methyltransferase 3B (*DNMT3B*) gene (OMIM 602900; ICF1) account for ~50% of ICF cases while ~30% of cases have mutations in the zinc finger and BTB domain containing 24 (*ZBTB24*) gene (OMIM 614064; ICF2)^{2, 5}. *DNMT3B* is a de novo DNA methyltransferase, primarily acting during early development with a preference for CpG dense regions⁷. ICF1 mutations in the catalytic domain of *DNMT3B* result in largely reduced methyltransferase activity^{8, 9}. The function of *ZBTB24* is unknown, but it belongs to a family of ZBTB proteins of which many have regulatory roles in hematopoietic differentiation^{10, 11}. Despite the successful identification of ICF genes, the pathophysiological mechanism underlying the syndrome remains largely unresolved.

Previous studies indicated further genetic heterogeneity in ICF syndrome². To identify the genetic cause in genetically unexplained cases (ICFX), we combined homozygosity mapping with whole exome sequencing. By using an autosomal recessive inheritance model and prioritizing homozygous variants in consanguineous families, we now identify four different homozygous and potentially damaging variants in the cell division cycle associated 7 (*CDCA7*) gene in five ICFX patients (now referred to as ICF3). In an additional five ICFX patients we identify compound heterozygous and homozygous variants in the helicase, lymphoid-specific (*HELLS*) gene (ICF4). We show that knock down of both new ICF genes leads to hypomethylation of juxtacentromeric heterochromatin repeats in a murine cell model. Our results emphasize the genetic heterogeneity of ICF syndrome, nonetheless provide evidence that all four ICF genes are involved in at least one common pathway.

Results

Missense mutations in CDCA7 cause ICF syndrome type 3

We selected 13 ICFX patients from 11 families, negative for mutations in *DNMT3B* or *ZBTB24*, of whom the clinicopathological characteristics are listed in **Table S1**. Hypomethylation of pericentromeric satellite type II (Sat II), common to all ICF patients, and centromeric α -satellite DNA repeats, shown to be affected only in ICF2 and ICFX, was shown before for a subset of patients^{2, 5, 6, 12}. For an additional set of ICFX patients we confirmed Sat II and α -satellite hypomethylation using Southern blot analysis (**Fig. S1**). In five ICFX patients from four families we identified and confirmed homozygous missense mutations in *CDCA7*, all near the first two zinc finger motifs in the conserved carboxyterminal 4-CXXC-type zinc finger domain (**Fig. 1A-E**). Segregation with disease

was confirmed in family D and all mutations were predicted to be pathogenic and have an allele frequency supporting pathogenicity (**Fig. 1E, Table S1**). *CDCA7* is involved in neoplastic transformation, MYC-dependent apoptosis, and hematopoietic stem cell emergence, however its molecular function is unknown^{13, 14}. All four zinc finger motifs are completely conserved in the highly homologous 4-CXXC zinc finger domain of the transcriptional repressor CDCA7-Like (*CDCA7L*) (**Fig. S2**). The repressive activity of

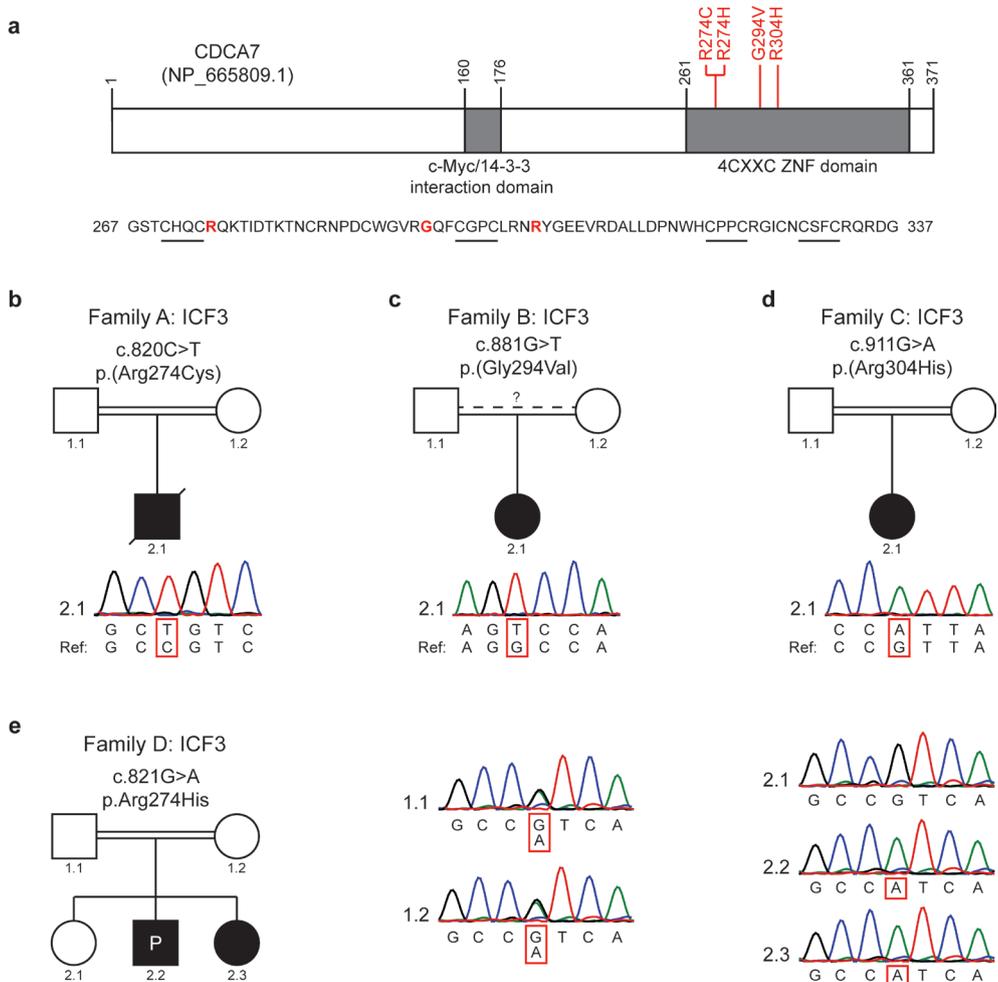


Figure 1: Homozygous missense mutations in *CDCA7* in five ICF3 patients.

A) Schematic representation of *CDCA7*, with the identified homozygous missense mutations in red. Sequence outtake: 4-CXXC type zinc finger domain, CXXC motifs are underlined, mutated residues in red. **B-D)** Sanger sequencing confirmation of missense mutations in *CDCA7* in families A-C. All variants were homozygous, the reference sequence is displayed for comparison. **E)** Sanger sequencing confirmation of the homozygous missense *CDCA7* mutation in patients 2.2 (proband) and 2.3 of family D. Both parents are heterozygous for the variant, sibling 2.1 is unaffected and homozygous for the wild type allele.

CDCA7L is dependent on its 4-CXXC domain and by homology CDCA7 mutations in ICF3 may disrupt a similar function¹⁵.

Mutations in HELLS cause ICF syndrome type 4

In an additional five ICFX patients from four families we identified mutations in *HELLS* that were predicted to be pathogenic and with allele frequencies supporting pathogenicity (**Fig. 2A, Table S1, ICF4**). In affected members of family E we identified a missense mutation in the conserved helicase domain (c.2096A>G; p.Gln699Arg) and an intronic mutation leading to destruction of the splice donor site in intron 5 (c.370+2T>A) (**Fig. 2B**). Different allelic origin is supported by analysis of maternal DNA, which carried only the splice site mutation (**Fig. 2C**). To analyse the effect of the lost splice donor site on mRNA processing, fibroblasts of both patients were treated with cycloheximide to inhibit nonsense mediated decay. Upon treatment, RT-PCR analysis showed increased levels of a splice variant with complete skipping of exon 5, leading to a frameshift followed by a premature stop codon in exon 6 (**Fig. 2C**).

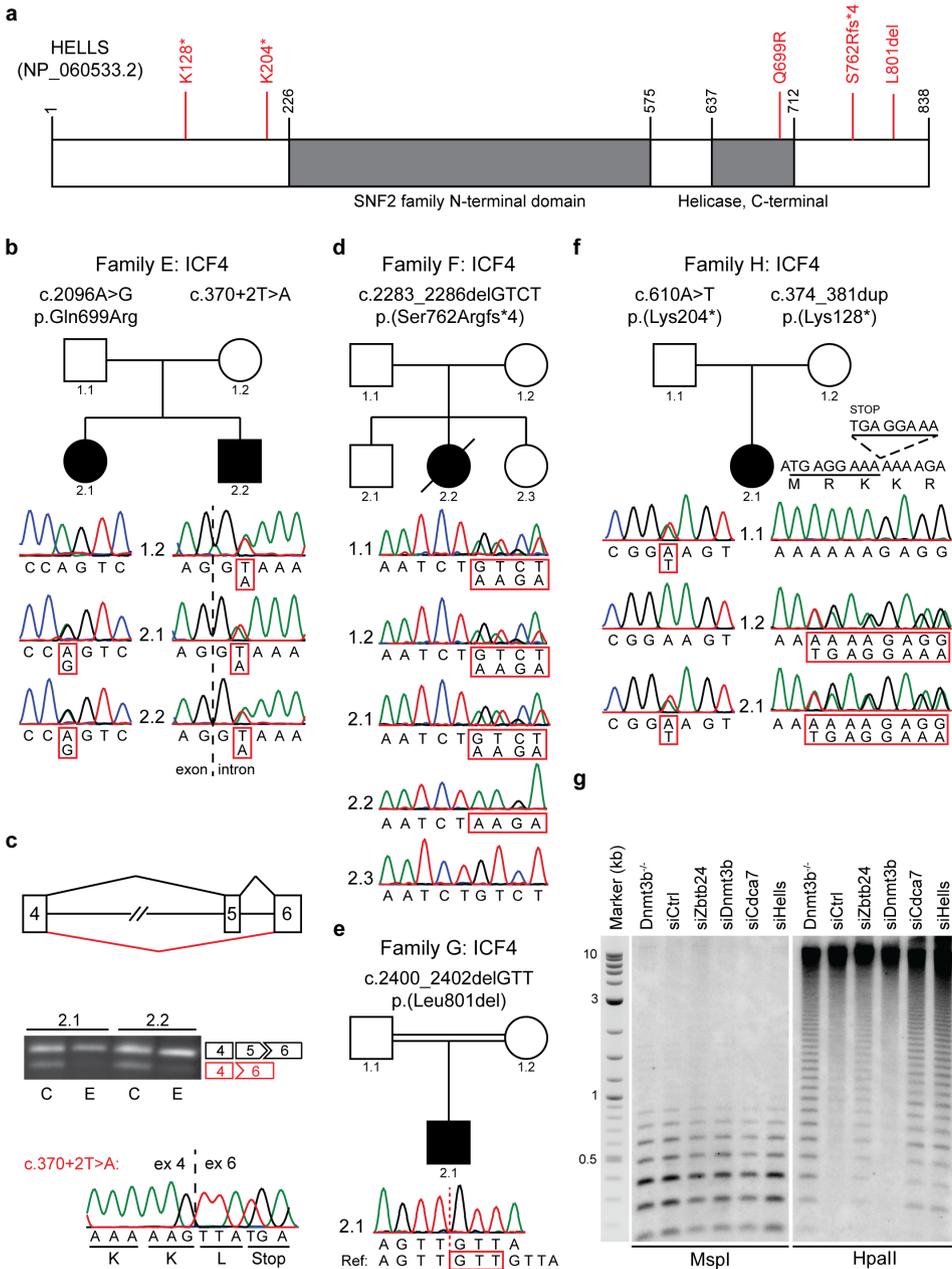
The patient in family F carries a homozygous out-of-frame deletion (c.2283_2286delGTCT; p.[Ser762Argfs*4]), resulting in a frameshift and introduction of a premature stop codon in exon 20. The unaffected siblings 2.1 and 2.3 were found to be heterozygous for the deletion or homozygous for the wild type allele, respectively (**Fig. 2D**). In family G we found a deleterious homozygous in-frame deletion in the C-terminal domain of *HELLS*, leading to the deletion of Leucine 801 (c.2400_2402delGTT; p.[Leu801del]) (**Fig. 2E**). The proband of family H carries a nonsense mutation (c.610A>T; p.[Lys204*]) and a duplication causing the insertion of a stop codon (c.374_381dup; p.[Lys128*]), suggesting that absence of *HELLS* is compatible with human life, whereas *Hells*^{-/-} mice die perinatally¹⁶. Different allelic origin of the mutations was confirmed in parental DNA (**Fig. 2F**).

ICF genes converge at centromeric DNA methylation regulation

In mouse, *HELLS* is required for T-cell proliferation and mediates de novo DNA methylation, through its interaction with DNMT3B, dependent on its ATPase domain¹⁶⁻¹⁹. Genome wide loss of CpG methylation, including centromeric repeats, has been observed in *Hells*^{-/-} mice, reminiscent of what has been described in *Dnmt3b* knockout mice and in mouse models for ICF^{120, 21}. We show that transient depletion of *HELLS*, *CDCA7* and *ZBTB24*, but not *DNMT3B*, resulted in decreased CpG methylation at centromeric repeats in wild type (wt) mouse embryonic fibroblasts (MEFs, **Fig. 2G, Fig. S2A-B**). This confirms that *DNMT3B* acts during establishment of centromeric CpG methylation²². Moreover, it supports a role for *ZBTB24* and *CDCA7* in maintenance of CpG methylation at centromeric repeats, and, combined with previously published work, suggests that *HELLS* may be involved in both processes¹⁹.

By identifying two new ICF syndrome genes this study highlights its genetic heterogeneity, and the identification of at least one additional disease gene is expected with still a few cases remaining genetically unresolved. The complex, but highly overlapping pathophysiology suggests that all ICF genes act in common or converging pathways involved in immunity, chromatin regulation and development. Convergence is supported

by hypomethylation of pericentromeric repeats, common to all ICF subgroups, however the result of different defective pathways in the establishment and/or maintenance of CpG methylation.



Methods

Patients

All samples were obtained in an anonymized fashion and all families gave consent for genetic analyses. The study was approved by the Medical Ethical Committee of the Leiden University Medical Center, the local ethics committee of Necker-Enfants Malades Hospital, Paris, France and the Kyushu University Institutional Review Board for Human Genome/Gene Research. The patient from family A (referred to as patient 2 in Kloeckener-Gruissem et. al., 2005), patients from families B and C (referred to as pC and pS, respectively, in Velasco et. al., 2014), patient 2.2 from family D, and the patients in family E (referred to as P4 and P8/P9, respectively, in De Greef et. al., 2011), as well as the three remaining ICFX cases (referred to as pN and P1 in Velasco et. al., 2014 and P4 in Kubota et. al., 2004) all show typical clinical features of ICF syndrome including hypo- or agammaglobulinemia in the presence of B cells, combined with the classical cytogenetic abnormalities involving chromosomes 1, 9 and 16, and hypomethylation of α -satellite DNA in addition to SatII hypomethylation. The critical features of all patients are summarized in supplementary table 1, further details can be found in previous descriptions of these patients^{2, 5, 6, 12, 23, 24}.

Gene identification by homozygosity mapping and sequencing

Homozygosity mapping was performed using the Sentrix HumanHap-300 Genotyping BeadChips (Illumina). To this end, 750ng genomic DNA was processed with the Infinium II Whole-Genome Genotyping Assay (Illumina). After DNA amplification, fragmentation, precipitation and resuspension, DNA was applied to the BeadChip and incubated overnight, followed by enzymatic base extension, fluorescent staining of the beads, and detection of fluorescent intensities by the BeadArray Reader (Illumina). To identify regions of homozygosity, B allele frequencies were assessed for all SNPs using BeadStudio version 3.2 (Illumina). For a subset of patients, whole exome sequencing was performed in the Neuromics project by deCODE Genetics (Reykjavik – Iceland) and analyzed using deCODE Clinical Sequence Miner. Recessive analysis for multiple cases and controls and gene variant effect count (with VEP consequences moderate to high) were used

Figure 2: Mutations in HELLS in five ICF4 patients.

A) Schematic representation of HELLS, with the identified mutations in red. **B)** Sanger sequencing confirmation of *HELLS* mutations in family E. Only c.370+2T>A was identified in maternal DNA, indicating different allelic origin of both mutations, or de novo occurrence of the second mutation. **C)** RT-PCR analysis of *HELLS* mRNA upon treatment of patient derived fibroblasts with cycloheximide (C) revealed that c.370+2T>A leads to complete skipping of exon 5 and disruption of the open reading frame. Ethanol treated samples (E) served as controls, alternative splicing was confirmed by Sanger sequencing in two independent experiments for both samples. **D)** Sanger sequencing confirmation of a homozygous out-of-frame deletion in *HELLS* in family F. Both parents, as well as unaffected sibling 2.1 are heterozygous for the deletion allele, unaffected sibling 2.3 is homozygous for the wildtype allele. **E)** Sanger sequencing confirmation of a homozygous in-frame deletion in *HELLS* in family G. Both parents are heterozygous for the deletion allele. **F)** Sanger sequencing confirmation of nonsense mutations in *HELLS* in family H. Different allelic origin was confirmed in parental DNA. **G)** Southern blot analysis of minor satellite DNA methylation in wt, *Dnmt3b*^{-/-} and siRNA-treated wt MEFs after digesting DNA with MspI or its methylation sensitive isoschizomer HpaII revealed CpG hypomethylation upon knockdown of Zbtb24, Cdca7 and Hells. Molecular weights of the 2-Log DNA size marker are in kilo-basepairs.

to identify possible recessive mutations. For an additional set of patients, DNA libraries for whole exome sequencing were constructed using the SureSelect Human All Exon V5 kit (Agilent Technologies) according to the manufacturer's instructions. Sequencing was performed on the Illumina HiSeq2500 platform to generate 100bp paired-end reads. Reads were mapped to the reference human genome (UCSC hg19) with the Burrows-Wheeler Alignment tool (BWA v0.7.4)²⁵. Duplicate reads were removed by Picard (v1.87). We called SNVs and indels using the Genome Analysis Toolkit (GATK v2.5-2)²⁶. Annotations of variants were made using ANNOVAR²⁷. For confirmation, relevant exons and flanking sequences of *CDCA7* and *HELLS* were amplified using standard PCR, products were purified and analysed by Sanger sequencing. Sequence tracks were analysed and visualized using ContigExpress (Vector NTI, Invitrogen-Life Technologies). PCR primers are listed in **table S2**.

Cycloheximide treatment and RT-PCR analysis

Early passage ($p < 6$) patient derived primary fibroblast were maintained in DMEM F12 (31331) supplemented with 20% heat inactivated Fetal Calf Serum, 1% pen-strep, 1% sodium pyruvate and 1% HEPES (all Invitrogen-Life Technologies). Cells were treated with 250 $\mu\text{g ml}^{-1}$ cycloheximide (dissolved in ethanol, 01810, Sigma-Aldrich) for 4 hours using equal volumes of ethanol as control. After treatment, cells were harvested in Qiazol lysis reagent and RNA was isolated using the miRNeasy mini kit (both Qiagen) all according to manufacturer's instructions. 2 μg of total RNA was used for random primed cDNA synthesis using the RevertAid first-strand cDNA synthesis kit (Thermo scientific). Transcripts were amplified by standard PCR with primers listed in supplementary table 2, separated by standard gel electrophoresis and sequenced by Sanger sequencing.

Knockdown of gene expression in MEFs

The use of animal work has been reviewed by the Animal Experimentation Ethical Committee Buffon (CEEA-40), Paris, France, and approved under the number CEB-06-2012. Female C57BL/6N pregnant mice, age 3 to 6 months, were sacrificed at 12.5 dpc, and individual mouse embryonic fibroblasts (MEF) clones isolated from each embryo of the litter. Primary MEFs were cultivated for no more than 2-3 passages in complete media (DMEM glutamax supplemented with 10% FBS, 100 U mL^{-1} penicillin, 100 $\mu\text{g mL}^{-1}$ streptomycin, all from Life Technologies). Two rounds of transfection with synthetic siRNAs at a final concentration of 20 nM were performed using Interferin transfection reagent (Polyplus-transfection) following the manufacturer's instructions: a first round on MEFs in suspension and the second when cells were allowed to adhere to the plate. Cells were collected after 48 hours for genomic DNA extraction. Sequences of siRNAs used in the study are listed in **Table S2**.

Satellite DNA methylation analysis by Southern blotting

Satellite II and α -satellite methylation in whole blood DNA was analysed by Southern blot analysis of 2 μg genomic DNA digested with methylation-sensitive restriction enzymes HhaI for analysis of Sat- α or BstBI for analysis of SatII repeats. Both enzymes were purchased from Fermentas. After overnight digestion, DNA was separated by electrophoresis using 0.8% agarose gels. DNA was transferred overnight to Hybond-N+

membranes (GE Healthcare) and hybridized with a radioactive probes recognizing the α -satellite and satellite II repeats, respectively. Signals were detected by phosphoimaging or by exposure to X ray films. For analysis of murine satellite repeats, genomic DNA was extracted and purified from MEFs using the NucleoSpin®Tissue kit (Macherey-Nagel) according to manufacturer's instructions. The DNA pellet was eluted in TE containing 20 $\mu\text{g mL}^{-1}$ RNase A. Genomic DNA from MEFs (500 ng) was digested with 20 units of MspI or HpaII (New England Biolabs) for 16h to analyze the DNA methylation patterns of centromeric minor satellite repeats. The digested DNA fragments were separated by electrophoresis using 1% agarose gels and transferred overnight to Hybond-N+ membranes (GE Healthcare) in 20XSSC. After UV-crosslink, the membranes were pre-hybridized in 6X SSC, 5X Denhardt and 0.1% SDS and then hybridized with 32P-labeled minor satellite oligonucleotide probe:

(5'-ACATTTCGTGGAAACGGGATTTGTAGAACAGTGTATATCAATGAGTTACAATGAGAAACAT). Pre-hybridization and hybridization were carried out at 42°C for 1h. The membranes were washed 3 times in 6X SSC and 0.1%SDS at 37°C and signals detected by phosphorimaging using FLA 7000 phosphorimager (Fuji). Uncropped scans of the Southern blots are presented in **Figs. S4-S5**.

Quantification of knockdown efficiency by qRT-PCR

Total RNA from MEFs was isolated using TRIzol® Reagent (Life Technologies) according to manufacturer's instructions. Contaminant genomic DNA was eliminated with TURBO DNA-free kit (Ambion). Reverse transcription was carried out using 1 μg DNA-free RNA and 50 μM random hexamers, 20U of RNase Out and 100U of RevertAid reverse transcriptase (Life Technologies). Complementary DNA reactions were used as templates for PCR reactions. Real-time PCR was performed using the light cycler-DNA MasterPLUS SYBR Green I mix (Thermo Scientific) supplemented with 0.5 μM of specific primer pairs (listed in supplementary table 2). Real-time quantitative PCRs were run on a light cycler rapid thermal system (LightCycler®480 2.0 Real time PCR system, Roche) with 20 sec of denaturation at 95°C, 20 sec of annealing at 60°C and 20 sec of extension at 72°C for all primers, and analyzed by the comparative CT (ΔCT) method.

References

- Hagleitner, M.M. et al. Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). *J. Med. Genet.* 45, 93-99 (2008).
- Weemaes, C.M. et al. Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. *Eur. J. Hum. Genet.* 21, 1219-1225 (2013).
- Jeanpierre, M. et al. An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Hum. Mol. Genet.* 2, 731-735 (1993).
- Tiepolo, L. et al. Multibranching chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. *Hum. Genet.* 51, 127-137 (1979).
- de Greef, J.C. et al. Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. *Am. J. Hum. Genet.* 88, 796-804 (2011).
- Jiang, Y.L. et al. DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. *Hum. Mutat.* 25, 56-63 (2005).
- Auclair, G., Guibert, S., Bender, A., & Weber, M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol.* 15, 545 (2014).
- Gowher, H. & Jeltsch, A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J. Biol. Chem.* 277, 20409-20414 (2002).
- Moarefi, A.H. & Chedin, F. ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. *J. Mol. Biol.* 409, 758-772 (2011).
- Nitta, H. et al. Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. *J. Hum. Genet.* 58, 455-460 (2013).
- Lee, S.U. & Maeda, T. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. *Immunol. Rev.* 247, 107-119 (2012).
- Velasco, G. et al. Germline genes hypomethylation and expression define a molecular signature in peripheral blood of ICF patients: implications for diagnosis and etiology. *Orphanet. J. Rare. Dis.* 9, 56 (2014).
- Gill, R.M., Gabor, T.V., Couzens, A.L., & Scheid, M.P. The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation. *Mol. Cell Biol.* 33, 498-513 (2013).
- Guiu, J. et al. Identification of Cdc7 as a novel Notch transcriptional target involved in hematopoietic stem cell emergence. *J. Exp. Med.* (2014).
- Chen, K., Ou, X.M., Wu, J.B., & Shih, J.C. Transcription factor E2F-associated phosphoprotein (EAPP), RAM2/CDCA7L/JPO2 (R1), and simian virus 40 promoter factor 1 (Sp1) cooperatively regulate glucocorticoid activation of monoamine oxidase B. *Mol. Pharmacol.* 79, 308-317 (2011).
- Geiman, T.M. & Muegge, K. Lsh, an SNF2/helicase family member, is required for proliferation of mature T lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4772-4777 (2000).
- Ren, J. et al. The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. *Nucleic Acids Res.* 43, 1444-1455 (2015).
- Yu, W. et al. Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. *Genome Res.* 24, 1613-1623 (2014).
- Zhu, H. et al. Lsh is involved in de novo methylation of DNA. *EMBO J.* 25, 335-345 (2006).
- Ueda, Y. et al. Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development* 133, 1183-1192 (2006).
- Velasco, G. et al. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl. Acad. Sci. U. S. A.* 107, 9281-9286 (2010).
- Okano, M., Bell, D.W., Haber, D.A., & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247-257 (1999).
- Kloeckener-Gruissem, B., Betts, D.R., Zankl, A., Berger, W., & Gungor, T. A new and a reclassified ICF patient without mutations in DNMT3B and its interacting proteins SUMO-1 and UBC9. *Am. J. Med. Genet. A* 136, 31-37 (2005).
- Kubota, T. et al. ICF syndrome in a girl with DNA hypomethylation but without detectable DNMT3B

- mutation. *Am. J. Med. Genet. A* 129A, 290-293 (2004).
25. Li,H. & Durbin,R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 26, 589-595 (2010).
 26. McKenna,A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297-1303 (2010).
 27. Wang,K., Li,M., & Hakonarson,H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164 (2010).

Acknowledgements

We thank all patients and their families for their consent to contribute to our study. We thank K. Ichiyonagi and L. Daxinger for helpful advice, staff of Kyushu University International Legal Office for help in legal and ethical affairs, and T.P. Pham, S.L. Höcker, M. Miyake and H. Furuumi for technical assistance. This work was supported by KAKENHI (22134006, 23249019, and 26253020), the National Institutes of Health/National Institute of Allergy and Infectious Diseases (R21 AI090135), European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 2012-305121 "Integrated European –omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)" and French National Research Agency – Program on physiopathology of rare diseases (ANR-09-GENO-035).

Competing financial interests

The authors declare to have no competing financial interests.

Accession Codes

Sequence data have been deposited to the Leiden Open Variation Database (LOVD):
<http://databases.lovd.nl/shared/individuals/CDCA7>
<http://databases.lovd.nl/shared/individuals/HELLS>

Supplementary information

Fig. S1: Repeat hypomethylation in ICF patients from families E, F and G

Fig. S2: Alignment of the 4-CXXC domains of CDCA7 and CDCA7L

Fig. S3: Confirmation of minor satellite hypomethylation through siRNA mediated knockdown in wildtype MEFs

Fig. S4: Uncropped images of agarose gels and Southern blots corresponding to Figure 2g

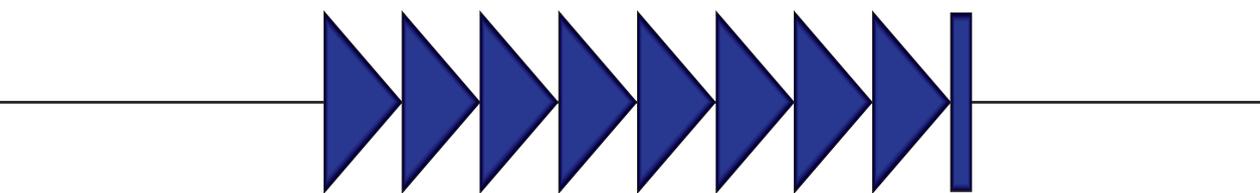
Fig. S5: Uncropped images of agarose gels and Southern blots corresponding to Supplementary Figure 3b

Table S1: Demographic and clinical characteristics of patients

Table S2: List of primers and siRNAs

All supplementary information belonging to this chapter can be accessed through <http://goo.gl/ynEKdX> or by using the QR-code below.





General Discussion

8



The studies described in this thesis aimed at better understanding the genetic and epigenetic contributors to two clinically unrelated, but epigenetically related diseases: the muscular dystrophy FSHD and the primary immunodeficiency ICF syndrome. Common to both disorders is the epigenetic dysregulation of repetitive DNA. FSHD, in most cases an *in cis* epigenetic disorder, is caused by chromatin relaxation of the D4Z4 macrosatellite repeat array in somatic cells and misexpression of the DUX4 transcription factor in skeletal muscle. In ICF syndrome, an *in trans* epigenetic disorder, the most prominent epigenetic characteristic is the loss of CpG methylation at (peri-) centromeric satellite repeats. How this causes pathology or contributes to the observed immunodeficiency is currently not known.

Repressing D4Z4: balancing repeat size and chromatin modifiers

In the majority of cases, FSHD is caused by contraction of the D4Z4 macrosatellite repeat on chromosome 4q35. Upon this contraction, a changed chromatin organization at the repeat leads to the derepression of the *DUX4* transcription factor in somatic cells. These two sentences summarize almost 20 years of scientific publications about the etiology of FSHD: from the linkage of FSHD to 4q35 and D4Z4 contraction in the early 1990's until the unifying disease mechanism in 2010¹⁻³. In between, key publications described the specific association of FSHD with the 4qA haplotype, the involvement of epigenetic dysregulation at D4Z4 and the transcriptional activity of the D4Z4 repeat⁴⁻¹¹. During this process, in absence of the proof for the unifying genetic mechanism of FSHD, alternative pathomechanisms were described, mainly involving proximal gene dysregulation¹²⁻¹⁴.

Recurring throughout the years, and important to all the proposed disease mechanisms, is the epigenetic dysregulation of D4Z4 in FSHD. The partial deletion of the D4Z4 repeat (FSHD1), or mutations in chromatin modifiers of D4Z4 (FSHD2), lead to a changed chromatin conformation at the repeat in somatic cells, which was proposed to spread or loop proximally and affect proximal gene regulation through an *in cis* mechanism. Involvement of the transcriptional activation of *FRG1* and *FRG2* was often studied, however never unequivocally proven to play a role in the FSHD disease mechanism¹²⁻¹⁸. The work in this thesis strengthens a role for epigenetic dysregulation at D4Z4 and derepression of *DUX4* as essential contributors to FSHD pathology. Earlier reports showed reduced binding of the H3K9me3-HP1-Cohesin network to the D4Z4 repeat in FSHD¹¹. Indeed, in **chapter 2** we firmly establish that there is a decreased chromatin compaction at D4Z4 (expressed as ChCS) in FSHD, either attributable to increased H3K4me2 or decreased H3K9me3 levels. However, our data in **chapter 4** could not confirm a causal role for H3K9me3 loss in the derepression of *DUX4* in a myogenic context. In patient derived myotubes, characterized by an increase in *DUX4* expression, the relative amount of H3K9me3:H3 at D4Z4 was similar to that observed in controls. Moreover, depletion of SUV39H1, shown to establish H3K9me3 at D4Z4 in HeLa cells and in immortalized human myoblasts^{11, 19}, or Cohesin subunits in control myotubes was not sufficient to activate *DUX4*. With this in mind, we attribute the decreased ChCS observed in **chapter 2** to an increased level of H3K4me2 in FSHD derived myoblasts, rather than a reduction in H3K9me3. This could simply reflect the increased expression

or “poised” state of the *DUX4* gene in FSHD and therefore the ChCS serves great purpose as a biomarker, but is uninformative in deciphering the epigenetic mechanism underlying *DUX4* derepression.

Chapter 4 highlights *SMCHD1* as the most potent known epigenetic regulator of D4Z4 to date as its ectopic expression reverses derepression of *DUX4* in both FSHD1 and FSHD2 derived myotubes. This is in line with genetic analyses showing 1) a causal role for *SMCHD1* mutations in FSHD2 and 2) a modifier effect of *SMCHD1* on disease severity in FSHD1²⁰⁻²². Depletion of *SMCHD1* in control myotubes mimics FSHD2 with *DUX4* becoming derepressed. In addition, we show in **chapter 4** that *SMCHD1* is indeed partially lost from D4Z4 upon its ectopic depletion and results in increased levels of PRC2 components and the PRC2 associated histone marker H3K27me3 at the D4Z4 repeat. This is also reflected in FSHD2 patient derived myotubes, but not in FSHD1 derived cells. Moreover, chemical inhibition of EZH2, the catalytic subunit of PRC2, leads to increased *DUX4* expression, but only in FSHD2 cells. These data indicate that, although highly similar, there are differences between FSHD1 and FSHD2 in the epigenetic regulation of D4Z4. Where in both forms of the disease *SMCHD1* is a repressor of the repeat, involvement of the PRC2 complex in D4Z4 regulation could only be detected in FSHD2.

A potential confounder in our ChIP-qPCR-based analyses is the selective involvement of the contracted D4Z4 repeat in FSHD1 and the epigenetic dysregulation of all four D4Z4 repeat arrays on chromosomes 4 and 10 in FSHD2. Our ChIP approach does not allow the selective analysis of the contracted allele in FSHD1 cells and therefore the vast majority of signal in ChIP-qPCR data generated in FSHD1 cells originates from the three (larger) non-affected alleles. Nonetheless, we are able to detect a decreased ChCS at D4Z4 in FSHD1 chromatin, supporting sensitivity of the assay. Moreover, with regard to the selective involvement of PRC2 in FSHD2, chemical inhibition of EZH2 should have still affected FSHD1 cells if increased levels of PRC2 at D4Z4 in FSHD1 would have been missed. The generation of isogenic cell lines in a D4Z4-free background, carrying a single 4qA type D4Z4 allele in both control and FSHD1 size range, can be used to further address this apparent difference between FSHD1 and FSHD2. To further study the *SMCHD1*-dependent enrichment of PRC2 complexes at D4Z4, the *SMCHD1* locus in these isogenic cell lines could be genetically engineered, e.g. through CRISPR-Cas9 genomic editing.

Alternatively, subtle sequence differences between 10q derived repeats and those derived from 4qA or 4qB could be exploited in patients carrying only one 4qA allele. A challenging aspect in this approach is the repetitive nature of D4Z4 and the technically challenging sequence composition of the distal end of the repeat. This could be overcome by a combination of single molecule real time sequencing (SMRT, PacBio) and massive parallel sequencing (Illumina HiSeq), provided that efficient D4Z4 enrichment strategies can be developed. The exact sequence composition of the four D4Z4 repeat arrays in a given sample can be determined through a combination of these approaches after which the smaller ChIP derived DNA sequence fragments can be superimposed on this. This allows to map the origin (which allele, or even which D4Z4 unit) of the

fragments enriched during the ChIP. This approach allows the fine mapping of known chromatin regulators and/or histone modifications at D4Z4.

Recently, CRISPR-Cas9 technology was employed to target both a transcriptional activator and repressor to the D4Z4 repeat, leading to increased and decreased DUX4 expression, respectively²³. This approach could be used to uncover proteins involved in regulating D4Z4 in a more unbiased (non-candidate driven) way. Upon targeting of an inactive Cas9 enzyme to the D4Z4 repeat, immunoprecipitation (IP) of the enzyme allows the enrichment of the D4Z4 chromatin template which can be subsequently analysed by proteomics and/or transcriptomics techniques. Together, these approaches can yield a more comprehensive and complete picture of the proteins and/or histone modifications present at D4Z4 and thereby facilitates the identification of new therapeutic targets and yet unknown FSHD2 disease genes²⁴.

In **chapter 2** we could not find a significant correlation between the derepression of D4Z4 and the age corrected clinical severity in FSHD patients, although a trend was observed in fibroblasts. In contrast, a correlation between CpG methylation levels at D4Z4 and clinical severity and/or penetrance has been demonstrated in multiple independent studies. In FSHD2 patients, the methylation level attributed to the shortest D4Z4 repeat of a single CpG in the proximal unit of D4Z4 (the FseI site) showed a significant correlation to disease severity²⁵. In two other reports, CpG methylation levels throughout the D4Z4 repeat, measured by bisulfite sequencing of specific domains within D4Z4, were shown to be indicative for disease penetrance. Non-affected carriers of an FSHD-sized repeat on a 4qA allele showed methylation levels comparable to control individuals, but are however more susceptible to ectopic *DUX4* derepression than controls^{26, 27}. These observations fit with other known genetic and epigenetic characteristics of D4Z4 in FSHD pathology. Firstly, FSHD1 shows the highest penetrance with residual repeat sizes of below 7 units and patients carrying only 1-3 residual repeat units are usually the most severely affected ones²⁸⁻³³, although a recent study showed high phenotypic variability in this patient group as well³⁴. Secondly, asymptomatic carriers of D4Z4 repeats between 7 and 10 units have a higher methylation level than expected based on repeat size²⁵. Thirdly, the size of the residual D4Z4 repeat positively correlates with the CpG methylation levels at single CpGs proximal to the repeat³⁵. Finally, FSHD2 patients typically carry D4Z4 repeats in the lower size range of controls.

Overall, a concept emerges that not the size of the repeat per se, but the ability of the muscle cell to maintain repression at D4Z4 determines clinical outcome. The main contributors to this repression are the size of the D4Z4 repeat and the activity and/or presence of chromatin repressors at D4Z4. FSHD1 is mainly a problem of repeat size, with an important contribution of chromatin repressors, whereas the opposite is true for FSHD2 (**Fig. 1**). In this respect it should be noted that the patients in which FSHD1 and FSHD2 co-exist carry residual D4Z4 repeat arrays of at least 6 units, as was the case for the FSHD1 cell lines studied in **chapter 4** in which we could rescue DUX4 expression by ectopic SMCHD1 expression. Conversely, mutations in these repressor genes are only harmful for individuals carrying D4Z4 repeat arrays in the lower control size range (<20

units) (Fig. 1). This indicates that the ability of SMCHD1, and/or other yet unknown modifiers, to repress the D4Z4 macrosatellite increases with the number of repeat units. It also suggests that to rescue the effect of repeat contraction, modifiers require a minimal amount of repeat units to exert their effect. The insufficient repression by e.g. SMCHD1 with lower repeat numbers could be caused by impaired functionality of the protein at shorter repeats or with inefficient binding to short repeats. The latter would be consistent with the observed preferential binding of SMCHD1 at longer telomeres compared to short telomeres³⁶. Although not experimentally proven, this model for penetrance and severity of FSHD must include the assumption that with decreasing capacity to repress D4Z4, the frequency of sporadic activation of *DUX4* and/or the expression levels upon activation increase, and that this leads an earlier onset and more progressive phenotype.

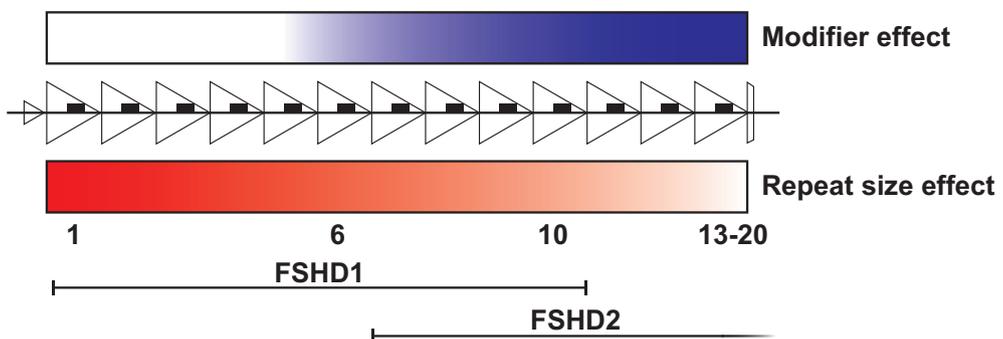


Figure 1: Schematic overview of the relative contribution of D4Z4 repeat size and chromatin modifiers in FSHD1 and FSHD2

Triangles represent D4Z4 repeat units. A D4Z4 repeat array of less than ten units is diagnostic for FSHD1, however there is a considerable effect of modifiers on repression of repeats of more than six units in length. Conversely, in the higher repeat size range (>20 units), the effect of mutations in modifiers like *SMCHD1* does not necessarily lead to pathology.

Polycomb repression at constitutive heterochromatin: targeted failsafe or random reshuffling?

There is a striking similarity between the changed chromatin organization of subtelomeres during cellular senescence and that of D4Z4 in FSHD2. In **chapter 3** we have shown that a decrease in subtelomeric CpG methylation (and H3K9me3) coincides with increased levels of H3K27me3. In **chapter 4** we report the same exchange of CpG methylation and H3K27me3 at D4Z4 in FSHD2 versus control derived cell lines. Moreover, increased H3K27me3 at D4Z4 was also observed in control cells upon depletion of SMCHD1.

These observations are in line with reports where inactivation of *Dnmt's*, or the chromatin remodeler *Hells*, in murine model systems leads to a similar anti-correlation between CpG methylation and H3K27me3³⁷⁻³⁹. Loss of methylation, through knocking out *Dnmt1*, at already lowly methylated CpG islands in majority leads to decreased levels of H3K27me3 at these promoters³⁸. In contrast, CpG poor regions are characterized by

increased presence of H3K27me3³⁷⁻³⁹. The mechanism behind this is unclear, but most likely relies on the affinity of PRC2 for unmethylated CpGs. The redistribution of this mark upon global loss of CpG methylation is proposed to be the effect of a random dilution though the increased genome wide abundance of unmethylated CpGs³⁸.

Based on the strong negative correlation between CpG methylation and gene expression, global disruption of CpG methylation patterns could be expected to greatly influence gene expression patterns. However, only minor changes in global gene expression were observed: upon loss of CpG methylation repressed genes remain repressed and transcribed genes remain transcribed^{37, 38}. Mainly genes which already display promoter hypomethylation were shown to be upregulated, correlating to decreased levels of H3K27me3 at these promoters³⁸. At regions which normally display high levels of CpG methylation, the increased levels of H3K27me3 maintains the transcriptionally repressed state. In **chapter 3** we indeed observed that the exchange of these marks did not have an effect on transcriptional activity of the loci under study (no CpG islands were included except the TERRA promoters). This supports the model delineated above: loss of CpG methylation correlates to the increase of H3K27me3 by which a repressive chromatin environment is maintained.

In **chapter 4** we observed an increase of H3K27me3 in FSHD2 derived cells at D4Z4 concomitant with hypomethylation of the repeat. In this case, however, the switch coincides with the sporadic derepression of *DUX4*. At first sight, this result opposes the model introduced above in which the loss of CpG methylation (and the associated repressive mechanisms) is compensated by increased Polycomb silencing. However, with the notion that only few cells escape silencing and show bursts of *DUX4* expression, the repression of D4Z4 by PRC2 and H3K27me3 still seems rather efficient. Moreover, an actual repressive effect of PRC2 at D4Z4 in FSHD2 is supported by chemical inhibition of EZH2, the catalytic subunit of PRC2, which led to increased levels of *DUX4* expression.

These results either suggest that, at D4Z4, CpG methylation reflects a more potent repressive mechanism than the PRC2 protein complex and its associated histone modifications. Alternatively, the proposed mechanism of PRC2 dilution may lead to a limited pool of PRC2 complexes which are “available” to repress D4Z4 upon loss of CpG methylation at the repeat. Both scenarios would fit with the occasional burst of *DUX4* expression in sporadic myonuclei of patients. However, only the latter would explain the discrepancy between FSHD1 and FSHD2 with regard to the involvement of PRC2 and H3K27me3 at the repeat. Most likely, changes in the chromatin organization in FSHD1 cells are confined to the contracted D4Z4 repeat. In contrast, mutations in *SMCHD1* in FSHD2, may have a much broader genome-wide effect on chromatin organization, as suggested by the genome-wide transcriptional dysregulation and the failure of X-inactivation in mouse *Smchd1* knock out cells⁴⁰⁻⁴². Although CpG hypomethylation at some other repeats was ruled out in FSHD2⁴³, studying the genome wide consequences of *SMCHD1* loss on CpG methylation and H3K27me3 may reveal an inverse correlation as was observed for *Dnmt's* and *Hells*.

Intriguingly, we could mimic the FSHD2 chromatin environment at D4Z4 by depletion

of SMCHD1 in control cells, however it is unclear whether this includes CpG hypomethylation at the repeat. By logic we can rule out extensive hypomethylation through passive demethylation: our experimental setup (depleting SMCHD1 during in vitro myogenic differentiation) strongly limits the number of cell divisions which would be needed for passive demethylation. Besides, *SMCHD1* has so far not been linked to active demethylation through the *TET* family of demethylases, so overall extensive hypomethylation is not expected at D4Z4 upon short term SMCHD1 depletion. Nonetheless, in this setup, depleting SMCHD1 does lead to the accumulation of PRC2 and H3K27me3 at D4Z4. Firstly, this suggests that the hypomethylated state of D4Z4 upon partial SMCHD1 loss in FSHD2 patients is determined during earlier stages of (muscle) development, or by prolonged absence of SMCHD1 from the repeat. Secondly, it suggests an order of events in which the absence of CpG methylation most likely leads to decreased SMCHD1 which in turn leads to increased PRC2 levels at D4Z4. This argues against a direct interaction between (absence of) CpG methylation and PRC2 recruitment and suggests the requirement of additional factors, including SMCHD1, to mediate the switch between CpG methylation and H3K27me3.

The technology to target CRISPR-Cas9 to the D4Z4 repeat described above offers new possibilities to mechanistically study the changes in the chromatin organization at D4Z4 which underlie *DUX4* activation. For example, tethering TET enzymes to D4Z4 through fusion with the inactive Cas9 enzyme may lead to local demethylation at the repeat and allows mapping of downstream molecular effects of CpG hypomethylation at D4Z4. Conversely, the effect of re-establishing CpG methylation at D4Z4 could be studied by targeting e.g. DNMT3B to the repeat in FSHD2 muscle cells. Using this approach in the mono-allelic 4qA D4Z4 cell lines discussed above would allow investigating whether similar mechanisms are active in FSHD1 as now described for FSHD2.

The mechanisms of proximal gene regulation revisited

Over the past decades, numerous studies have reported on potential deregulation of genes proximal to D4Z4 upon its contraction. Either deregulation was never unequivocally proven (*FRG1*), contribution to the FSHD pathomechanism was not clear (*FRG2*), or both (*FAT1*). Few studies have described a potential role for deregulation of *DUX4c* in the pathogenesis of FSHD. *DUX4c* is encoded by an inverted and incomplete D4Z4 repeat unit proximal of *FRG2*. It was reported to be deregulated in FSHD derived samples and to inhibit muscle cell differentiation by interfering with myogenic transcription factors like Myf5 and MyoD^{44,45}. However, in contrast to *DUX4*, its ectopic expression in *Xenopus laevis* did not reveal any developmental abnormalities⁴⁶. Moreover, since it is lost on 4q35 in patients with proximally extended deletions of D4Z4, a major contribution to FSHD pathology is unlikely⁴⁷. Detailed genetic analyses and the subsequent discovery of *DUX4* derepression, yet the most likely and best supported disease mechanism, further challenged the contribution of proximal genes to FSHD. The reported deregulation of both *FRG1* and *FRG2* was attributed to spreading of the derepression at D4Z4 and/or a changed higher order chromatin structure at 4q35. In **chapter 5** we have demonstrated that this is an unlikely explanation for the observed transcriptional upregulation of

FRG2. The combination of transcriptional analysis, ChIP analysis and Luciferase reporter assays revealed that *FRG2* is in fact a target gene of DUX4 and is as such activated during muscle cell differentiation in FSHD derived samples.

More recently, a similar study showed that the same mechanism results in the upregulation of *FRG1* in FSHD⁴⁸. The direct binding of DUX4 to the *FRG1* locus was again shown by ChIP and ectopic modulation of DUX4 levels support its role as transcriptional activator of *FRG1*. Intriguingly, whereas the DUX4 binding site at *FRG2* is upstream of the transcriptional start site (TSS), the DUX4 responsive peak in *FRG1* lies in intron 2 of the gene. At LTR repeats, but also at non repetitive targets like *ZSCAN4*, DUX4 has been shown to often create a new TSS at its exact binding site, leading to alternative transcripts from the adjacent target genes⁴⁹. Unfortunately the exact transcriptional consequences of DUX4 binding at *FRG1* and the nature of the DUX4-induced *FRG1* transcripts, as well as protein function, were not studied. Thus, both studies provide a central role for the DUX4 protein in the transcriptional activation of *FRG1* and *FRG2* in FSHD, however are not conclusive about the potential role for both genes in the pathology of FSHD.

So far, it is difficult to predict the possible role for DUX4 in regulating *FAT1*, the third gene upstream of D4Z4 which was implicated in FSHD pathology. A detailed study of the interaction between DUX4 and the *FAT1* locus will be pivotal to further evaluate an “D4Z4-independent” role for *FAT1* in the pathogenesis of FSHD. So far, FSHD samples were shown to have decreased *FAT1* expression levels, which in a mouse model was shown to affect muscle development^{50, 51}. Decreased *FAT1* expression could be the consequence of DUX4 expression, although the data are currently inconclusive. On the one hand, DUX4 was shown to decrease *FAT1* expression by at least twofold upon its overexpression in human myoblasts^{50, 52}. DUX4 ChIP analysis revealed a strong binding peak in intron 2 of *FAT1*, suggesting a direct effect of DUX4 binding on *FAT1*. Intriguingly, if the effect would indeed be direct, repression of *FAT1* would be in disagreement with the currently known function of DUX4 as transcriptional activator. In contrast, inhibition of DUX4 expression through shRNA expression did not affect *FAT1* expression levels, but resulted in decreased expression levels of well-established DUX4 target genes⁵¹. This discrepancy may be caused by the relatively easy way to detect activated target genes which are normally absent versus the diluted effect of restoring *FAT1* expression in only a minority of cells. A small number of individuals which do not display D4Z4 hypomethylation and/or D4Z4 contraction, but do show an FSHD-like phenotype were reported to carry possibly damaging *FAT1* variants⁵³. These variants were often polymorphisms which are present in the normal population with low frequency⁵³. DUX4 expression independent of D4Z4 contraction and/or hypomethylation should be analysed in these patients to exclude that the effect of *FAT1* deregulation is secondary to DUX4.

Altogether, FSHD specific deregulation of *FRG2* is now explained through DUX4 activity, while it is not inconceivable that DUX4 also affects *FRG1* and *FAT1*. All this is in support for a *in trans* effect of DUX4, as opposed to an *in cis* effects of the changed chromatin

organization at the D4Z4 repeat in FSHD. Additionally, strong genetic evidence supporting the *in trans* model came from the identification of an FSHD1 patient carrying a contracted D4Z4 repeat array on chromosome 10, of which the distal end consists of chromosome 4 derived D4Z4 units¹. If proximal gene deregulation would be mediated through an *in cis* effect, 10q26 genes would be affected. This was not studied, however if true it would argue against a causal role for *FAT1* and *FRG1* in the pathogenesis of FSHD as these genes are not present on 10q26. Although the deregulation of the upstream 4q35 genes is shown to be secondary to *DUX4* activation, a role in the pathology cannot be excluded and warrants further study.

Recent studies have shown an effect of telomere length on the regulation of 4q35 genes through a telomeric position effect (TPE). In mammals TPE is a poorly defined mechanism of chromatin spreading: telomeric heterochromatin leads to the silencing of genes in close proximity to the telomere and this silencing is lost upon progressive telomere shortening⁵⁴. Expression of *DUX4* and *FRG2*, and more recently *SORBS2*, was shown to be increased with decreasing telomere length in FSHD derived cells^{55, 56}. In **chapter 5** we have shown that the observed effect on *FRG2* is most likely secondary to increased *DUX4* expression as a result of telomere shortening. *SORBS2* is located ~4.5 Mb upstream of D4Z4, is expressed in skeletal muscle and so far described as a signal transduction and/or structural protein in cardiac muscle cells⁵⁵. The increase of *SORBS2* expression is mediated through a long range interaction of *SORBS2* with *FRG1*, which is lost upon extensive telomere shortening in FSHD cells only⁵⁵. Interestingly, several *DUX4* binding peaks were identified near or in *SORBS2*, however there is no other evidence for an *in trans* effect of *DUX4* on *SORBS2*, as observed for the previously described upstream genes. Overall, there may be a contribution of telomere shortening on the expression of 4q35 genes, including *DUX4* itself, however this effect is small in comparison to the effect of either D4Z4 contraction and/or mutations in modifiers like *SMCHD1*. For example, carriers of residual repeats of 1-3 units are often severely affected in early childhood where extensive telomere shortening is not yet expected. To better establish a role for TPE in FSHD, the results obtained in cellular models in which telomere length was ectopically modified should be translated to *in vivo* studies or large patient populations. This may prove difficult because the subtelomeric position of these genes may be primate specific and because of the large variation in telomere length in the general population.

SORBS2 is one of the four genes reported so far which are deregulated through disturbed long range interactions with extensive telomere shortening. This mechanism has been described as “TPE over long distances” (TPE-OLD) and differs from the classic TPE. In TPE-OLD, the 3D conformation of the genome rather than individual gene location determines the transcriptional response to extensive telomere shortening⁵⁷. It remains unclear at this point if *DUX4* transcription is affected in a similar mechanism or is affected through classic TPE. The independent observations that *SMCHD1* preferentially localizes at longer telomeres and that it is the major repressor of D4Z4 may suggest its involvement in TPE and/or TPE-OLD and might link telomere length to *DUX4* expression³⁶.

The final result of telomere shortening is a tumor suppressive mechanism called cellular senescence. Cells exit the cell cycle and over the recent years it has become clear that senescent cells undergo genome wide changes in chromatin organization. For example, recent reports showed genome wide redistribution of CpG methylation, H3K4me3 and H3K27me3 upon senescence^{58, 59}. Moreover, global nuclear organization is changed upon the formation of senescence associated heterochromatic foci containing large heterochromatic regions normally associated with the nuclear lamina^{60, 61}. Indeed, in **chapter 3** we observed decreased levels of subtelomeric CpG methylation, H3K9me3 and H4K16ac with a concomitant increase of H3K27me3 and H3K36me3 upon telomere induced senescence. Reading out *DUX4* transcript levels was not possible as these observations were done in fibroblasts without a 4qA allele. In the system used to identify targets of TPE-OLD, senescence signalling was excluded, however the changed 3D conformation of the genome upon TPE-OLD raises the possibility that both mechanisms are related. A careful overlap between genomic regions affected by TPE-OLD and cellular senescence can yield more insights into whether or not these are two independent mechanisms.

In any case, the possible activation of *DUX4* with decreased telomere length and/or senescence may form the molecular basis for the progressive nature of FSHD. With age, average telomere length declines variably but significantly. Moreover, there is build-up of senescent cells throughout aging tissues⁶². Both scenarios may lead to increased transcription of *DUX4* with age and would thereby contribute to the progressive nature and general late onset observed in FSHD. In this view, telomere length could be considered as a modifier of disease, but to support this claim larger studies in patient cohorts and even in specific tissues are necessary.

FSHD disease models: fighting against evolution

The generation of faithful animal models is currently indispensable for translational research. In other words: identifying and developing potential therapeutic interventions for FSHD requires the generation of an in vivo model for FSHD. In **chapter 6** we described the generation of two transgenic mouse models which recapitulate key genetic and epigenetic features of FSHD. The D4Z4-2.5 mouse carries an FSHD sized 4qA D4Z4 allele, which is characterized by relative chromatin derepression in somatic cells, as compared to the D4Z4-12.5 mouse line carrying a 4qA D4Z4 allele in the size range of control individuals. In agreement with the chromatin derepression, several (muscle) tissues and primary muscle cells derived from the D4Z4-2.5 mouse express detectable amounts of *DUX4*, whereas *DUX4* remains largely repressed in somatic tissues of the D4Z4-12.5 mouse. These data show that the mechanism of repeat length dependent D4Z4 repression is conserved between mouse and man even though the D4Z4 repeat is not present in the mouse genome.

The conservation of D4Z4 repression in our mouse models has great potential to study the epigenetic mechanisms involved in (de)repression of D4Z4. In **chapter 6** we observed that in somatic cells the D4Z4-2.5 mouse displays decreased levels of DNA methylation and a lower ChCS than the D4Z4-12.5 mouse. This is in good agreement

with the data we have obtained in human cell cultures in **chapters 2 and 4** and the published data on D4Z4 hypomethylation in FSHD^{5, 10, 25, 26, 63}. It would be interesting to see whether the observed negative correlation between DNA methylation and PRC2/H3K27me3 enrichment in FSHD2 is present at D4Z4 in the D4Z4-2.5 mouse as compared to the D4Z4-12.5 mouse. The absence of additional D4Z4 repeat (-like) sequences in both mouse lines eliminates the confounding effect of the additional (unaffected) repeat arrays in FSHD cells and may reveal whether the PRC2 enrichment is truly FSHD2 specific as shown in **chapter 4**. The limitations of this approach are that the transgenes in both lines have integrated at different sites in the genome and only one founder line of each is available. This creates the potential of interpreting data on the chromatin structure as an intrinsic property of D4Z4 whereas it actually could partially rely on the local chromatin environment at the site of integration. One way to overcome this problem, and confirm the intrinsic chromatin regulation of D4Z4, is to ectopically induce repeat contraction in the D4Z4-12.5 mouse line, for example by CRISPR-Cas9 technology. With this approach, the resulting mouse lines will be isogenic, but discordant for D4Z4 repeat length.

The D4Z4-12.5 mouse line carries a D4Z4 repeat in the size range of those observed in FSHD2 individuals. Crossbreeding this mouse line with a mouse line carrying mutations in *Smchd1* essentially generates offspring with an FSHD2 genotype. Moreover, the same approach in the D4Z4-2.5 mouse creates a genotype in which the genetic requirements of both FSHD1 and FSHD2 are met. In humans this genotype leads to an aggravated phenotype and led to the identification of *SMCHD1* as a disease modifier in FSHD1²². The generation of these models would yield more mechanistic insight in the effect of *SMCHD1* on the D4Z4 repeat, but also offers additional models to study therapeutic interventions aiming at *SMCHD1*. However, it first remains to be determined whether these mice display a similar interaction between *Smchd1* and D4Z4 as observed in humans.

Although genetic, epigenetic and transcriptional features of FSHD are recapitulated in our models, the D4Z4-2.5 mouse does not display a clinically relevant muscle phenotype. Approximately half of the D4Z4-2.5 mice do develop an abnormal eye phenotype of yet unclear etiology. This is in contrast to a published animal model relying on ectopic expression of DUX4. Intramuscular delivery of DUX4-expressing adeno-associated viruses (AAV) resulted in profound local muscle damage through the induction of p53 dependent apoptosis⁶⁴. Mice developed muscle weakness, but also showed quick recovery after these ectopic bursts of DUX4 expression. The expression of DUX4 in this model is localized to the site of injection and does not reflect the typical pattern of sporadic nuclei expressing DUX4. Moreover, the distinct regulatory mechanism including chromatin derepression of a repeat array is not recapitulated in this model⁶⁴. Similar detrimental effects of DUX4 expression, muscle specific or body wide, were observed in *Danio rerio*, highlighting the in vivo toxicity of DUX4^{64, 65}. Gross developmental muscle abnormalities and the muscle degeneration seen in *D. rerio* and AAV injected mouse muscle are not representing the muscle histology observed in FSHD patients⁶⁶, suggesting that the observed phenomena are generated through

species-specific molecular pathways.

More recently, a doxycycline inducible transgenic mouse model was generated in which the *DUX4* transgene was inserted in an euchromatic region of the mouse genome, upstream of the ubiquitously expressed *Hprt* gene on the X chromosome⁶⁷. Unexpectedly, leaky expression in non-doxycycline-treated animals led to early lethality in male mice only. Females displayed several non-muscular phenotypes, including a striped pattern of scaly skin, possibly as a consequence of random X-inactivation. Rarely surviving male carriers displayed runting, delayed muscle development and had a homogenous scale skin. *DUX4* was detectable in several tissues including testis, retina and brain, but was hardly detectable in muscle samples⁶⁷. In concordance with the D4Z4-2.5 mice described in **chapter 6**, cultured muscle cells showed sporadic bursts of *DUX4* expression. Strikingly, this mouse model confirms our observation of absence of an obvious muscular muscle phenotype upon systemic *DUX4* transgenesis. What is also in common is the detrimental effect of *DUX4* on the eye, which may be related to the observed extra-muscular phenotype of FSHD patients. Again, although low levels of *DUX4* expression can be tolerated in the mouse, the virtually absent muscle phenotype suggests different molecular consequences of *DUX4* expression between mouse and man.

The absence of a muscle phenotype in the two *DUX4* transgenic mouse models published to date may be a consequence of the evolutionary distance between mouse and man. *DUX4* has no orthologue in the mouse, however a paralogue, *Dux*, has been identified to be present in a tandem array in the mouse. *DUX4* in human cells has the propensity to bind repetitive elements, in particular specific subclasses of LTRs, which have expanded specifically in the primate genome. This correlates well with the presence of *DUX4/DUXC* in primates while absent in rodents. This suggests co-evolutionary events in which *DUX4*, normally expressed in germline cells, has a specifically evolved set of target genes to exert its yet unknown normal function. This selective pressure on *DUX4* binding sites is absent in the mouse. Therefore, the transcriptional consequences of *DUX4* in mouse cells have, are far less obvious than in human cells, as was reported in **chapter 6** and elsewhere⁶⁸. Consequently, the small overlap in transcriptional targets of *DUX4* between mouse and man is a plausible explanation for the absence of muscle phenotypes and the manifestation of abnormalities not seen in FSHD patient.

The recent observation that *FRG1* is a *DUX4* target gene in human cells, but not in the mouse, was suggested to explain the absence of a muscle phenotype in the D4Z4-2.5 mouse⁴⁸. Although it may have some relevance, the transcriptional activation of *FRG1* in response to *DUX4* expression in human cells is low compared to the minimum levels of *FRG1* associated with a muscle phenotype in the previously published transgenic *FRG1* mouse¹³. Generating the FSHD2 mouse model described above could reveal a possible role for the activation of human *FRG1* in recapitulating FSHD like muscle symptoms in the mouse, as the transgene in the D4Z4-12.5 mouse model includes *FRG1*. As *FRG1* is only one of the many target genes discordant between mouse and man, and it is currently not known which of the pathways deregulated in humans by *DUX4* are causal

to muscle pathology, it is difficult to predict biological consequence of the DUX4-FRG1 axis.

This discrepancy between mouse and man does not necessarily mean that the generated mouse lines lose their scientific and translational value. The D4Z4-2.5 and D4Z4-12.5 are promising tools to study epigenetic contributors to FSHD and, given a good measurable outcome, may become very useful to test therapeutic strategies aimed at inhibiting *DUX4* transcription. The doxycycline inducible transgenic *DUX4* mouse, as well as the viral delivery of *DUX4* in mouse muscle, allows for the induction of high levels of *DUX4* expression and can become valuable considering therapeutic strategies aiming at *DUX4* transcript and/or protein reduction.

Instead of expressing the *DUX4* transgene in mouse muscle, Zhang *et al.* described a completely different approach by xenografting human control and FSHD patient derived muscle biopsies into immunodeficient recipient mice⁶⁹. This approach is similar to our previous work using isogenic muscle cell clones discordant for D4Z4 repeat size⁷⁰. In both models, the human derived muscle cells succeeded to form new myofibers in the mouse muscle and specifically FSHD patient derived material was shown to express *DUX4*. Although the evolutionary distance between mouse and man is overcome in these models, they have their own limitations. First of all, in these models a contribution of the immune system to FSHD pathology cannot be analysed. More importantly, the xenografting procedure does not allow high-throughput studies and relies on the availability of donor material. Additionally, scoring performance of these animals in functional tests is uninformative with only single muscles being affected by the grafting. In conclusion, every conceivable disease model relying on *DUX4* expression in non-primate species will suffer from the primate specific evolution of *DUX4* and its targets, while more humanized models come with their own constraints. With these limitations in mind, the scientific community should adjust their demands and expectations regarding animal models for FSHD.

ICF syndrome: identifying the point of functional convergence of different disease genes

In **chapter 7** we have described the identification of two new disease genes underlying ICF syndrome. ICF syndrome is characterized by a triad of seemingly non-related phenotypes. Firstly, patients suffer from recurring infections of the gastro-intestinal and respiratory tracts due to a- or hypogammaglobulinemia, in the presence of B-cells. Secondly, cultured blood cells from patients display chromosomal instability at the centromeres of chromosomes 1, 9 and 16, which is correlated to CpG hypomethylation of centromeric repeats. Finally, almost all patients display a distinct but variable set of facial dysmorphisms, often including hypertelorism, flat nasal bridge and epicanthus. So far, mutations in four different genes - *DNMT3B*, *ZBTB24*, *CDCA7* and *HELLS* - have been identified to underlie the syndrome. Although the phenotype can vary between patients, it cannot be used to classify the patients into one of the genetic subtypes a priori⁷¹. The overlap between the different patients with regard to the three phenotypic features suggests that the four causative genes functionally converge at one or multiple

points during (B-cell-) development. Identification of these shared, similar or redundant pathways will be a key step to unravel the complex disease mechanism underlying ICF syndrome.

For two of the four genes, *DNMT3B* and *HELLS*, a clear functional connection has been previously reported. The chromatin remodelling activity of *HELLS* is required for proper functioning of the de novo methyltransferase *DNMT3B* during early development⁷². Loss of *Hells* in murine cells mainly leads to defects in CpG methylation at repetitive elements, but also affects a significant number of genic CpG sites^{37, 73-75}. These effects were reported to rely on the role of *HELLS* in enabling the proper establishment of CpG methylation through interacting with *DNMT3A* and *DNMT3B*, however the involvement of *HELLS* in maintaining CpG methylation is unclear yet. Arguing against a role for *Hells* in maintenance of CpG methylation is that the methylation status of episomal DNA can be maintained in absence of *HELLS*⁷⁶. In contrast, *HELLS* associates with late replicating DNA, interacts with *DNMT1* and *Hells* deficiency was initially reported to affect CpG methylation at an imprinted region, all implicating a role for *HELLS* in maintaining CpG methylation patterns⁷⁷⁻⁷⁹. In **chapter 7** we provide additional evidence for a role of *HELLS* in maintaining CpG methylation at murine satellite repeats. siRNA mediated knockdown of *HELLS* in mouse embryonic fibroblasts (MEFs), in which the establishment of DNA methylation is completed, led to decreased CpG methylation of satellite DNA. Considering the molecular function of *HELLS*, a chromatin remodeler necessary for *DNMT* functioning, it may not be surprising that it promotes both the establishment and maintenance of CpG methylation patterns.

The hypomethylated state of centromeric repeats is a hallmark of the disease and is shared by all patients. As for *Hells*, the data in **chapter 7** also revealed a role for *Zbtb24* and *Cdca7* in the maintenance of CpG methylation at centromeric minor satellites. It remains unclear at this point whether these genes are only involved in maintaining, or also plays a role in establishing CpG methylation at centromeric repeats. More importantly, the observed effects on minor satellite CpG methylation cannot be linked to any other known functional aspect of *ZBTB24* or *CDCA7*, mainly because no clear molecular function has been described for both genes.

By homology, *ZBTB24* belongs to a family of BTB-domain transcription factors of which some are involved in lymphocyte development⁸⁰. The presence of 8 tandem zinc finger domains suggests that *ZBTB24* has DNA binding capacity and can likely act as a transcription factor. However, we have recently discovered that *ZBTB24* promotes the repair of DNA double strand breaks during immunoglobulin class switching in B-cells and that this depends on its ZNF domain (unpublished observations). During this process, *ZBTB24* binds and stabilizes poly (ADP-ribose) chains on the DNA damage signaling protein poly (ADP-ribose) polymerase 1 (*PARP1*) and thereby promotes repair. Through identification of the molecular function of *ZBTB24* these data for the first time mechanistically explain the immunodeficiency in ICF2 syndrome. Given the high overlap in the immunological phenotype of all ICF patients, it can be anticipated that *DNMT3B*, *CDCA7* and *HELLS* converge with *ZBTB24* at some point during development of antibody

producing B-cells. Although our observations establish a plausible explanation for the immunodeficiency in ICF2, it is currently unclear how deficiency in *ZBTB24* would lead to the two other phenotypic hallmarks of ICF syndrome and how loss of its molecular function would result in CpG hypomethylation at (peri-) centromeric repeats.

So far, no molecular function for *CDCA7* has been established, although it has been implicated in several processes. *CDCA7* has been shown to interact with *Myc* and thereby plays a role in neoplastic transformation⁸¹. More recently, *Cdca7* was identified in a screen for Notch target genes involved in hematopoietic stem cell emergence⁸². *CDCA7* contains a 4-CXXC zinc finger domain, which is conserved but only shared with its close homologue *CDCA7L*. It has been shown that *CDCA7L* acts as a transcriptional repressor for monoamine oxidases (MAO) and that this likely depends on the DNA binding capacity of the 4-CXXC zinc finger⁸³. Based on the high conservation of this domain, a similar function could be expected of *CDCA7*, however the localization of *CDCA7L* to chromatin depends on its N-terminal p75 binding domain which is not present in *CDCA7*⁸⁴.

To address the etiology of ICF syndrome it will be of great interest to better characterize the molecular functions of *ZBTB24* and *CDCA7*. At the same time, it is pivotal to delineate how the absence of these functionalities, and that of *DNMT3B* and *HELLS*, leads to all phenotypic characteristics of ICF syndrome. One way to address these questions is by generating mouse models harboring ICF-like genotypes. Mouse models in which *Dnmt3b* or *Hells* were knocked out, were proven very useful to characterize the molecular function of both proteins, however early lethality in both models greatly impairs studying immunological and developmental features characteristic of ICF syndrome⁸⁵.⁸⁶. Moreover, the immunological phenotype observed in ICF mouse models, based on patient derived missense mutations in *Dnmt3b*, involves impaired T-cell function, rather than impaired B-cell function. Based on the genetics observed in ICF patients, a “classic” knockout model for *Zbtb24* should recapitulate ICF2. By using CRISPr-Cas9 technology, the identified missense mutations at conserved residues in *CDCA7* could be introduced into the mouse genome to model ICF3. To overcome (possible) lethality in mouse models for ICF syndrome, an alternative would be to isolate fetal liver cells from developing embryo’s deficient for one of the four genes and transplant those into immune compromised recipient mice to specifically follow the development of the immune cell repertoire⁸⁷.

In conclusion, the identification of four different genes to underlie ICF syndrome offers a great opportunity to better study the mechanisms underlying the disease. The shared phenotype between all patients calls for the identification of the molecular functions and especially the spatio-temporal expression of the disease genes. Only functions and/or (redundant) pathways shared by all four genes are likely to be disease causing. For example, it remains to be determined whether loss of *ZBTB24*, *CDCA7* and *HELLS* results in similar genome wide CpG hypomethylation as observed in ICF1 derived patient material⁸⁸. Overlapping these shared functions, pathways and phenotypes will likely result in shared defective pathways leading to disease and filter out effects mediated by

only a subset of the four genes which are unlikely to contribute largely to the disease mechanism.

FSHD and ICF syndrome: chromatin derepression at D4Z4 causing discordant phenotypes

The notion that there is the global defect in CpG methylation patterns in ICF1 patients, including CpG hypomethylation of the D4Z4 repeat, puts those patients at risk for developing FSHD. Indeed, while not extensively studied in ICF2-4 syndrome, ICF1 patients show a similar degree of D4Z4 hypomethylation as FSHD patients⁸⁹. Moreover, a recent report describes two families in which heterozygous mutations in *DNMT3B* segregate with D4Z4 hypomethylation and result in the development of FSHD2 or aggravation of FSHD1 (van den Boogaard et al., 2016). Although limited in the number of patients and families, these data confirm *DNMT3B* as a modifier of D4Z4 chromatin structure and, more importantly, as an FSHD2 disease gene, similar to what is observed for *SMCHD1*. Since heterozygous carriers of ICF1-like *DNMT3B* mutations are apparently at higher risk for developing FSHD, ICF patients and related heterozygous carriers should be carefully examined for the presence of FSHD associated symptoms, which was previously reported not to be the case⁹⁰. However, given the early onset and severity of ICF syndrome, the clinical signs of FSHD, usually with later onset, may have been missed. Conversely, it will be important to evaluate potential comorbidities in FSHD2 families with *DNMT3B* mutations, as the CpG hypomethylation may not be confined to the D4Z4 repeat array. This was suggested in the study of Van den Boogaard et al., as some individuals showed CpG hypomethylation of other repetitive elements in the genome. In concordance with a normal immune-phenotype in ICF1 carriers, this study did not reveal (subclinical) immunological defects in *DNMT3B* carriers from these FSHD2 families. The effect of mutations in *ZBTB24*, *CDCA7* or *HELLS* on genome wide, and more specifically D4Z4, chromatin organization is currently unclear. With similar approaches described in **chapter 4**, the involvement of these genes in D4Z4 repression could be studied to mechanistically support co-occurrence of FSHD and ICF syndrome.

Concluding remarks

The work described in this thesis highlights the relevance of chromatin organization at repetitive elements for human health. FSHD and ICF syndrome, two clinically unrelated diseases, are hallmarked by a loss of repression at specific repetitive elements. In FSHD, revolving around derepression of the D4Z4 repeat, this is primarily an *in cis* effect of partial deletion of the repeat. In ICF syndrome, (peri-) centromeric satellite repeats are derepressed and destabilized through an *in trans* mechanism mediated by at least four different genes. Our work, together with published observations, shows that repression of repetitive DNA elements is evolutionary conserved between mouse and man, although the phenotypic outcome of improper regulation of these elements can be different. Studying the commonalities and discrepancies between the different diseases and underlying genetic mutations can reveal common and specialized mechanisms to maintain repression of repetitive DNA in mammals.

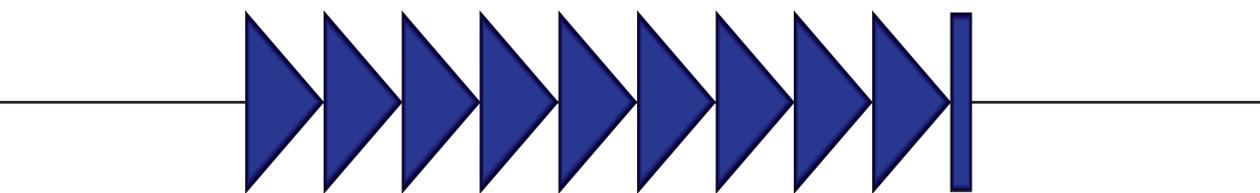
References

1. Lemmers,R.J. et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650-1653 (2010).
2. Wijmenga,C. et al. Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. *Genomics* 9, 570-575 (1991).
3. Wijmenga,C. et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* 2, 26-30 (1992).
4. Bodega,B. et al. Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FHD) locus and upregulation of FHD-related gene 1 (FRG1) expression during human myogenic differentiation. *BMC. Biol.* 7, 41 (2009).
5. de Greef,J.C. et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FHD. *Hum. Mutat.* 30, 1449-1459 (2009).
6. Lemmers,R.J. et al. Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 75, 1124-1130 (2004).
7. Lemmers,R.J. et al. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am. J Hum. Genet* 81, 884-894 (2007).
8. Snider,L. et al. RNA Transcripts, miRNA-sized Fragments, and Proteins Produced from D4Z4 Units: New Candidates for the Pathophysiology of Facioscapulohumeral Dystrophy. *Hum. Mol. Genet* 18, 2414-2430 (2009).
9. Snider,L. et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS. Genet.* 6, e1001181 (2010).
10. van Overveld,P.G. et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet* 35, 315-317 (2003).
11. Zeng,W. et al. Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats in facioscapulohumeral dystrophy (FHD). *PLoS. Genet* 7, e1000559 (2009).
12. Gabellini,D., Green,M., & Tupler,R. Inappropriate Gene Activation in FHD. A Repressor Complex Binds a Chromosomal Repeat Deleted in Dystrophic Muscle. *Cell* 110, 339-248 (2002).
13. Gabellini,D. et al. Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature* 439, 973-977 (2006).
14. Rijkers,T. et al. FRG2, an FHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FHD patients. *J Med Genet* 41, 826-836 (2004).
15. Jiang,G. et al. Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum Mol. Genet* 12, 2909-2921 (2003).
16. Klooster,R. et al. Comprehensive expression analysis of FHD candidate genes at the mRNA and protein level. *Eur. J. Hum. Genet.* 17, 1615-1624 (2009).
17. Winokur,S.T. et al. 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 Res* 2, 225-234 (1994).
18. Yao,Z. et al. DUX4-induced gene expression is the major molecular signature in FHD skeletal muscle. *Hum. Mol. Genet.*(2014).
19. Zeng,W. et al. Genetic and Epigenetic Characteristics of FHD-Associated 4q and 10q D4Z4 that are Distinct from Non-4q/10q D4Z4 Homologs. *Hum. Mutat.* 35, 998-1010 (2014).
20. Larsen,M. et al. Diagnostic approach for FHD revisited: SMCHD1 mutations cause FHD2 and act as modifiers of disease severity in FHD1. *Eur. J. Hum. Genet.* 23, 808-816 (2015).
21. Lemmers,R.J. et al. Digenic inheritance of an SMCHD1 mutation and an FHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* 44, 1370-1374 (2012).
22. Sacconi,S. et al. The FHD2 gene SMCHD1 is a modifier of disease severity in families affected by FHD1. *Am. J. Hum. Genet.* 93, 744-751 (2013).
23. Himeda,C.L., Jones,T.I., & Jones,P.L. CRISPR/dCas9-mediated Transcriptional Inhibition Ameliorates the Epigenetic Dysregulation at D4Z4 and Represses DUX4-fl in FSH Muscular Dystrophy. *Mol. Ther.*(2015).
24. Fujita,T. & Fujii,H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation

- (enChIP) using CRISPR. *Biochemical and Biophysical Research Communications* 439, 132-136 (2013).
25. Lemmers,R.J. et al. Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. *Hum. Mol. Genet.* 24, 659-669 (2015).
 26. Gaillard,M.C. et al. Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. *Neurology* 83, 733-742 (2014).
 27. Jones,T.I. et al. Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. *Clin. Epigenetics.* 7, 37 (2015).
 28. Goto,K., Nishino,I., & Hayashi,Y.K. Very low penetrance in 85 Japanese families with facioscapulohumeral muscular dystrophy 1A. *J Med Genet* 41, e12 (2004).
 29. Lunt,P.W. et al. Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). *Hum Mol Genet* 4, 951-958 (1995).
 30. Ricci,G. et al. Large scale genotype-phenotype analyses indicate that novel prognostic tools are required for families with facioscapulohumeral muscular dystrophy. *Brain* 136, 3408-3417 (2013).
 31. Statland,J.M. et al. Milder phenotype in facioscapulohumeral dystrophy with 7-10 residual D4Z4 repeats. *Neurology*(2015).
 32. Tawil,R. et al. Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. *Ann Neurol* 39, 744-748 (1996).
 33. Tonini,M.M.O. et al. Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). *Neuromuscular Disorders* 14, 33-38 (2004).
 34. Nikolic,A. et al. Clinical expression of facioscapulohumeral muscular dystrophy in carriers of 1-3 D4Z4 reduced alleles: experience of the FSHD Italian National Registry. *BMJ Open.* 6, e007798 (2016).
 35. van Overveld,P.G. et al. Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. *Ann. Neurol.* 58, 569-576 (2005).
 36. Grolimund,L. et al. A quantitative telomeric chromatin isolation protocol identifies different telomeric states. *Nat. Commun.* 4, 2848 (2013).
 37. Yu,W. et al. Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. *Genome Res.* 24, 1613-1623 (2014).
 38. Reddington,J.P. et al. Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol.* 14, R25 (2013).
 39. Brinkman,A.B. et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* 22, 1128-1138 (2012).
 40. Blewitt,M.E. et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* 40, 663-669 (2008).
 41. Gendrel,A.V. et al. Epigenetic functions of smchd1 repress gene clusters on the inactive x chromosome and on autosomes. *Mol. Cell Biol.* 33, 3150-3165 (2013).
 42. Mould,A.W. et al. Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics. Chromatin.* 6, 19 (2013).
 43. de Greef,J.C. et al. Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. *Neurology* 69, 1018-1026 (2007).
 44. Anseau,E. et al. DUX4c is up-regulated in FSHD. It induces the MYF5 protein and human myoblast proliferation. *PLoS. ONE.* 4, e7482 (2009).
 45. Bosnakovski,D. et al. DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. *Exp. Neurol.* 214, 87-96 (2008).
 46. Wuebbles,R.D., Long,S.W., Hanel,M.L., & Jones,P.L. Testing the effects of FSHD candidate gene expression in vertebrate muscle development. *Int. J. Clin. Exp. Pathol.* 3, 386-400 (2010).
 47. Lemmers,R.J. et al. D4F104S1 deletion in facioscapulohumeral muscular dystrophy: Phenotype, size, and detection. *Neurology* 61, 178-183 (2003).
 48. Ferri,G., Huichalaf,C.H., Caccia,R., & Gabellini,D. Direct interplay between two candidate genes in FSHD muscular dystrophy. *Hum. Mol. Genet.* 24, 1256-1266 (2015).
 49. Young,J.M. et al. DUX4 binding to retroelements creates promoters that are active in FSHD muscle

- and testis. *PLoS. Genet.* 9, e1003947 (2013).
50. Caruso, N. et al. Deregulation of the protocadherin gene FAT1 alters muscle shapes: implications for the pathogenesis of facioscapulohumeral dystrophy. *PLoS. Genet.* 9, e1003550 (2013).
 51. Mariot, V. et al. Correlation between low FAT1 expression and early affected muscle in facioscapulohumeral muscular dystrophy. *Ann Neurol.* 78, 387-400 (2015).
 52. Geng, L.N. et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38-51 (2012).
 53. Puppo, F. et al. Identification of variants in the 4q35 gene FAT1 in patients with a facioscapulohumeral dystrophy-like phenotype. *Hum. Mutat.* 36, 443-453 (2015).
 54. Baur, J.A., Zou, Y., Shay, J.W., & Wright, W.E. Telomere position effect in human cells. *Science* 292, 2075-2077 (2001).
 55. Robin, J.D. et al. SORBS2 transcription is activated by telomere position effect-over long distance upon telomere shortening in muscle cells from patients with facioscapulohumeral dystrophy. *Genome Res.* 25, 1781-1790 (2015).
 56. Stadler, G. et al. Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat. Struct. Mol. Biol.* 20, 671-678 (2013).
 57. Robin, J.D. et al. Telomere position effect: regulation of gene expression with progressive telomere shortening over long distances. *Genes Dev.* 28, 2464-2476 (2014).
 58. Hanzelmann, S. et al. Replicative senescence is associated with nuclear reorganization and with DNA methylation at specific transcription factor binding sites. *Clin. Epigenetics.* 7, 19 (2015).
 59. Shah, P.P. et al. Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev.* 27, 1787-1799 (2013).
 60. Chandra, T. et al. Independence of repressive histone marks and chromatin compaction during senescent heterochromatic layer formation. *Mol. Cell* 47, 203-214 (2012).
 61. Chandra, T. et al. Global reorganization of the nuclear landscape in senescent cells. *Cell Rep.* 10, 471-483 (2015).
 62. van Deursen, J.M. The role of senescent cells in ageing. *Nature* 509, 439-446 (2014).
 63. Hartweck, L.M. et al. A focal domain of extreme demethylation within D4Z4 in FSHD2. *Neurology* 80, 392-399 (2013).
 64. Wallace, L.M. et al. DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann. Neurol.* 69, 540-552 (2011).
 65. Mitsuhashi, H., Mitsuhashi, S., Lynn-Jones, T., Kawahara, G., & Kunkel, L.M. Expression of DUX4 in zebrafish development recapitulates facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* (2012).
 66. Statland, J. & Tawil, R. Facioscapulohumeral muscular dystrophy. *Neurol. Clin.* 32, 721-8, ix (2014).
 67. Dandapat, A. et al. Dominant lethal pathologies in male mice engineered to contain an X-linked DUX4 transgene. *Cell Rep.* 8, 1484-1496 (2014).
 68. Sharma, V., Harafuji, N., Belayew, A., & Chen, Y.W. DUX4 differentially regulates transcriptomes of human rhabdomyosarcoma and mouse C2C12 cells. *PLoS. ONE.* 8, e64691 (2013).
 69. Zhang, Y. et al. Human skeletal muscle xenograft as a new preclinical model for muscle disorders. *Hum. Mol. Genet.* 23, 3180-3188 (2014).
 70. Krom, Y.D. et al. Generation of isogenic D4Z4 contracted and noncontracted immortal muscle cell clones from a mosaic patient: a cellular model for FSHD. *Am. J. Pathol.* 181, 1387-1401 (2012).
 71. Weemaes, C.M. et al. Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. *Eur. J. Hum. Genet.* 21, 1219-1225 (2013).
 72. Ren, J. et al. The ATP binding site of the chromatin remodeling homolog Lsh is required for nucleosome density and de novo DNA methylation at repeat sequences. *Nucleic Acids Res.* 43, 1444-1455 (2015).
 73. Myant, K. et al. LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Res.* 21, 83-94 (2011).
 74. Tao, Y. et al. Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. *Proc. Natl. Acad. Sci. U. S. A* 108, 5626-5631 (2011).
 75. Xi, S. et al. Lsh controls Hox gene silencing during development. *Proc. Natl. Acad. Sci. U. S. A* 104, 14366-14371 (2007).

76. Zhu,H. et al. Lsh is involved in de novo methylation of DNA. *EMBO J.* 25, 335-345 (2006).
77. Dennis,K., Fan,T., Geiman,T., Yan,Q., & Muegge,K. Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* 15, 2940-2944 (2001).
78. Myant,K. & Stancheva,I. LSH cooperates with DNA methyltransferases to repress transcription. *Mol. Cell Biol.* 28, 215-226 (2008).
79. Yan,Q., Cho,E., Lockett,S., & Muegge,K. Association of Lsh, a regulator of DNA methylation, with pericentromeric heterochromatin is dependent on intact heterochromatin. *Mol. Cell Biol.* 23, 8416-8428 (2003).
80. Lee,S.U. & Maeda,T. POK/ZBTB proteins: an emerging family of proteins that regulate lymphoid development and function. *Immunol. Rev.* 247, 107-119 (2012).
81. Gill,R.M., Gabor,T.V., Couzens,A.L., & Scheid,M.P. The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation. *Mol. Cell Biol.* 33, 498-513 (2013).
82. Guiu,J. et al. Identification of Cdca7 as a novel Notch transcriptional target involved in hematopoietic stem cell emergence. *J. Exp. Med.*(2014).
83. Chen,K., Ou,X.M., Wu,J.B., & Shih,J.C. Transcription factor E2F-associated phosphoprotein (EAPP), RAM2/CDCA7L/JPO2 (R1), and simian virus 40 promoter factor 1 (Sp1) cooperatively regulate glucocorticoid activation of monoamine oxidase B. *Mol. Pharmacol.* 79, 308-317 (2011).
84. Maertens,G.N., Cherepanov,P., & Engelman,A. Transcriptional co-activator p75 binds and tethers the Myc-interacting protein JPO2 to chromatin. *J. Cell Sci.* 119, 2563-2571 (2006).
85. Geiman,T.M. & Muegge,K. Lsh, an SNF2/helicase family member, is required for proliferation of mature T lymphocytes. *Proc. Natl. Acad. Sci. U. S. A* 97, 4772-4777 (2000).
86. Ueda,Y. et al. Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. *Development* 133, 1183-1192 (2006).
87. Gudmundsson,K.O., Stull,S.W., & Keller,J.R. Transplantation of mouse fetal liver cells for analyzing the function of hematopoietic stem and progenitor cells. *Methods Mol. Biol.* 879, 123-133 (2012).
88. Heyn,H. et al. Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. *Epigenetics.* 7, 542-550 (2012).
89. Kondo,T. et al. Whole-genome methylation scan in ICF syndrome: hypomethylation of non-satellite DNA repeats D4Z4 and NBL2. *Hum Mol Genet* 9, 597-604 (2000).
90. van der Maarel,S.M. & Frants,R.R. The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 76, 375-386 (2005).



Appendices

A



Summary

The repressive chromatin organization of repetitive DNA is pivotal for maintaining cell homeostasis. Defects in repeat silencing are a hallmark of the two genetic diseases studied in this thesis: FSHD and ICF syndrome. In FSHD, the D4Z4 macrosatellite repeat gets derepressed through an *in cis* or *in trans* mechanism, leading to sporadic expression of the toxic transcription factor *DUX4*. ICF is a severe immunodeficiency syndrome, characterized by derepression of (peri-) centromeric satellite repeats and is genetically heterogeneous. Mutations in multiple genes lead to a highly similar phenotype, which suggests these genes to act in common or converging pathways. The studies described in this thesis aimed at better understanding the genetic and epigenetic mechanisms underlying these diseases.

In **chapter 2** we have studied a possible correlation between D4Z4 chromatin compaction in somatic cells, measured by relative abundance of H3K9me3 over H3K4me2 at the D4Z4 repeat, and disease severity. Overall, FSHD derived somatic cells are characterized by a reduced chromatin compaction at D4Z4, which is especially pronounced in FSHD2 cells. Although trending in fibroblasts, we did not find a significant correlation between the chromatin compaction of D4Z4 and the age corrected clinical severity. Strikingly, the muscle pathology score in the *vastus lateralis* muscle showed a clear correlation with clinical severity score.

Telomere position effects (TPE) could affect the regulation of the subtelomeric D4Z4 repeat, but we could not investigate the 4q subtelomere directly due to its repetitive nature. Therefore, in **chapter 3** we have studied the chromatin regulation at two single copy subtelomeres (7q and 11q), as a model for subtelomeres in general, in response to telomere shortening and cellular senescence. Subtelomeres are characterized by a loss of H3K9me3 and CpG methylation upon telomere shortening and senescence. In contrast, we did not detect transcriptional activation of subtelomeric loci at the transcriptional and chromatin level. Silencing of the studied subtelomeres is most likely maintained through increased levels of markers for facultative heterochromatin, mediated by PRC2. Both subtelomeres under study, as well as D4Z4, showed similar regulation, with the exception of TERRA promoters. Overall, the 7q TERRA promoter showed similar regulation to the more proximal subtelomere, whereas the 11q promoter was distinct from its more proximal subtelomere, in concordance with the observed difference in TERRA transcription from 7q and 11q.

Having established the generic principles of subtelomeric chromatin structure in human cells, in **chapter 4** we focused on the chromatin regulation of D4Z4 by SMCHD1 in control and FSHD derived myotubes. This study established a pivotal role for *SMCHD1* in both genetic forms of the disease. In FSHD cells, the activation of *DUX4* during *in vitro* muscle cell differentiation coincides with a decreased level of SMCHD1 at D4Z4. More importantly, moderate overexpression of SMCHD1 resulted in silencing of *DUX4* in both FSHD1 and FSHD2 cells. Ectopic inhibition of SMCHD1 expression in control cells leads to the accumulation of PRC2 and the PRC2 mediated H3K27me3 marker for facultative heterochromatin at D4Z4. Increased abundance of PRC2 and H3K27me3 at

D4Z4, and sensitivity to PRC2 inhibition, was also observed in FSHD2 patient derived myotubes, however not in FSHD1 myotubes. In conclusion, although SMCHD1 is able to silence DUX4 in both FSHD1 and FSHD2 derived myotubes, the epigenetic regulation of the D4Z4 repeat in both forms of the disease seem to be distinguished by aberrant Polycomb mediated chromatin regulation, which is only observed in FSHD2.

Over the past decades, the absence of evidence for sporadic *DUX4* activation resulted in numerous studies focusing on proximal gene dysregulation upon D4Z4 contraction as a potential disease mechanism. Models were presented in which D4Z4 could act as an insulator element to prevent TPE spreading in proximal direction, or in which the chromatin structure of D4Z4 itself could affect the transcriptional regulation of nearby genes through spreading or looping mechanisms. In **chapter 5** we show that the FSHD specific dysregulation of *FRG2*, the gene closest to the D4Z4 repeat array, is a direct consequence of DUX4 protein activity, rather than any of the above proposed models. This further strengthens the disease mechanism revolving around *DUX4* while at the same time giving a molecular explanation for a longstanding enigmatic observation of *FRG2* activation in FSHD patients.

In **chapter 6** we show that the epigenetic regulation of D4Z4 and its derepression upon contraction can be recapitulated in the mouse, independent from its subtelomeric location. We have generated a transgenic FSHD mouse model, carrying an FSHD1 sized D4Z4 repeat array of 2.5 units (D4Z4-2.5), and a transgenic “control” mouse carrying a 12.5 unit containing D4Z4 repeat (D4Z4-12.5). Our data shows that the D4Z4-2.5 mouse model, as compared to the D4Z4-12.5 mouse, recapitulates key aspects of FSHD: robust *DUX4* expression in the germ line, chromatin derepression of the repeat in somatic cells accompanied by sporadic *DUX4* activation in skeletal muscle and activation of (mouse specific) DUX4 target genes. Intriguingly, the D4Z4-2.5 mouse does not develop an overt muscle phenotype, suggesting that the consequences of DUX4 expression are not conserved between man and mouse. Nevertheless, our model can serve great purpose to assess therapeutic interventions aiming at inhibiting the mis-expression of *DUX4* in FSHD.

Finally, in **chapter 7** we describe the identification of two new ICF syndrome disease genes, *CDCA7* (ICF3) and *HELLS* (ICF4). Herewith we identify two novel chromatin modifiers of repetitive DNA structures in humans, and therefore potentially important factors for the regulation of D4Z4 and *DUX4*. For both newly identified genes, as well as for the ICF2 gene *ZBTB24*, we show a role in the maintenance of CpG methylation at centromeric repeats. The identification of two additional ICF disease genes renders new possibilities for mechanistic studies into the disease mechanism behind ICF syndrome, and to study the epigenetic regulation of (macrosatellite) repeats in general.

The progress made in understanding the (epi-)genetic mechanisms underlying FSHD and ICF syndrome offers new possibilities for translational and mechanistic studies focusing on chromatin organization of repetitive DNA. The common feature of repeat derepression in both diseases offers potential to study commonalities and differences of different mechanisms affecting repeat silencing during development and differentiation.

Samenvatting

De repressieve chromatine organisatie van DNA repeats is cruciaal voor het behoud van cellulaire homeostase. Defecten in de repressie van repeats zijn een kenmerk van de twee genetische ziekten bestudeerd in dit proefschrift: FSHD en ICF-syndroom. In FSHD is de repressie van de D4Z4 macrosatelliet repeat onvoldoende als gevolg van een *in cis* of *in trans*-mechanisme, wat leidt tot sporadische expressie van de toxische transcriptiefactor *DUX4*. ICF is een ernstig immunodeficiëntie syndroom, gekenmerkt door onvolledige repressie van (peri)-centromere satelliet repeats en is genetisch heterogeen. Mutaties in meerdere genen leiden tot een zeer vergelijkbaar fenotype, wat suggereert dat deze genen een rol spelen in gemeenschappelijke of convergerende moleculaire mechanismen. De studies beschreven in dit proefschrift hadden als doel een beter inzicht te krijgen in de genetische en epigenetische mechanismen die ten grondslag liggen aan deze ziekten.

In **hoofdstuk 2** hebben we een mogelijke correlatie tussen D4Z4 chromatine organisatie in somatische FSHD cellen, gemeten door de verhouding tussen H3K9me3 en H3K4me2 op de D4Z4 repeat, en de klinische ernst van de ziekte bestudeerd. In het algemeen, worden somatische cellen verkregen uit FSHD patiënten gekenmerkt door een verminderde chromatine compactheid op D4Z4, voornamelijk in FSHD2 cellen. Hoewel er een trend zichtbaar is in fibroblasten, hebben we geen significante correlatie kunnen vinden tussen de chromatine compactheid van D4Z4 en de klinische ernst gecorrigeerd voor leeftijd. Opvallend is dat de pathologie van de *vastus lateralis* spier significant correleert met de klinische ernst.

Telomere positie effecten (TPE) kunnen invloed hebben op de regulering van de subtelomere D4Z4 repeat, maar het repetitieve karakter van het 4q subtelomeer hinderde de analyse van dit locus. In **hoofdstuk 3** hebben we daarom de chromatine regulatie op twee niet repetitieve en unieke subtelomere gebieden (7q en 11q) bestudeerd, als model voor subtelomere gebieden in het algemeen, in reactie op telomeerverkorting en cellulaire veroudering. Tijdens telomeerverkorting en cellulaire veroudering worden subtelomere gebieden gekenmerkt door een verlies van H3K9me3 en CpG methylering. Noch op chromatine regulatie als op transcriptieel niveau, konden we activatie detecteren op deze loci. Repressie van de bestudeerde subtelomere gebieden wordt waarschijnlijk gehandhaafd door middel van verhoogde niveaus van markers voor facultatief heterochromatine, gemedieerd door PRC2. Beide bestudeerde subtelomere gebieden, evenals D4Z4, toonde vergelijkbare regulatie, met uitzondering van de TERRA promotoren. Over het geheel genomen, vertoonde de 7Q TERRA promotor vergelijkbare regulatie met de meer proximale subtelomere loci, terwijl de 11q promotor verschilde van de meer proximale subtelomere loci, in overeenstemming met het waargenomen verschil in TERRA transcriptie van 7q en 11q.

Na de generieke principes van subtelomere chromatine structuur in menselijke cellen te hebben vastgesteld, hebben we ons in **hoofdstuk 4** gericht op de regulering van het D4Z4 chromatine door *SMCHD1*, in gedifferentieerde spiercellen van zowel controles als FSHD patiënten. Deze studie stelde een centrale rol vast voor *SMCHD1* in beide

genetische vormen van de ziekte. In FSHD cellen valt de activatie van *DUX4* tijdens in vitro spierceldifferentiatie samen met een verlaagd niveau van SMCHD1 op D4Z4. Nog belangrijker is dat matige overexpressie van SMCHD1 resulteerde in repressie van *DUX4* in zowel FSHD1 en FSHD2 cellen. Ectopische depletie van SMCHD1 expressie in controlecellen leidt tot de accumulatie van PRC2 en H3K27me3, markers voor facultatief heterochromatine, op D4Z4. Verhoogde niveaus van PRC2 en H3K27me3 op D4Z4, en gevoeligheid voor PRC2 inhibitie, werden ook waargenomen in gedifferentieerde spiercellen van FSHD2 patiënten, echter niet in FSHD1 cellen. Dus, hoewel SMCHD1 in staat is om expressie van *DUX4* te onderdrukken in gedifferentieerde spiercellen van zowel FSHD1 en FSHD2 patiënten, is de epigenetische regulatie van de D4Z4 repeat tussen beide vormen van de ziekte verschillend door afwijkende Polycomb gemedieerde chromatine regulatie, uniek voor FSHD2.

In de afgelopen decennia, resulteerde het gebrek aan bewijs voor sporadische *DUX4* activatie in tal van studies, gericht op ontregeling van proximale genen ten gevolge van D4Z4 contractie als een potentieel ziektemechanisme. In deze modellen zou D4Z4 fungeren als een buffer, ter voorkoming van proximale spreiding van TPE, of de chromatine structuur van D4Z4 zelf zou een rol hebben in de transcriptionele regulatie van de nabijgelegen genen door positie effecten of formatie van chromatine loops. In **hoofdstuk 5** laten we zien dat de FSHD specifieke ontregeling van het *FRG2* gen, het dichtstbij t.o.v. de D4Z4 repeat, een direct gevolg is van *DUX4* eiwitactiviteit, in plaats van de hierboven voorgestelde modellen. Dit herbevestigt de centrale rol van *DUX4* in het FSHD ziektemechanisme en tegelijkertijd geeft het een moleculaire verklaring voor de raadselachtige en lang bestaande observatie van *FRG2* activatie in FSHD patiënten.

In **hoofdstuk 6** laten we zien dat de epigenetische regulatie van D4Z4 en de derepressie na contractie kunnen worden gerecapituleerd in de muis, onafhankelijk van de subtelomere locatie. We hebben een transgeen FSHD muismodel gegenereerd, deze muis draagt een D4Z4 repeat van 2,5 units (FSHD lengte, D4Z4-2,5), en een transgene "controle" muis met een D4Z4 repeat van 12,5 units (controle lengte, D4Z4-12,5). Onze data tonen aan dat de D4Z4-2,5 muis, in vergelijking met de D4Z4-12,5 muis, de belangrijkste aspecten van FSHD recapituleert: robuuste *DUX4* expressie in kiembaancellen, chromatine derepressie van de repeat in somatische cellen, sporadische *DUX4* activatie in de skeletspier en activatie van (muis-specifieke) *DUX4* targetgenen. De D4Z4-2,5 muizen ontwikkelen opmerkelijk genoeg geen duidelijk spierfenotype, wat suggereert dat de gevolgen van *DUX4* expressie niet zijn geconserveerd tussen mens en muis. Toch kan ons model goed gebruikt worden om therapeutische interventies, gericht op het remmen van de mis-expressie van *DUX4*, in FSHD te testen.

Tenslotte beschrijven we in **hoofdstuk 7** de identificatie van twee nieuwe ziektegenen voor ICF, *CDCA7* (ICF3) en *HELLS* (ICF4). Hiermee identificeren we twee nieuwe eiwitten betrokken bij de chromatine regulatie van repeat DNA in de mens en dus potentieel belangrijke factoren voor de regulering van D4Z4 en *DUX4*. We tonen voor zowel deze nieuw geïdentificeerde genen, als voor het ICF2 gen *ZBTB24*, een rol aan in het handhaven van CpG methylering op centromere DNA repeats. De identificatie van twee

extra ziektegenen voor ICF, creëert nieuwe mogelijkheden voor het doen van studies naar het mechanisme onderliggend aan ICF syndroom, en naar de epigenetische regulatie van (macrosatelliet) repeats in het algemeen.

De vooruitgang die is geboekt in ons begrip van de (epi) genetische mechanismen die ten grondslag liggen aan FSHD en ICF-syndroom biedt nieuwe mogelijkheden voor translationeel en mechanistisch onderzoek gericht op chromatine organisatie van repetitief DNA. Het gemeenschappelijke kenmerk van repeat-derepressie in beide ziekten biedt mogelijkheden om de overeenkomsten en verschillen te bestuderen tussen de diverse mechanismen die van invloed zijn op de repressie van repeats tijdens ontwikkeling en differentiatie.

Curriculum Vitae

Peter Thijssen was born on December 6th 1983 in Zoetermeer, The Netherlands. He successfully completed pre-university education (VWO) at Het Alfrink College in Zoetermeer in 2002 with a focus on Science and Technology (natuur en techniek) supplemented with biology and economics. In September 2002, he started studying Life, Science & Technology at the University of Leiden and the Delft University of Technology. To finish his Bachelor's programme he did an internship at the department of Nephrology of the Leiden University Medical Center (LUMC), focusing on the role of C1q on the inflammation status of macrophages, under supervision of Dr. Wei Xu and Prof. Dr. Cees van Kooten. After graduating in February 2006, he enrolled in the master's programme of the Biomedical Sciences programme at Leiden university. During his master's programme, his first internship focused on the effect of D4Z4 repeat length on *FRG1* expression and was supervised by Dr. Yvonne Krom and Prof. Dr. Ir. Silvère van der Maarel at the department of Human Genetics of the LUMC. His second and final internship was conducted at TNO Quality of Life, Leiden, under supervision of Ing. Margreet de Vries and Prof. Dr. Paul Quax, and focused on the role of epigenetic dysregulation during vascular restenosis. He received his master's degree with honors (cum laude) in August 2008. In september 2008 he started working as a Ph.D. student at the departments of Human Genetics and Molecular Epidemiology of the LUMC, under supervision of Prof. Dr. Ir. Silvère van der Maarel and Prof. Dr. P. Eline Slagboom. Since January 2013 he has been working as a post-doctoral researcher at the department of Human Genetics of the LUMC under supervision of Dr. Haico van Attikum and Prof. Dr. Ir. Silvère van der Maarel.

List of Publications

Discussed in this thesis:

Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4.

Balog J[#], Thijssen PE[#], Shadle S[#], Straasheijm KR, Van Der Vliet PJ, Krom YD, Van Den Boogaard ML, De Jong A, Lemmers RJ, Tawil R, Tapscott SJ, Van Der Maarel SM.

Epigenetics. 2015 Dec

Mutations in CDCA7 and HELLS cause immunodeficiency–centromeric instability–facial anomalies syndrome.

Thijssen PE, Ito Y, Grillo G, Wang J, Velasco G, Nitta H, Unoki M, Yoshihara M, Suyama M, Sun Y, Lemmers RJ, de Greef JC, Gennery A, Picco P, Kloeckener-Gruissem B, Gungör T, Reisli I, Picard C, Kebaili K, Roquelaure B, Iwai T, Kondo I, Kubota T, van Ostaijen-Ten Dam MM, van Tol MJ, Weemaes C, Francastel C, van der Maarel SM, Sasaki H.

Nature Communications 2015 Jul

DUX4 promotes transcription of FRG2 by directly activating its promoter in Facioscapulohumeral muscular dystrophy.

Thijssen PE, Balog J, Yao Z, Pham TP, Tawil R, Tapscott SJ, Van der Maarel SM.

Skeletal Muscle 2014 Oct

Chromatin remodeling of human subtelomeres and TERRA promoters upon cellular senescence: commonalities and differences between chromosomes.

Thijssen PE, Tobi EW, Balog J, Schouten SG, Kremer D, El Bouazzaoui F, Henneman P, Putter H, Eline Slagboom P, Heijmans BT, van der Maarel SM.

Epigenetics. 2013 May

Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD.

Krom YD[#], Thijssen PE[#], Young JM, den Hamer B, Balog J, Yao Z, Maves L, Snider L, Knopp P, Zammit PS, Rijkers T, van Engelen BG, Padberg GW, Frants RR, Tawil R, Tapscott SJ, van der Maarel SM.

PLoS Genet. 2013 Apr

Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD.

Balog J, Thijssen PE, de Greef JC, Shah B, van Engelen BG, Yokomori K, Tapscott SJ, Tawil R, van der Maarel SM.

Epigenetics. 2012 Jun 1

#: authors contributed equally

Publications as co-author:

The de-ubiquitylating enzymes USP26 and USP37 regulate homologous recombination by counteracting RAP80.

Typas D, Luijsterburg MS, Wiegant WW, Diakatou M, Helfricht A, **Thijssen PE**, van de Broek B, Mullenders LH, van Attikum H.

Nucleic Acids Res. 2015 Jun

DNA methylation signatures link prenatal famine exposure to growth and metabolism

Tobi EW, Goeman JJ, Monajemi R, Gu H, Putter H, Zhang Y, Slieker RC, Stok AP, **Thijssen PE**, Müller F, van Zwet EW, Bock C, Meissner A, Lumey LH, Eline Slagboom P, Heijmans BT.

Nature Communications 2014 Nov

Leukocyte telomere length associates with prospective mortality independent of immune-related parameters and known genetic markers.

Deelen J, Beekman M, Codd V, Trompet S, Broer L, Hägg S, Fischer K, **Thijssen PE**, Suchiman HE, Postmus I, Uitterlinden AG, Hofman A, de Craen AJ, Metspalu A, Pedersen NL, van Duijn CM, Jukema JW, Houwing-Duistermaat JJ, Samani NJ, Slagboom PE.

Int J Epidemiol. 2014 Jun

MuSK IgG4 autoantibodies cause myasthenia gravis by inhibiting binding between MuSK and Lrp4.

Huijbers MG, Zhang W, Klooster R, Niks EH, Friese MB, Straasheijm KR, **Thijssen PE**, Vrolijk H, Plomp JJ, Vogels P, Losen M, Van der Maarel SM, Burden SJ, Verschuuren JJ.

Proc Natl Acad Sci U S A. 2013 Dec

Adipocyte telomere length associates negatively with adipocyte size, whereas adipose tissue telomere length associates negatively with the extent of fibrosis in severely obese women.

El Bouazzaoui F, Henneman P, **Thijssen PE**, Visser A, Koning F, Lips MA, Janssen I, Pijl H, Willems van Dijk K, van Harmelen V.

Int J Obes (Lond). 2013 Sep.

Remodeling and spacing factor 1 (RSF1) deposits centromere proteins at DNA double-strand breaks to promote non-homologous end-joining.

Helfricht A, Wiegant WW, **Thijssen PE**, Vertegaal AC, Luijsterburg MS, van Attikum H.

Cell Cycle. 2013 Sep

Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects.

Weemaes CM, van Tol MJ, Wang J, van Ostaijen-Ten Dam MM, van Eggermond MC, **Thijssen PE**, Aytekin C, Brunetti-Pierri N, van der Burg M, Graham Davies E, Ferster A, Furthner D, Gimelli G, Gennery A, Kloeckener-Gruissem B, Meyn S, Powell C, Reisli I, Schuetz C, Schulz A, Shugar A, van den Elsen PJ, van der Maarel SM.

Eur J Hum Genet. 2013 Mar

Gene set analysis of GWAS data for human longevity highlights the relevance of the insulin/IGF-1 signaling and telomere maintenance pathways.

Deelen J, Uh HW, Monajemi R, van Heemst D, **Thijssen PE**, Böhringer S, van den Akker EB, de Craen AJ, Rivadeneira F, Uitterlinden AG, Westendorp RG, Goeman JJ, Slagboom PE, Houwing-Duistermaat JJ, Beekman M.

Age (Dordr). 2011 Nov

Genetic variation in PCAF, a key mediator in epigenetics, is associated with reduced vascular morbidity and mortality: evidence for a new concept from three independent prospective studies.

Pons D, Trompet S, de Craen AJ, **Thijssen PE**, Quax PH, de Vries MR, Wierda RJ, van den Elsen PJ, Monraats PS, Ewing MM, Heijmans BT, Slagboom PE, Zwinderman AH, Doevendans PA, Tio RA, de Winter RJ, de Maat MP, Iakoubova OA, Sattar N, Shepherd J, Westendorp RG, Jukema JW; PROSPER study group; WOSCOPS study group; GENDER study group.

Heart. 2011 Jan

SOX antibodies in small-cell lung cancer and Lambert-Eaton myasthenic syndrome: frequency and relation with survival.

Titulaer MJ, Klooster R, Potman M, Sabater L, Graus F, Hegeman IM, **Thijssen PE**, Wirtz PW, Twijnstra A, Smitt PA, van der Maarel SM, Verschuuren JJ.

J Clin Oncol. 2009 Sep

Also see: <http://goo.gl/D7gnJj> or scan the QR-code below





