Expanding the mutation spectrum in FSHD and ICF syndrome

Boogaard, T.L. van den

EXPANDING THE mutation spectrum IN FSHD AND ICF SYNDROME

Marlinde van den Boogaard

Layout & cover design:	Design your thesis	www.designyourthesis.com
Printing:	Ridderprint	www.ridderprint.nl
ISBN:	978-94-6299-827-8	

Publication of this thesis was financially supported by the Leiden University Medical Center

Expanding the mutation spectrum in FSHD and ICF syndrome

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 13 februari 2018
klokke 15.00 uur

door

Tamar Linda van den Boogaard
geboren te Woerden in 1989

Promoter
Prof. Dr. Ir. S.M. van der Maarel

Co-promotoren
Dr. R.J.L.F. Lemmers
Dr. J. Balog

Leden promotiecommissie

Prof. Dr. F. Baas	
Prof. Dr. T.N. Wijmenga	University Medical Center Groningen, Groningen
Prof. Dr. J.H.L.M. van Bokhoven	Radboud University Medical Center, Nijmegen

TABLE OF CONTENTS

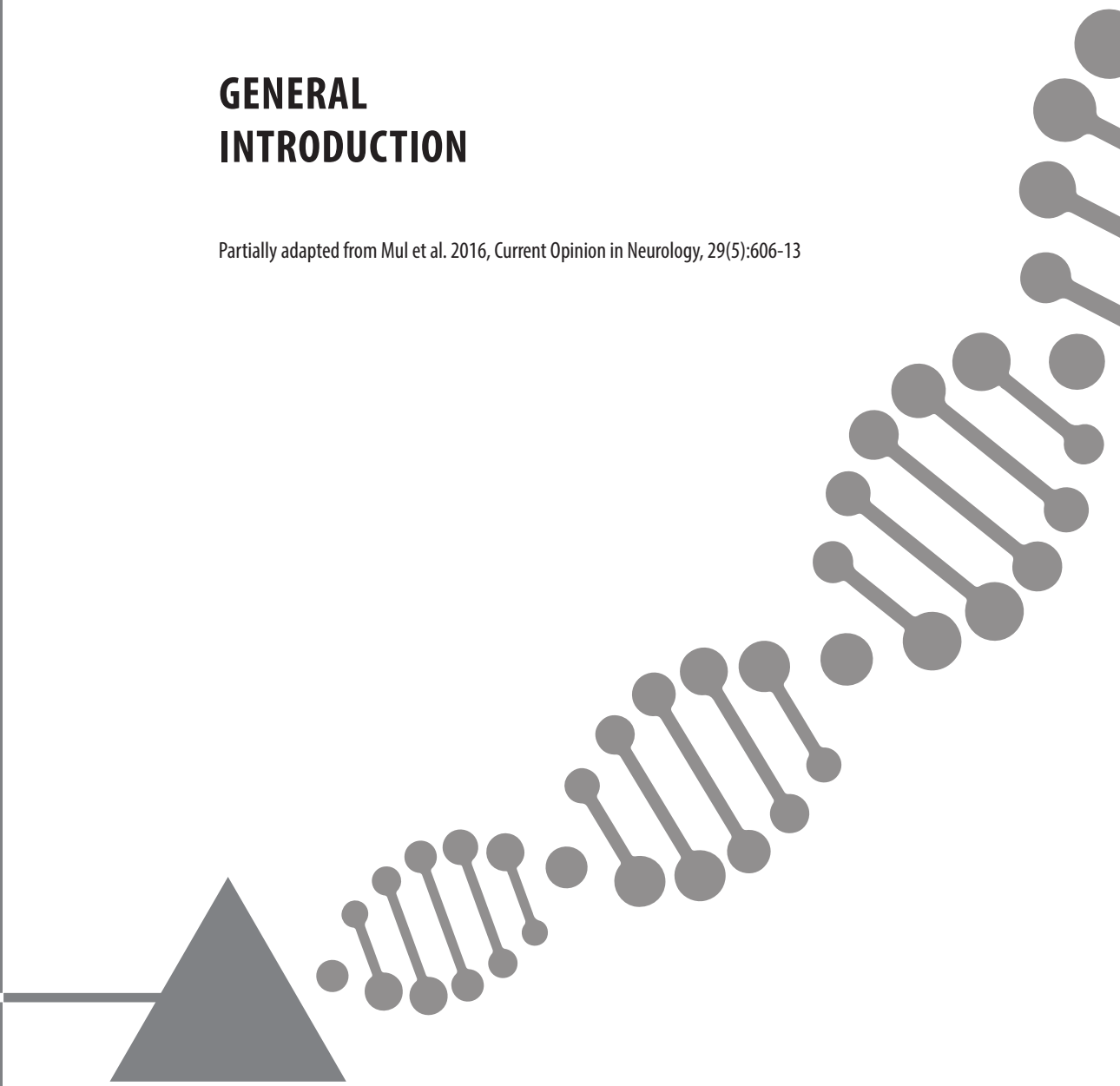
Karlien Mul[1], Marlinde L van den Boogaard[2], Silvère M van der Maarel[2], Baziel GM van Engelen[1]

1 Department of Neurology, Radboud University Medical Center, Nijmegen, The Netherlands
2 Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands

1

GENERAL INTRODUCTION

Partially adapted from Mul et al. 2016, Current Opinion in Neurology, 29(5):606-13



INTRODUCTION

Facioscapulohumeral muscular dystrophy (FSHD) is mainly characterized by progressive and often asymmetric weakness and wasting of muscles of the face, shoulder girdle, and upper arms[1]. Immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome is mainly characterized by an absence or strong reduction of immunoglobulins and facial anomalies, often combined with developmental delay[2]. Although FSHD and ICF have very distinct phenotypes, both diseases are related to errors in chromatin modifiers that are involved in the epigenetic regulation of repetitive DNA[3; 4]. In this thesis the mutation spectrum in these chromatin modifiers involved in FSHD and ICF syndrome is expanded.

Epigenetic regulation of the genome

In the nucleus DNA is organized and packed into a proteinaceous structure called chromatin. DNA is wrapped around an octet of core histone proteins, called the nucleosome. Nucleosomes are composed of pairs of 4 histone proteins (H2A, H2B, H3 and H4) that form the core of the nucleosome with their histone tails protruding outwards[5; 6]. Histones can be post-translationally modified mostly at their histone tails, and this generally affects the accessibility of the DNA in the chromatin[7; 8]. DNA accessibility can also be influenced by the spacing of nucleosomes, or by replacement of core histones with histone variants[9; 10]. The DNA itself can be modified by methylation of primarily CpG dinucleotides or in rare cases by other chemical modifications such as hydroxymethylation[11; 12].

Although almost every somatic cell in the human body contains essentially the same genome, the epigenetic regulation of the chromatin structure differs between cell types resulting in the differential expression of genes thereby creating cell diversity. Epigenetics is defined as 'the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence'[13]. Epigenetic regulation is established by chromatin modifiers, proteins which can deposit, remove, or bind to modifications on DNA or histones, or remodel the positioning of nucleosomes (Fig. 1)[14].

Writers, erasers and readers of chromatin modifications

Epigenetic marks are written (deposited), erased (removed) or read (recognized) by a variety of chromatin modifier proteins[15]. Nucleosomes are post-translationally modified on their histone tails and these modifications include methylation, acetylation, phosphorylation, ubiquitination and many others, which are established by chromatin modifiers referred to as 'writers' of histone modifications. These modifications can

also be removed by 'erasers' of histone modifications. Acetylation of histone tails, for example, is established by histone acetyltransferases (HATs) and occurs on lysine residues, neutralizing their positive charge. Histone acetylation is associated with increased accessibility of the DNA to the transcription machinery. After transcription, lysine acetylation is removed by histone deacetylases (HDACs), re-establishing the positive charge and attracting negatively charged DNA to tighten DNA-histone contacts. This process of acetylation and deacetylation is highly dynamic[8]. Methylation of histone tails is most well studied on lysine residues, but can also occur on arginine and histidine residues. Lysine methyltransferases and lysine demethylases write and erase methyl groups from histone tails, respectively[14]. Lysines can be methylated once (monomethylation, me1), twice (dimethylation, me2), or three times (trimethylation, me3)[16]. Depending on the lysine residues that are mono/di/trimethylated it might be associated with e.g. actively transcribed loci (e.g. H3 Lysine 4 trimethylation, H3K4me3) or repressed loci (H3K27me3)[8].

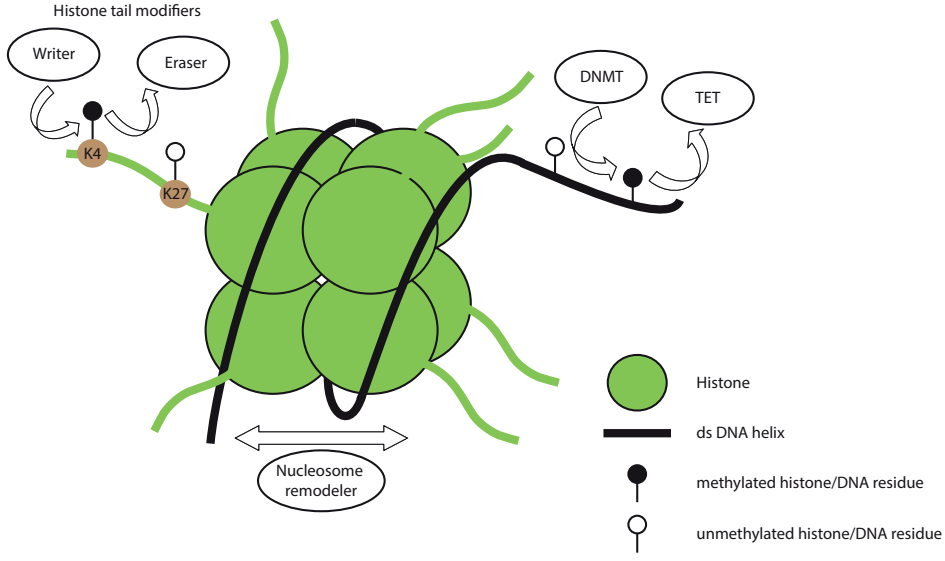


FIGURE 1. Schematic representation of epigenetic regulation of the genome by chromatin modifiers. DNA is wrapped around nucleosomes, which consists of 8 histones (2x H2A, 2x H2B, 2x H3, 2x H4). The tails of the histones are pointing outwards and can be modified by writers and erasers of histone tail modifications (e.g. histone methyltransferases and histone demethylases) at e.g. lysine 4 (K4) or lysine 27 (K27). DNA methyltransferases (DNMTs) methylate the DNA, and this modification can be removed by Ten eleven translocation (TET) proteins. Nucleosome remodeling enzymes determine the positioning, occupancy and composition of nucleosomes.

Similarly, there are chromatin modifiers responsible for the establishment and removal of DNA methylation. The family of DNA methyltransferases DNMT1, DNMT3A, DNMT3B, and DNMT3L, establish DNA methylation, or act as a cofactor in case of DNMT3L[12]. DNMT3A and DNMT3B are *de novo* methyltransferases, mainly responsible for the establishment of DNA methylation during early development. Last year another de novo methyltransferase (Dnmt3C), which is only conserved in muroid rodents, was identified in mouse[17]. DNMT1 is a maintenance methyltransferase, which preferably methylates hemimethylated DNA during replication. However, it recently has been recognized that both DNMT3 and DNMT1 enzymes have a function in both the establishment and maintenance of DNA methylation[12]. Ten eleven translocation proteins TET1, TET2, and TET3 remove DNA methylation by oxidizing methylgroups at cytosines, as a first step in the removal process[18]. DNA methylation is generally related to gene silencing, but is also identified in actively transcribed gene bodies[19].

DNA accessibility is also defined by the positioning of nucleosomes, which is important for the regulation of gene expression. Most genomic DNA is occupied by nucleosomes, however nucleosome occupancy is lower for regulatory regions, such as promoters and enhancers. Nucleosome positioning is determined by nucleosome remodelling enzyme complexes in combination with the DNA sequence and transcription factor binding[10]. Nucleosome remodelling enzyme complexes, such as the SWI/SNF (SWItch/Sucrose NonFermentable) complex, use ATP hydrolysis to slide, dissociate, or replace histones and thereby change the chromatin accessibility for e.g. transcription factors[20]. Nucleosome remodelling enzyme complexes are also involved in the exchange of histones for histone variants, which can have small or large amino acid differences compared to the 'canonical' histones. Histone variants affect chromatin organization and function by changing the structure and stability of nucleosomes, and influencing their post-translational modifications and protein interactions[9].

Histone modifications and DNA methylation can be recognized by 'reader' proteins, which bind to specific combinations of modifications. These readers can be architectural proteins, nucleosome remodelling enzymes, chromatin modifiers establishing or erasing other modifications, or proteins involved in processes such as transcription, replication or DNA repair[21]. Some reader proteins create crosstalk between different histone modifications and/or DNA methylation to reinforce chromatin states. Two examples are SET domain bifurcated 1 (SETDB1), which methylates H3K9 only in the absence of H3K4me3[22], and methyl-CpG binding protein 2 (MECP2), which binds to methylated DNA and interacts with histone deacetylases and histone methyltransferases[23].

Another class of epigenetic regulators are long non-coding RNAs (lncRNAs) that form thermodynamically stable secondary structures, and contribute to epigenetic regulation by binding and recruiting chromatin modifying proteins to their target sites.

Both lncRNAs that act *in cis* (at the site of transcription) and *in trans* (at a target away from their site of transcription) have been identified, where they modulate chromatin structure and gene expression[24]. An example of a lncRNA that functions *in cis* is *HOTTIP* (HOXA transcript at the distal tip). *HOTTIP* is transcribed from the 5' of the *HOXA* locus and recruits the WD repeat-containing protein 5/mixed lineage leukemia (WDR5/MLL) complex to *HOXA* genes. Recruitment of the WDR5/MLL complex drives H3K4 trimethylation, creating a positive feedback loop for active transcription of the *HOXA* genes[25].

Euchromatic and heterochromatic regions

DNA and histone modifications along with reader proteins generate three major types of chromatin: euchromatin, facultative heterochromatin, and constitutive heterochromatin. Euchromatin is relatively accessible chromatin, which is in general gene rich, and contains actively transcribed loci. Common histone marks at euchromatic regions include mono/di/tri-methylation of H3K4 (H3K4me1/2/3) and acetylation of histone residues[14]. Conversely, heterochromatin is found in gene poor regions and genes in heterochromatic regions are generally silenced. Heterochromatin is characterized by histone hypoacetylation and DNA methylation. Two types of heterochromatin can be distinguished: facultative and constitutive heterochromatin[26]. Facultative heterochromatin contains high levels of H3K27me3, which is established by Polycomb repressive complex 2 (PRC2) and bound by Polycomb repressive complex 1 (PRC1) at these regions. Facultative heterochromatin is often found at genes which are transcriptionally silent in certain cell types, while active in others[27]. Constitutive heterochromatin is more stable then facultative heterochromatin and contains high levels of H3K9me3 and is mainly identified at repetitive regions. Silencing of these repetitive regions is important for maintaining genome stability[26; 28].

Repetitive DNA

The human genome is comprised of more than 50% repetitive DNA[29]. Two major constituents of repetitive DNA are the interspersed repeats and tandem repeats. Interspersed repeats are identical, or nearly identical, DNA sequences dispersed throughout the genome. Interspersed repeats include long terminal repeat elements (LTRs) and long/short interspersed elements (LINES/SINES), which amplify over the genome by retrotransposition, as well as DNA transposons, which transpose directly from one location in the genome to another[30]. Tandem repeats are organized in a head to tail fashion and are found near telomeres and centromeres. Based on the size of the repetitive unit, they are classified as microsatellites, minisatellites, or macrosatellites. Microsatellite and minisatellites have a repetitive unit of 1-8 basepairs (bp) and 9-100

bp, respectively and occur in both coding and non-coding regions of the genome[31; 32]. Macrosatellites have a repetitive unit of >100 bp, but are usually multiple kilobases (kb) per unit and often located at subtelomeric or centromeric regions[33]. Some macrosatellites contain protein-coding sequences in every unit[34].

Chromatin modifiers in disease

Regulation of the chromatin structure by chromatin modifiers is essential for correct gene expression patterns. Therefore, it is not surprising that errors in epigenetic regulation of the genome can result in human disease. The most well-known example is cancer, where variants in genes encoding chromatin modifying proteins are often involved[35]. In cancer cells, DNA methylation patterns are often greatly disturbed by global CpG hypomethylation, which is often combined with hypermethylation of CpG islands of certain promoters[36]. In addition, histone modifications, as well as the positioning, occupancy, and composition of nucleosomes can be altered in cancer cells as a result of variants in chromatin modifiers[35]. These epigenetic changes can activate the expression of oncogenes, silence tumor suppressor genes, and create chromosomal instability[36].

Errors in epigenetic regulation of repetitive DNA, such as DNA hypomethylation of repeats, can lead to aberrant transcription and genome instability, which can cause disease. This thesis focuses on expanding the variant spectrum in chromatin modifiers involved in two epigenetic diseases that share a common feature of hypomethylation at repetitive DNA: Facioscapulohumeral muscular dystrophy (FSHD) and Immunodeficiency Centromeric instability and Facial anomalies (ICF) syndrome. FSHD is characterized by hypomethylation of the D4Z4 macrosatellite repeat array in the subtelomere of chromosome 4q. In ICF there is genome-wide hypomethylation of satellite repeats, mainly at centromeres. Both disorders and the chromatin modifiers involved will be further introduced below.

FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY

Prevalence

FSHD is one of the most common forms of muscular dystrophy, and was first described in 1885 by Landouzy and Dejerine[37; 38]. Since the introduction of DNA diagnostics for FSHD, its prevalence was found to be on average 5 per 100 000. A recent study in the Netherlands, however, suggests a prevalence of 12 per 100 000, which is more than twice as high as previously thought[38]. There are two genetic forms of FSHD, called

FSHD1 (OMIM 158900) and FSHD2 (OMIM 158901). FSHD1 is the most common form which affects about 95% of the patients. FSHD1 and FSHD2 share many of the same clinical characteristics and are described below as 'FSHD'.

Clinical features

FSHD is characterized by a typical distribution of muscle weakness[1]. In most cases, FSHD presents with progressive and irreversible weakness of facial muscles and/or shoulder girdle and upper arm muscles. As the disease progresses, muscles of the trunk and lower extremities may become involved as well. Age at onset is typically in the second decade of life, but varies from early childhood to onset beyond age 60. The weakness is often asymmetrical and slowly progresses over life, but life expectancy is generally not reduced. Severity and rate of progression are largely variable, ranging from isolated facial weakness to severe generalized weakness, with approximately 20% of patients eventually requiring a wheelchair[39]. There is a high percentage (20–30%) of asymptomatic or minimally affected gene carriers. Even within the same family, large differences in disease severity can occur. In most cases, FSHD is inherited in an autosomal dominant pattern, but 10–30% of cases are caused by de-novo deletions, which often occur post-fertilization and result in gonosomal mosaicism[40]. Systemic involvement is rare but can include retinal vasculopathy, sensorineural hearing loss, restrictive lung disease and (incomplete) right bundle branch block, although all are most often subclinical[41]. Patients with early onset severe disease are more at risk of developing symptomatic extramuscular manifestations.

Chromatin relaxation at the D4Z4 repeat

FSHD is genetically linked to the D4Z4 macrosatellite repeat located in the subtelomere of chromosome 4q[42]. Each D4Z4 repeat unit is 3.3 kb in size and contains a copy of the *double homeobox 4* (*DUX4*) retrogene[43]. Two almost equally common major haplotypes of the 4q subtelomere exist, 4qA and 4qB [42; 44]. FSHD is specifically linked to the 4qA (FSHD-permissive) haplotype which contains a somatic *DUX4* polyadenylation signal (PAS) following the most distal repeat unit, which is essential for stable expression of DUX4 in somatic cells (Fig. 2)[43]. Although every D4Z4 repeat unit contains a copy of the first two exons of the *DUX4* gene, stable DUX4 expression occurs most efficiently from the last repeat unit which is followed by a PAS, and is much less prevalent from internal D4Z4 units[45]. In FSHD there is chromatin relaxation of the D4Z4 repeat associated with expression of *DUX4* in skeletal muscle, which is considered as the molecular hallmark of FSHD. In muscle cells of FSHD patients there is variegated expression of *DUX4* since relatively high levels of DUX4 protein are only detected in a small percentage of myonuclei[3]. In controls, the D4Z4 repeat is transcriptionally silent in somatic cells,

and is marked by both heterochromatic (H3K9me3 and H3K27me3) and euchromatic histone marks (H3K4me2 and H3Ac). H3K9me3 at D4Z4 is established by the H3K9 methyltransferase SUV39H1 (suppressor of variegation 3-9 homolog 1) and bound by HP1γ (heterochromatic protein 1γ) and cohesin[46]. Furthermore, the GC rich D4Z4 repeat (73%) is highly methylated in controls[47; 48]. In FSHD patients the chromatin of the D4Z4 repeat is less compacted as indicated by a lower chromatin compaction score (H3K9me3:H3K4me2 ratio) and reduced CpG methylation levels compared to controls[48; 49]. Additionally, HP1γ and cohesin are also reduced in FSHD patients[46]. Together these changes emphasize the chromatin relaxation at D4Z4 in FSHD patients.

Genetic mechanism

The two genetic forms of FSHD, FSHD1 and FSHD2, are clinically almost indistinguishable[50]. Both forms share the chromatin relaxation of the D4Z4 macrosatellite repeat array on chromosome 4 and the transcriptional derepression of the D4Z4 encoded *DUX4* gene in skeletal muscle. In the most common form, FSHD1, D4Z4 chromatin relaxation and *DUX4* expression are associated with a contraction of the D4Z4 repeat to an array of 1–10 units (Fig. 2)[51]. This repeat varies between 8 and 100 units in the control population, with FSHD-sized alleles of 8-10 units found in 1-3% of the control population[52; 53]. The D4Z4 contraction must occur on a 4qA permissive chromosome (containing a *DUX4* PAS for stable expression) to cause FSHD. Contractions on 4qB chromosomes or of the very homologous D4Z4 repeat on chromosome 10 are associated with chromatin relaxation, but they are not associated with FSHD since these repeats lack a *DUX4* PAS. FSHD2 is characterized by D4Z4 chromatin relaxation of both 4q chromosomes as well as the very homologous D4Z4 repeats on chromosome 10, but without repeat contraction. FSHD2 patients carry at least one permissive 4qA allele. The median size of the shortest permissive allele in FSHD2 patients is 16 units, which is smaller than the median size in the control population (23 units)[54]. Heterozygous variants in the chromatin modifier *structural maintenance of chromosome flexible hinge domain containing 1* (*SMCHD1*) gene have been identified as the most common cause of FSHD2 (Fig. 2)[54-59]. The mutation spectrum of *SMCHD1* includes missense variants, insertions, deletions, nonsense and splice site variants and they are distributed over the entire *SMCHD1* gene. Haploinsufficiency as well as expected loss of function variants are found in FSHD2 patients, indicating that both types of functional outcomes of *SMCHD1* variants can cause FSHD2. Heterozygous *SMCHD1* variants explain about 85% of the FSHD2 patients. In some FSHD2 families, characterized by their clinical phenotype and D4Z4 hypomethylation, no exonic *SMCHD1* variant that affects function could be identified. This suggests that FSHD2 is genetically heterogeneous.

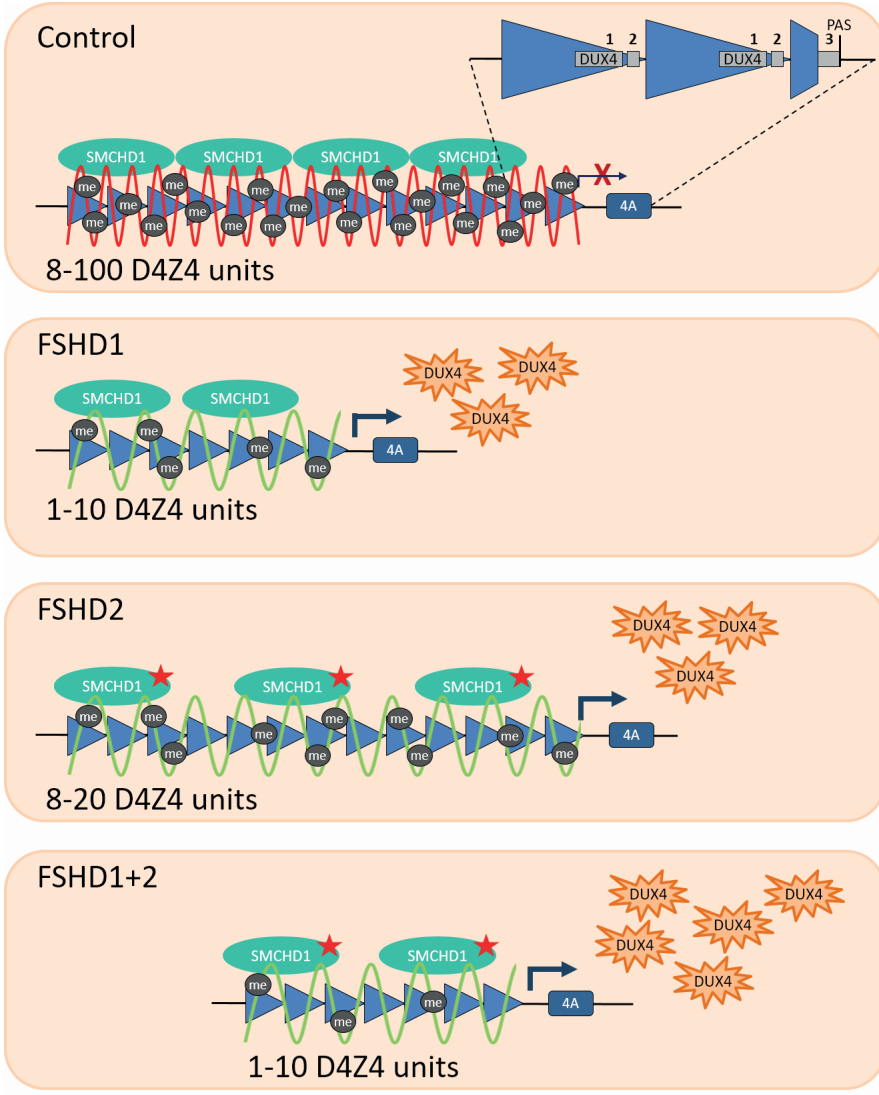


FIGURE 2. Genetic mechanisms of FSHD. In controls the D4Z4 repeat array (triangles) on chromosome 4q varies between 8 and 100 units and adopts a repressed chromatin structure (red wavy lines) characterized by high CpG methylation (me). Each D4Z4 repeat contains a copy of the DUX4 gene, without a polyadenylation signal (PAS). Only on 4qA chromosomes (and not on 4qB chromosomes) the last repeat unit is followed by a third exon which contains a PAS which can stabilize the DUX4 transcript from the last repeat unit. FSHD is related to chromatin relaxation (green wavy lines) of the D4Z4 repeat array, facilitating the stable expression of DUX4 from a 4qA chromosome. This chromatin relaxation is associated with reduced CpG methylation. FSHD1 is caused by a contraction of the D4Z4 repeat array to a size of 1-10 units. FSHD2 is caused by a variant in a chromatin modifier, most often SMCHD1. Variants in SMCHD1 can also act as a modifier of disease severity in FSHD1.

SMCHD1 is an epigenetic regulator of D4Z4

SMCHD1 is an atypical member of the SMC gene superfamily of proteins, which are involved in chromosome organization and dynamics [60;61]. SMCHD1 contains a functional GHKL (gyrase, heat-shock protein 90, histidine kinase, MutL) type ATPase domain and a hinge domain flanked by a coiled-coil domain on both sides. The hinge domain is involved in homodimerization of SMCHD1 and essential for chromatin binding[61;62].

Smchd1 was originally identified as regulator of epigenetic silencing in an N-ethyl-N-nitrosourea (ENU) mutagenesis screen in mice[63]. Female embryos carrying a homozygous nonsense variant in *Smchd1* show mid-gestation lethality and a defect in X-chromosome inactivation[64]. X-chromosome inactivation ensures dosage compensation in female mammals, by epigenetic silencing of one of the two X-chromosomes. It is initiated by expression of Xist (X-inactive specific transcript) lncRNA, which covers the X-chromosome *in cis* and recruits PRC1 and PRC2. These complexes establish repressive histone modifications on the Xist bound X-chromosome to from facultative heterochromatin, resulting in silencing of most genes on the inactive X-chromosome[65]. Smchd1 has been shown to play an important role in methylation of CpG islands and gene silencing on the inactive X-chromosome[64; 66; 67]. Additionally, SMCHD1 interacts with XIST associated H3K27me3 domains and compacts the X-chromosome by an interaction with HBiX1 bound to HP1 associated H3K9me3 domains[68]. Besides its role in X-chromosome inactivation, Smchd1/SMCHD1 is involved in silencing of several autosomal loci, including imprinted genes and the clustered protocadherin genes[67; 69-71].

With respect to FSHD, chromatin immunoprecipitation (ChIP) showed that SMCHD1 binds directly to D4Z4 and this binding is reduced in muscle cells of FSHD2 patients[55]. Knockdown of SMCHD1 in control myotubes with a permissive D4Z4 allele activates *DUX4* expression and increases *DUX4* expression in FSHD1 myotubes[55; 56]. Conversely, moderate overexpression of SMCHD1 with lentiviral systems in FSHD1 and FSHD2 myoblasts reduces *DUX4* expression. Interestingly, SMCHD1 protein levels decrease during muscle cell differentiation, suggesting that muscle might be more prone to D4Z4 derepression[72]. Together this indicates that SMCHD1 has an important role in silencing the D4Z4 repeat in somatic tissue such as skeletal muscle and that *DUX4* derepression as observed in FSHD muscle is reversible.

Transcriptional (de)repression at D4Z4

Detailed knowledge about the mechanisms of (de)repression at D4Z4 is thought to be essential for the development of therapeutic strategies aimed at silencing *DUX4* in FSHD patients. As discussed above, SMCHD1 has been identified as a key chromatin modifier for D4Z4. Besides SMCHD1, other chromatin modifiers which bind to D4Z4

include SUV39H1, HP1γ and cohesin. However, knockdown of these proteins did not increase *DUX4* expression in primary myotube cultures[72]. Additionally, members of the PRC2 complex are also binding to D4Z4 in control and FSHD proliferating myoblasts[73; 74]. Specifically in FSHD2 increased levels of H3K27me3 and the PRC2 protein SUZ12 were detected compared to controls. Inhibition of a member of the PRC2 complex increased *DUX4* expression but only in FSHD2 myotubes, indicating that although the epigenetic (mis)regulation of D4Z4 is highly similar in FSHD1 and FSHD2, there are also differences[72].

The transcription of small and long non-coding RNA fragments from D4Z4 also influences D4Z4 (de)repression. A lncRNA transcribed from the D4Z4 repeat was identified specifically in FSHD patients. This lncRNA, DBE-T (D4Z4 binding element transcript), recruits the H3K36 methyltransferase ASH1L (ASH1 like histone lysine methyltransferase) and activates *DUX4* expression[74]. Small RNA fragments (siRNAs and miRNAs) are bidirectionally transcribed from the D4Z4 repeat in fibroblasts and muscle cells of controls and FSHD patients[45]. It was discovered that these siRNAs also contribute to D4Z4 silencing, since knockdown of DICER and AGO2 (argonaute2), components of the endogenous siRNA pathway, upregulates *DUX4* expression in control muscle cells. On the other hand, exogenous siRNAs targeted against the region upstream of the *DUX4* transcription start site increase AGO2 recruitment and H3K9me2 at the D4Z4 repeat and repress *DUX4* expression[75].

The opening of the chromatin structure at D4Z4 in FSHD patients might render the repeat more accessible for proteins that are not present at the repeat in controls. Two examples are PARP1 (poly(ADP-ribose) polymerase 1) and DNMT1, which interact with the *DUX4* promoter only in immortalized myoblasts from an FSHD patient, but not in the myoblasts from an unaffected sibling. Treatment with PARP1 inhibitors can suppress *DUX4* expression in these FSHD immortalized myoblasts, indicating that PARP1 might modulate *DUX4* expression[76].

DUX4 is a transcription factor toxic to skeletal muscle

Variegated *DUX4* expression in skeletal myocytes is common to both forms of FSHD and is thought to be necessary for the disease to develop. DUX4 is a transcription factor, which is normally expressed in the luminal cells of the testis and in the thymus, and suppressed in most other somatic tissues[3; 77]. Because of the D4Z4 chromatin relaxation, sporadic *DUX4* expression can be detected in a small and variable percentage of myonuclei from FSHD patients, and sometimes also at very low levels in some control muscle biopsies[3; 78]. How *DUX4* expression leads to muscle weakness is under active investigation, but ectopic expression in myocytes leads to a complex cascade of events eventually leading to muscle cell death[79]. Close homologues of *DUX4* have only been

identified in primates and afrotheria[80; 81]. Nevertheless, overexpression of *DUX4* induces cell death not only in human cell lines[82], but also in adult mouse muscle,[83] in *Xenopus laevis*[84], in Zebrafish[45; 83] and in *Drosophila*[85]. This DUX4 induced apoptosis seems to be dependent on P53[83; 86]. DUX4 contains a N-terminal double homeobox domain involved in DNA binding and a C-terminal activation domain. The N-terminal double homeobox domain can bind DNA in the absence of the C-terminal domain, but it lacks the ability to activate DUX4 target genes and its overexpression is not toxic[87; 88]. The C-terminal activation domain of DUX4 interacts with p300/CBP (CREB-binding protein) and DUX4 might upregulate the expression of (part of) its target genes by recruiting p300/CBP, which induces acetylation of H3K27 and activates transcription[89].

As DUX4 is a transcription factor, several studies analysed the transcriptional consequences of the presence of DUX4 in muscle cells. Differentially expressed genes include both direct targets of DUX4, as indicated by DUX4 binding sites identified by ChIP-sequencing, as well as indirect targets[79; 87; 90]. A comparison of transcriptional profiles of muscle cells expressing *DUX4* from different origins (endogenous DUX4, lentiviral transduced DUX4, or doxycycline inducible DUX4) revealed highly correlated gene expression signatures. The transcriptional consequences of *DUX4* expression include activation of germ line genes, suppression of the innate immune response, and disruption of RNA metabolism, as well as other pathways[79; 87; 90]. In *DUX4* expressing myoblasts there is accumulation of alternative transcripts containing premature stop codons and aberrant spliced exons, which might be related to the inhibition of nonsense-mediated decay (NMD) by DUX4, as its expression triggers the degradation of UPF1 (up-frameshift protein 1), a central component of NMD[79; 86]. Since DUX4 is also a target of NMD, the inhibition of NMD stabilizes DUX4 mRNA and this feedback loop might contribute to the burst of *DUX4* expression[86]. Overexpression of *DUX4* in immortalized myoblast cultures also increases the level of reactive oxygen species and induces DNA damage, which possibly affects myogenic differentiation of FSHD myoblasts[91]. Furthermore, both endogenous and exogenous *DUX4* expression in myotubes induce coaggregation of two DNA/RNA binding proteins FUS and TDP43, and disruption of the structure of some nuclear bodies (PML bodies and SC35 speckles)[92; 93]. These alterations in nuclear body structure might disturb (epi)genetic gene regulation in *DUX4* expressing myotubes[93]. The contributions of these processes to FSHD pathology need to be studied further.

Besides *DUX4*, other genes proximal to D4Z4, such as *FRG1 (FSHD region gene 1*), *FRG2* (*FSHD region gene 2*) and *FAT1* (*FAT atypical cadherin 1*) might also be deregulated in FSHD either by DUX4 directly or independently, or a combination thereof[94-96]. Their clinical relevance to disease warrants further investigation as some FSHD families genetically exclude a direct role for proximal 4qter genes[43; 97; 98].

Phenotype-(epi)genotype relation

The high clinical variability in FSHD raises the question whether (epi)genetic factors influence disease severity. Indeed, in FSHD1, there is a rough inverse correlation between residual D4Z4 repeat array size and disease severity, which was recently confirmed in Korean and Chinese cohorts[99; 100]. Individuals with D4Z4 repeats of 7–10 units typically have milder phenotype or remain non-penetrant and have lower disease grades for their muscle biopsies[101; 102]. In addition, D4Z4 repeats of 7–10 units on disease permissive 4qA chromosomes are found in 1–3% of the control population, illustrating the reduced penetrance of these alleles[52; 53]. However, unaffected carriers exist for smaller repeat sizes (1–3 units) as well, indicating that repeat size seems not always predictive for disease severity, which can only be partly attributed to somatic mosaicism[103]. Furthermore, in some FSHD1 families with borderline repeat sizes and D4Z4 hypomethylation, *SMCHD1* variants were identified that modify disease severity (Fig. 2). While carriers of both an FSHD1 sized allele and an *SMCHD1* variant were typically severely affected, familial carriers of only one of these conditions showed a very mild phenotype[56; 57].

The relevance of the epigenetic state of the D4Z4 repeat for disease severity was shown in multiple studies, which use D4Z4 methylation as a measure of chromatin relaxation[54; 104-106]. A correlation between D4Z4 methylation and D4Z4 repeat array size was identified at different sites in the repeat using two methods: methylation sensitive Southern blotting and bisulfite sequencing. Digestion of the D4Z4 repeat with the methylation sensitive digestion enzyme FseI followed by Southern Blot is used to determine the methylation at the FseI site in a sequence identical to the *DUX4* promoter region of the most proximal unit of the D4Z4 repeats on chromosomes 4 and 10 simultaneously[55]. A significant correlation between the total number of D4Z4 repeat units on chromosomes 4 and 10 and the methylation level at this FseI site was found in PBMCs of controls, FSHD1 and FSHD2 patients[54]. With this correlation the methylation can be predicted based on repeat size. The Delta1 value is defined as the observed methylation minus the predicted methylation in controls. The delta1 value varies between -10% and 10% in controls and FSHD1 patients and is <-21% in FSHD2 patients[54].

In contrast to Southern blotting, D4Z4 bisulfite sequencing enables the analysis of the methylation status of multiple CpGs simultaneously. The first study that used D4Z4 bisulfite sequencing identified specific regions in D4Z4 with reduced methylation in FSHD patients[107]. A recent study performed bisulfite sequencing specifically in the region immediately distal to D4Z4 on 4qA alleles and identified a correlation between D4Z4 methylation at a specific CpG (called 'CpG6') in the PAS region and the repeat array size of the 4qA allele. An additive effect of D4Z4 repeat array size and *SMCHD1* variants was seen in FSHD2 patients, which showed much lower methylation at this CpG

compared to FSHD1 patients and controls with the same 4qA length[106]. Furthermore, a linear correlation was found between the age-corrected clinical severity score and the methylation at the FSHD allele in FSHD1 and FSHD2 patients (individuals with a 4qA and 4qB allele)[106]. These studies indicate that D4Z4 methylation is dependent on D4Z4 repeat array size, and that FSHD2 patients have lower D4Z4 methylation than expected based on repeat array size.

In addition, a significantly reduced Delta1 value was detected in affected but not in unaffected carriers of an FSHD1 sized allele of 7-10 units. This indicates that D4Z4 methylation is lower than predicted based on the sizes of the repeats for affected carriers of a FSHD1 sized allele of 7-10 units[54]. In another study, bisulfite sequencing identified lower methylation levels in myocytes from FSHD1 patients compared to myocytes from unaffected relatives with an FSHD1 sized allele at a region upstream of the DUX4 open reading frame (DR1) and at the gene body in the distal D4Z4 repeat[105]. These studies emphasize that FSHD1 affected individuals are epigenetically more susceptible to *DUX4* expression and disease presentation. Indeed, myocytes from FSHD1 affected individuals were more responsive to treatment with compounds that relieve epigenetic repression of *DUX4* than myocytes from their unaffected relatives[105]. It is possible that expression levels of, or polymorphisms in chromatin modifiers, such as *SMCHD1*, could influence this epigenetic susceptibility.

For FSHD2, the size of the smallest permissive repeat is suggested to contribute to the epigenetic susceptibility to disease presentation[54]. In FSHD2 patients the D4Z4 methylation attributed to the smallest permissive allele correlates with disease severity, with shorter repeats being associated with lower methylation and a higher disease severity[54]. In addition, the nature of the *SMCHD1* variant influences the extent of hypomethylation with open reading frame (ORF) preserving variants resulting in significantly lower D4Z4 methylation than ORF disrupting variants[54]. SMCHD1 forms homodimers through its hinge domain[61; 62], and open reading frame preserving variants might result in malfunctioning dimers, which could explain why these variants have a more severe effect on methylation. Altogether, this suggests that FSHD1 and FSHD2 are representatives of the same disease spectrum, where a combination of the D4Z4 repeat array size and its epigenetic state influence disease severity.

IMMUNODEFICIENCY CENTROMERIC INSTABILITY AND FACIAL ANOMALIES SYNDROME

Clinical characteristics

Immunodeficiency, centromeric instability and facial anomalies syndrome (ICF) is a rare, autosomal recessive disease. The first patients were described in two reports from the late 1970s and the acronym ICF was introduced in 1988[108-110]. The immunodeficiency of ICF patients is characterized by hypo- or a-gammaglobunemia (an absence or strong reduction of immunoglobulins), in the presence of circulating B-cells. However, only naive B-cells have been detected in blood of ICF patients, while memory B-cells were absent, which suggests a defect in B-cell maturation in ICF patients[111]. The immunodeficiency causes susceptibility to severe infections of the respiratory and/or gastro-intestinal tract, which can be fatal[2]. Most ICF patients are treated with immunoglobulin replacement therapy (intravenous immunoglobulin (IVIG)). In some patients with severe infections hematopoietic stem cell transplantation was successfully performed[112; 113]. Facial anomalies are observed in almost all patients, and most often include flat nasal bridge, hypertelorism, and epicanthus. Besides the immunodeficiency and facial anomalies, ICF patients often have a delay in motor and/or speech development[2].

Centromeric instability

The cytogenetic hallmark of ICF is centromeric instability of chromosomes 1, 9, and 16. In mitogen stimulated lymphocytes of ICF patients, and at lower frequencies in other cell types, a variety of chromosomal anomalies involving the pericentromeric regions of mainly chromosomes 1 and 16 can be recognized. Chromosome 9 is involved to a lesser extent. These chromosomal anomalies include whole-arm deletions, pericentromeric breaks, multibranched chromosomes, and decondensed chromosomes[114]. Similar chromosomal anomalies can be found in lymphocytes cultures treated with demethylating agents, which provided the first clue that ICF patients suffer from a DNA methylation defect[115]. Indeed, ICF patients show hypomethylation of satellite repeats, which are highly abundant at the centromeric region of chromosomes 1, 9, and 16, and are involved in the chromosomal anomalies[115; 116].

Genetic heterogeneity

Genetically ICF is a heterogeneous disease. The first gene defect for ICF was discovered in 1999 and is defined by recessive variants in the *DNA methyltransferase 3B* (*DNMT3B)* gene[117; 118]. However, only about half of the ICF patients can be explained by variants in *DNMT3B* (ICF1; OMIM 602900). Furthermore, *DNMT3B* variant-negative patients showed

additional hypomethylation of alpha satellite DNA[119], while hypomethylation of the D4Z4 macrosatellite repeat seems to be specific to ICF1 patients. In 2011 variants in the *zinc-finger and BTB domain containing 24* (ZBTB24) gene were identified in the majority of *DNMT3B* variant negative patients (ICF2; OMIM 614064)[50]. Recently, variants in the *cell division cycle associated 7* (*CDCA7*, ICF3; OMIM 616910) gene and variants in the *helicase, lymphoid-specific* (HELLS, ICF4; OMIM 616911) gene have been identified in ICF patients negative for both *DNMT3B* and *ZBTB24* variants[120].

ICF1

About 50% of the ICF patients carry recessive variants in *DNMT3B* (ICF1)[2]. DNMT3B is considered a de-novo methyltransferase, which is responsible for the establishment of DNA methylation during early development, together with DNMT3A and the stimulatory cofactor DNMT3L[12]. It contains a PWWP domain, an ADD (ATRX, DNMT3, DNMT3L) domain and a catalytic domain located at the C-terminus. The PWWP domain is involved in targeting of DNMT3B to chromatin marked by H3K36me3[121]. The ADD domain binds to histone H3 when it is unmethylated at lysine 4 (H3K4), this binding is blocked by H3K4me3[122]. Together these domains target the catalytic activity of DNMT3B to regions of the genome with H3K36me3 and without H3K4me3[12]. Most ICF1 variants are missense variants in the catalytic domain of DNMT3B, leading to reduced methyltransferase activity *in vitro*[123; 124]. Some patients carry a nonsense variant, but this is always combined with a missense variant[2]. This suggests that homozygous nonsense variants might be embryonic lethal, as is observed for homozygous Dnmt3b knockout mice. In Dnmt3b knockout embryos, but not in Dnmt3a knockout embryos, reduced methylation levels of centromeric minor satellite repeats were detected, indicating that these repeats are specific targets of Dnmt3b[125]. Mice carrying homozygous or compound heterozygous ICF1 variants show hypomethylation of three repetitive sequences: major satellite repeats, minor satellite repeats, and C-type retrovirus. Their methylation levels are however higher than in knockout mice indicating that mutated Dnmt3b retains some functionality[126]. Mice with homozygous or compound heterozygous hypomorphic Dnmt3b variants also show hypomethylation and upregulation of germ line genes as well as reduced viability, facial anomalies, runting phenotype, and T-cell apoptosis[126-128]. Interestingly, in contrast to ICF patients, no B-cell defect is detected in the ICF1 mouse models. Similarly, to the ICF1 mouse models, ICF1 patients show hypomethylation of various repetitive regions such as centromeric satellite repeats, Alu repeats[116], interspersed LINE-1 repeat elements on the X-chromosome[129], the D4Z4 macrosatellite repeat, and NBL2 repeat[130].

ICF2

ICF2 patients carry recessive variants in *ZBTB24*, often nonsense variants, suggesting that these patients lack full length ZBTB24 protein. In addition, some patients with homozygous or compound heterozygous missense variants have been reported. In contrast to humans, where the complete absence of ZBTB24 seems to be compatible with life, a homozygous Zbtb24 BTB domain deletion mouse, which lacks functional Zbtb24 protein, was found to be embryonic lethal[131].

The function of ZBTB24 is largely unknown. ZBTB24 is a member of the BTB-ZF family of proteins and contains an N-terminal BTB domain suggested to be involved in dimerization and protein-protein interaction, an AT hook which is predicted to bind AT-rich sequences, and eight C-terminal zinc fingers, typically involved in DNA binding. Many BTB-ZF proteins are involved in the development of the immune system and they are thought to function as transcriptional suppressors. Recently, a role of ZBTB24 as transcriptional repressor was discovered in human B-cell line Raji cells, where downregulation of ZBTB24 upregulates the expression of two transcriptional factors (*Interferon regulatory factor 4* (*IRF-4*) and *B lymphocyte-induced maturation protein-1* (Blimp-*1*))[132]. However, a role for ZBTB24 as transcriptional activator of a subset of genes, most prominently the ICF3 gene *CDCA7* was recently reported as well[131]. Mouse Zbtb24 was also shown to colocalize with HP1α, a heterochromatic protein, in NIH3T3 cells[133], and it has a role in the maintenance of CpG methylation at centromeric repeat in MEFs[120]. These properties fit with the hypomethylation at centromeric and pericentromeric repeats in ICF2.

ICF3 and ICF4

The recent discovery of two new disease genes genetically confirmed ICF syndrome in 8 families negative for *DNMT3B* and *ZBTB24* variants. Homozygous variants in *CDCA7* were identified in four ICF families. These variants were all missense variants and affect residues near the first two zinc-finger motifs in the conserved carboxyterminal 4-CXXC-type zinc finger domain[120]. The molecular function of CDCA7 is unknown, however some interaction partners and transcriptional activators of Cdca7 have been identified. Cdca7 is a transcriptional target and interaction partner of Myc and is involved in neoplastic transformation[134; 135]. *Cdca7* was also identified as a Notch target gene and contributes to hematopoietic stem cell emergence[136]. In mouse embryonic fibroblasts transiently depleted for Cdca7 a reduction of CpG methylation at minor satellite repeats was detected, indicating that Cdca7 also has a role in the maintenance of CpG methylation[120]. This is consistent with the hypomethylation detected in ICF patients.

In four other ICF families recessive variants in *HELLS* were identified, including a missense variant, nonsense variants and an in frame deletion of an amino acid[120]. Two of the ICF4 patients carry homozygous or compound heterozygous nonsense variants in HELLS, suggesting that the absence of HELLS is compatible with human life[120]. This is in contrast to homozygous deletion of *Hells* in mice, which causes perinatal lethality[137]. HELLS is an ATP dependent chromatin remodelling protein of the SNF2 family of helicases and is involved in DNA methylation during development. HELLS/Hells is an interaction partner of DNMT3B/Dnmt3b, and in mouse embryonic stem cells Hells recruits Dnmt3b to repeat elements[138-140]. This implies a functional connection between the ICF1 and ICF4 genes in the establishment of CpG methylation at repeat elements, which is reduced in ICF patients. Additionally, transient depletion of Hells in mouse embryonic fibroblasts reduced CpG methylation at minor satellite repeats, suggesting that Hells has an additional role in the maintenance of CpG methylation[120]. There are still few genetically unexplained ICF patients, suggesting that there is at least one more gene ICF gene involved.

Functional studies

Most functional studies on ICF syndrome have been performed on a limited set of samples, mainly of ICF1 patients and in a heterogeneous set of cell types. They include multiple studies aimed at defining methylation and expression profiles. Besides the hypomethylation of repetitive regions in ICF1 patients, there is a genome wide loss of methylation and upregulation of expression at various non-repetitive regions, including multiple coding genes involved in e.g. immune function, development, neurogenesis, and germ line pathways[141-145]. Hypomethylated positions were identified in both genic and intergenic regions as well as promoters[141]. For some hypomethylated genes a concomitant loss of repressive histone marks (H3K37me3 and H3K9me3) and increase in active histone marks (H3K4me3 and H3K9ac) was identified as well, indicating that the epigenetic changes in ICF1 syndrome are not limited to CpG hypomethylation[142]. In ICF1 patient cell lines hypomethylation of subtelomeric regions was associated with decreased telomere lengths in these patients[146]. Additionally, higher levels of H3K4me2 and reduced levels of H3K9me3 were measured at subtelomeres, co-occurring with increased expression of the lncRNA TERRA (telomeric repeat containing RNA) that originates from subtelomeric regions[146; 147]. The genome wide loss of methylation is also present in induced pluripotent stem cells (iPSCs) derived from ICF1 patients. A global loss of non-CG methylation and a loss of CpG methylation was detected at some promoters and enhancers, as well as at large regions in the centromeric and subtelomeric regions. Almost all of the alterations remain present in iPSC-derived mesenchymal stem cells (MSCs), which suggests that induced pluripotent stem cells (iPSCs) might

be a useful tool to study ICF syndrome[144]. Although the telomeres in ICF1-iPSCs are elongated compared to their parental fibroblasts, high TERRA levels and subtelomeric hypomethylation were still present in ICF1-iPSCs. When iPSCs are differentiated to fibroblast-like cells, they maintain high TERRA levels and accelerated shortening of telomere[148].

Besides changes in methylation patterns, the nuclear architecture is also altered in ICF syndrome. The average area of juxtacentromeric heterochromatin from chromosome 1 and the association of these regions with the nuclear periphery were reduced in B-lymphoblastoid cell nuclei from an ICF1 and ICF2 patient compared to controls. The changes in nuclear organisation might contribute to alterations in gene expression independent of DNA methylation differences as exemplified by two genes from chromosome 1 that are upregulated in ICF and show reduced co-localisation with juxtacentromeric heterochromatin compared to controls[149]. Also in fibroblasts from a female ICF1 patients an altered nuclear localization of some genes at the inactive X chromosome was identified, including genes more distant from the site of hypomethylation[150]. The changes in the chromatin structure and nuclear architecture might be responsible for changes in replication timing. In ICF1 patient cell lines advanced replication timing of satellite 2[151] and telomeres[146], as well as genes in facultative (Inactive X-chromosome)[152] and constitutive heterochromatin[153], was detected.

The functional studies did not yet reveal a plausible explanation for the immune defect, as well as other comorbidities. Furthermore, since almost all these studies were performed in ICF1 patient cells, the hypomethylation and expression profiles of different ICF subtypes cannot be compared to identify common changes which might play a role in the pathogenic mechanism.

OUTLINE OF THIS THESIS

This thesis focuses on expanding the mutation spectrum in chromatin modifiers involved in two epigenetic diseases that share a common feature of hypomethylation of repetitive DNA: FSHD and ICF syndrome. In chapters 2, 3 and 4 we expand the *SMCHD1* mutation spectrum in FSHD2. In chapter 5 we identified a new FSHD2 gene, *DNMT3B*, which was previously shown to cause ICF syndrome. In chapter 6 we expand the mutation spectrum of ICF1 and ICF2.

In chapter 2 we describe three FSHD2 families in which the proband carries two potentially damaging *SMCHD1* variants. We investigated whether these variants were located *in cis* or *in trans* and whether they contribute to D4Z4 hypomethylation independently. In one of the families the variants were located in trans and contribute

synergistically to D4Z4 hypomethylation and disease penetrance in FSHD2. This family also shows that carrying two *SMCHD1* missense variants *in trans* is compatible with life in males and females.

In chapter 3 we describe two FSHD2 families with a large deletion spanning several genes on chromosome 18p, including *SMCHD1*. These families show that a deletion of *SMCHD1* can also cause FSHD2 and this might put 18p deletion carriers (with a deletion including *SMCHD1*) at risk of developing FSHD2. We identified D4Z4 hypomethylation in these 18p deletion cases and combined with the distribution of D4Z4 repeat array size and frequency of permissive alleles, we estimated that 1:8 of the 18p deletion cases is at risk of developing FSHD2.

In chapter 4 we identified two FSHD families with a (deep) intronic variant in *SMCHD1*. In both families this results in alternative transcripts with premature stopcodons. In one family an intronic variant in *SMCHD1* leads to the inclusion of the last 14 base pairs of intron 13 into the transcript, which is predicted to disrupt the open reading frame. In the other family a deep intronic variant in *SMCHD1* creates a splice acceptor site, which results in exonisation of 53 base pairs from intron 34 into the transcript, which is predicted to disrupt the open reading frame. This variant modifies disease severity in an FSHD1 family.

Because some FSHD2 cases cannot be explained by variants in *SMCHD1*, we performed whole exome sequencing in eight families with D4Z4 hypomethylation without a pathogenic *SMCHD1* variant. In chapter 5 we describe that in two of these families we identified missense variants in *DNMT3B* segregating with D4Z4 hypomethylation and increased penetrance of FSHD. Since recessive *DNMT3B* variants were previously shown to cause ICF1, we investigated ICF characteristics in these FSHD families, as well as *DUX4* expression in ICF1 patients. We identified some mild epigenetic characteristics of ICF in one *DNMT3B* variant carrier from these FSHD families, but without an immunophenotype. Additionally, we identified DUX4 expression in myoblasts from one ICF1 patient as well as in MyoD transduced fibroblasts from two other ICF1 patients. Both the short life expectancy of ICF1 patients as well as the absence of a short permissive allele in their parents might explain why a muscle phenotype has so far not been reported in ICF1 families.

Finally, in chapter 6 we expand the mutation spectrum of the two most common types of ICF, ICF1 and ICF2. We describe seven new ICF1 cases from four different families, which carry in total six different *DNMT3B* missense variants, which have not previously been reported in ICF syndrome. Additionally, we described five ICF2 cases from five different families, including one ICF2 patient that carries a homozygous deletion including *ZBTB24*. This patient confirms that complete absence of ZBTB24 protein is compatible with life. Additionally, we analysed the prevalence of the different

ICF subtypes in the literature and this study. This revealed that 56% of the cases have ICF1, 31% ICF2, 7% ICF3 and 7% ICF4. Interestingly, in ICF2, we noticed a gender bias, with 79% male patients and only 21% females. This suggests that *ZBTB24* variants are more deleterious in females.

REFERENCES

1. Mul, K., Lassche, S., Voermans, N.C., Padberg, G.W., Horlings, C.G., and van Engelen, B.G. (2016). What's in a name? The clinical features of facioscapulohumeral muscular dystrophy. Pract Neurol 16, 201-207.
2. Weemaes, C.M., van Tol, M.J., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C., Thijssen, P.E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E., et al. (2013). Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. European journal of human genetics : EJHG 21, 1219-1225.
3. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., et al. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics 6, e1001181.
4. Carpenter, N.J., Filipovich, A., Blaese, R.M., Carey, T.L., and Berkel, A.I. (1988). Variable immunodeficiency with abnormal condensation of the heterochromatin of chromosomes 1, 9, and 16. The Journal of pediatrics 112, 757-760.
5. Eickbush, T.H., and Moudrianakis, E.N. (1978). The histone core complex: an octamer assembled by two sets of protein-protein interactions. Biochemistry 17, 4955-4964.
6. McGinty, R.K., and Tan, S. (2015). Nucleosome structure and function. Chem Rev 115, 2255-2273.
7. Mersfelder, E.L., and Parthun, M.R. (2006). The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. Nucleic acids research 34, 2653-2662.
8. Zentner, G.E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. Nature structural & molecular biology 20, 259-266.
9. Buschbeck, M., and Hake, S.B. (2017). Variants of core histones and their roles in cell fate decisions, development and cancer. Nature reviews Molecular cell biology.
10. Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nature structural & molecular biology 20, 267-273.
11. Branco, M.R., Ficz, G., and Reik, W. (2011). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nat Rev Genet 13, 7-13.
12. Jeltsch, A., and Jurkowska, R.Z. (2016). Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. Nucleic acids research 44, 8556-8575.
13. Riggs AD, M.R., Russo VEA. (1996). Epigenetic Mechanisms of Gene Regulation. C.S.H. Cold Spring Harbor Laboratory Press, NY, ed. (
14. Allis, C.D., Jenuwein,T., Reinberg,D., & Caparros,M. (2007). Epigenetics.(Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).
15. Torres, I.O., and Fujimori, D.G. (2015). Functional coupling between writers, erasers and readers of histone and DNA methylation. Curr Opin Struct Biol 35, 68-75.

16. Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. Nat Rev Genet 13, 343-357.

17. Barau, J., Teissandier, A., Zamudio, N., Roy, S., Nalesso, V., Herault, Y., Guillou, F., and Bourc'his, D. (2016). The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. Science 354, 909-912.

18. Rasmussen, K.D., and Helin, K. (2016). Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev 30, 733-750.

19. Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13, 484-492.

20. Masliah-Planchon, J., Bieche, I., Guinebretiere, J.M., Bourdeaut, F., and Delattre, O. (2015). SWI/SNF chromatin remodeling and human malignancies. Annu Rev Pathol 10, 145-171.

21. Yun, M., Wu, J., Workman, J.L., and Li, B. (2011). Readers of histone modifications. Cell research 21, 564-578.

22. Binda, O. (2013). On your histone mark, SET, methylate! Epigenetics : official journal of the DNA Methylation Society 8, 457-463.

23. Du, Q., Luu, P.L., Stirzaker, C., and Clark, S.J. (2015). Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics 7, 1051-1073.

24. Mercer, T.R., and Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. Nature structural & molecular biology 20, 300-307.

25. Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature 472, 120-124.

26. Nishibuchi, G., and Dejardin, J. (2017). The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. Chromosome Res.

27. Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Mol Cell 28, 1-13.

28. Saksouk, N., Simboeck, E., and Dejardin, J. (2015). Constitutive heterochromatin formation and transcription in mammals. Epigenetics & chromatin 8, 3.

29. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.

30. Smit, A.F. (1996). The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6, 743-748.

31. Richard, G.F., Kerrest, A., and Dujon, B. (2008). Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev 72, 686-727.

32. Gemayel, R., Vinces, M.D., Legendre, M., and Verstrepen, K.J. (2010). Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu Rev Genet 44, 445-477.

33. Balog, J., Miller, D., Sanchez-Curtailles, E., Carbo-Marques, J., Block, G., Potman, M., de Knijff, P., Lemmers, R.J., Tapscott, S.J., and van der Maarel, S.M. (2012). Epigenetic regulation of the X-chromosomal macrosatellite repeat encoding for the cancer/testis gene CT47. European journal of human genetics : EJHG 20, 185-191.

34. Warburton, P.E., Hasson, D., Guillem, F., Lescale, C., Jin, X., and Abrusan, G. (2008). Analysis of the largest tandemly repeated DNA families in the human genome. BMC Genomics 9, 533.

35. Koschmann, C., Nunez, F.J., Mendez, F., Brosnan-Cashman, J.A., Meeker, A.K., Lowenstein, P.R., and Castro, M.G. (2017). Mutated Chromatin Regulatory Factors as Tumor Drivers in Cancer. Cancer Res.

36. Portela, A., and Esteller, M. (2010). Epigenetic modifications and human disease. Nat Biotechnol 28, 1057-1068.

37. Landouzy, L., and Dejerine, J. (1885). De la myopathie atrophique progressive. Rev Med 5, 253-366.

38. Deenen, J.C., Arnts, H., van der Maarel, S.M., Padberg, G.W., Verschuuren, J.J., Bakker, E., Weinreich, S.S., Verbeek, A.L., and van Engelen, B.G. (2014). Population-based incidence and prevalence of facioscapulohumeral dystrophy. Neurology 83, 1056-1059.

39. Statland, J.M., and Tawil, R. (2014). Risk of functional impairment in Facioscapulohumeral muscular dystrophy. Muscle Nerve 49, 520-527.

40. Lemmers, R.J., van der Wielen, M.J., Bakker, E., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2004). Somatic mosaicism in FSHD often goes undetected. Annals of neurology 55, 845-850.

41. Tawil, R., Kissel, J.T., Heatwole, C., Pandya, S., Gronseth, G., Benatar, M., Guideline Development, D., Implementation Subcommittee of the American Academy of, N., Practice Issues Review Panel of the American Association of, N., and Electrodiagnostic, M. (2015). Evidence-based guideline summary: Evaluation, diagnosis, and management of facioscapulohumeral muscular dystrophy: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology and the Practice Issues Review Panel of the American Association of Neuromuscular & Electrodiagnostic Medicine. Neurology 85, 357-364.

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

43. Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329, 1650-1653.

44. Lemmers, R.J., van der Vliet, P.J., van der Gaag, K.J., Zuniga, S., Frants, R.R., de Knijff, P., and van der Maarel, S.M. (2010). Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete interchromosomal sequence transfers in human evolution. American journal of human genetics 86, 364-377.

45. Snider, L., Asawachaicharn, A., Tyler, A.E., Geng, L.N., Petek, L.M., Maves, L., Miller, D.G., Lemmers, R.J., Winokur, S.T., Tawil, R., et al. (2009). RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. Human molecular genetics 18, 2414-2430.

46. Zeng, W., de Greef, J.C., Chen, Y.Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A., et al. (2009). Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS genetics 5, e1000559.

47. Hewitt, J.E., Lyle, R., Clark, L.N., Valleley, E.M., Wright, T.J., Wijmenga, C., van Deutekom, J.C., Francis, F., Sharpe, P.T., Hofker, M., et al. (1994). Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Human molecular genetics 3, 1287-1295.

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

49. Balog, J., Thijssen, P.E., de Greef, J.C., Shah, B., van Engelen, B.G., Yokomori, K., Tapscott, S.J., Tawil, R., and van der Maarel, S.M. (2012). Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics : official journal of the DNA Methylation Society 7, 579-584.

50. de Greef, J.C., Lemmers, R.J., Camano, P., Day, J.W., Sacconi, S., Dunand, M., van Engelen, B.G., Kiuru-Enari, S., Padberg, G.W., Rosa, A.L., et al. (2010). Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology 75, 1548-1554.

51. Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nature genetics 2, 26-30.

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

53. Scionti, I., Fabbri, G., Fiorillo, C., Ricci, G., Greco, F., D'Amico, R., Termanini, A., Vercelli, L., Tomelleri, G., Cao, M., et al. (2012). Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. Journal of medical genetics 49, 171-178.

54. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2014). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics.

55. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics 44, 1370-1374.

56. Sacconi, S., Lemmers, R.J., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. American journal of human genetics 93, 744-751.

57. Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., and Muller, C.R. (2014). Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. European journal of human genetics : EJHG.

58. Mitsuhashi, S., Boyden, S.E., Estrella, E.A., Jones, T.I., Rahimov, F., Yu, T.W., Darras, B.T., Amato, A.A., Folkerth, R.D., Jones, P.L., et al. (2013). Exome sequencing identifies a novel SMCHD1 mutation in facioscapulohumeral muscular dystrophy 2. Neuromuscular disorders : NMD 23, 975-980.

59. Winston, J., Duerden, L., Mort, M., Frayling, I.M., Rogers, M.T., and Upadhyaya, M. (2014). Identification of two novel SMCHD1 sequence variants in families with FSHD-like muscular dystrophy. European journal of human genetics : EJHG.

60. Hirano, T. (2005). SMC proteins and chromosome mechanics: from bacteria to humans. Philosophical transactions of the Royal Society of London Series B, Biological sciences 360, 507-514.

61. Brideau, N.J., Coker, H., Gendrel, A.V., Siebert, C.A., Bezstarosti, K., Demmers, J., Poot, R.A., Nesterova, T.B., and Brockdorff, N. (2015). Independent Mechanisms Target SMCHD1 to Trimethylated Histone H3 Lysine 9-Modified Chromatin and the Inactive X Chromosome. Molecular and cellular biology 35, 4053-4068.

62. Chen, K., Czabotar, P.E., Blewitt, M.E., and Murphy, J.M. (2016). The hinge domain of the epigenetic repressor Smchd1 adopts an unconventional homodimeric configuration. Biochem J 473, 733-742.

63. Blewitt, M.E., Vickaryous, N.K., Hemley, S.J., Ashe, A., Bruxner, T.J., Preis, J.I., Arkell, R., and Whitelaw, E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proceedings of the National Academy of Sciences of the United States of America 102, 7629-7634.

64. Blewitt, M.E., Gendrel, A.V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., et al. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nature genetics 40, 663-669.

65. Wutz, A. (2011). Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. Nat Rev Genet 12, 542-553.

66. Gendrel, A.V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Developmental cell 23, 265-279.

67. Gendrel, A.V., Tang, Y.A., Suzuki, M., Godwin, J., Nesterova, T.B., Greally, J.M., Heard, E., and Brockdorff, N. (2013). Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Molecular and cellular biology 33, 3150-3165.

68. Nozawa, R.S., Nagao, K., Igami, K.T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nature structural & molecular biology 20, 566-573.

69. Mould, A.W., Pang, Z., Pakusch, M., Tonks, I.D., Stark, M., Carrie, D., Mukhopadhyay, P., Seidel, A., Ellis, J.J., Deakin, J., et al. (2013). Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. Epigenetics & chromatin 6, 19.

70. Liu, R., Chen, K., Jansz, N., Blewitt, M.E., and Ritchie, M.E. (2016). Transcriptional profiling of the epigenetic regulator Smchd1. Genom Data 7, 144-147.

71. Massah, S., Hollebakken, R., Labrecque, M.P., Kolybaba, A.M., Beischlag, T.V., and Prefontaine, G.G. (2014). Epigenetic characterization of the growth hormone gene identifies SmcHD1 as a regulator of autosomal gene clusters. PloS one 9, e97535.

72. Balog, J., Thijssen, P.E., Shadle, S., Straasheijm, K.R., van der Vliet, P.J., Krom, Y.D., van den Boogaard, M.L., de Jong, A., RJ, F.L., Tawil, R., et al. (2015). Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4. Epigenetics : official journal of the DNA Methylation Society 10, 1133-1142.

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

74. Cabianca, D.S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y., and Gabellini, D. (2012). A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell 149, 819-831.

75. Lim, J.W., Snider, L., Yao, Z., Tawil, R., Van Der Maarel, S.M., Rigo, F., Bennett, C.F., Filippova, G.N., and Tapscott, S.J. (2015). DICER/AGO-dependent epigenetic silencing of D4Z4 repeats enhanced by exogenous siRNA suggests mechanisms and therapies for FSHD. Human molecular genetics 24, 4817-4828.

76. Sharma, V., Pandey, S.N., Khawaja, H., Brown, K.J., Hathout, Y., and Chen, Y.W. (2016). PARP1 Differentially Interacts with Promoter region of DUX4 Gene in FSHD Myoblasts. J Genet Syndr Gene Ther 7.

77. Das, S., and Chadwick, B.P. (2016). Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. PloS one 11, e0160022.

78. Jones, T.I., Chen, J.C., Rahimov, F., Homma, S., Arashiro, P., Beermann, M.L., King, O.D., Miller, J.B., Kunkel, L.M., Emerson, C.P., Jr., et al. (2012). Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. Human molecular genetics 21, 4419-4430.

79. Rickard, A.M., Petek, L.M., and Miller, D.G. (2015). Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Human molecular genetics 24, 5901-5914.

80. Clapp, J., Mitchell, L.M., Bolland, D.J., Fantes, J., Corcoran, A.E., Scotting, P.J., Armour, J.A., and Hewitt, J.E. (2007). Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. American journal of human genetics 81, 264-279.

81. Leidenroth, A., Clapp, J., Mitchell, L.M., Coneyworth, D., Dearden, F.L., Iannuzzi, L., and Hewitt, J.E. (2012). Evolution of DUX gene macrosatellites in placental mammals. Chromosoma 121, 489-497.

82. Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O., et al. (2007). The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscular disorders : NMD 17, 611-623.

83. Wallace, L.M., Garwick, S.E., Mei, W., Belayew, A., Coppee, F., Ladner, K.J., Guttridge, D., Yang, J., and Harper, S.Q. (2011). DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. Annals of neurology 69, 540-552.

84. Wuebbles, R.D., Long, S.W., Hanel, M.L., and Jones, P.L. (2010). Testing the effects of FSHD candidate gene expression in vertebrate muscle development. Int J Clin Exp Pathol 3, 386-400.

85. Jones, T.I., Parilla, M., and Jones, P.L. (2016). Transgenic Drosophila for Investigating DUX4 and FRG1, Two Genes Associated with Facioscapulohumeral Muscular Dystrophy (FSHD). PloS one 11, e0150938.

86. Feng, Q., Snider, L., Jagannathan, S., Tawil, R., van der Maarel, S.M., Tapscott, S.J., and Bradley, R.K. (2015). A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. Elife 4.

87. Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., et al. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Developmental cell 22, 38-51.

88. Geng, L.N., Tyler, A.E., and Tapscott, S.J. (2011). Immunodetection of human double homeobox 4. Hybridoma (Larchmt) 30, 125-130.

89. Choi, S.H., Gearhart, M.D., Cui, Z., Bosnakovski, D., Kim, M., Schennum, N., and Kyba, M. (2016). DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. Nucleic acids research 44, 5161-5173.

90. Yao, Z., Snider, L., Balog, J., Lemmers, R.J., Van Der Maarel, S.M., Tawil, R., and Tapscott, S.J. (2014). DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. Human molecular genetics 23, 5342-5352.

91. Dmitriev, P., Bou Saada, Y., Dib, C., Ansseau, E., Barat, A., Hamade, A., Dessen, P., Robert, T., Lazar, V., Louzada, R.A., et al. (2016). DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. Free Radic Biol Med 99, 244-258.

92. Homma, S., Beermann, M.L., Boyce, F.M., and Miller, J.B. (2015). Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. Ann Clin Transl Neurol 2, 151-166.

93. Homma, S., Beermann, M.L., Yu, B., Boyce, F.M., and Miller, J.B. (2016). Nuclear bodies reorganize during myogenesis in vitro and are differentially disrupted by expression of FSHD-associated DUX4. Skeletal muscle 6, 42.

94. Ferri, G., Huichalaf, C.H., Caccia, R., and Gabellini, D. (2015). Direct interplay between two candidate genes in FSHD muscular dystrophy. Human molecular genetics 24, 1256-1266.

95. Mariot, V., Roche, S., Hourde, C., Portilho, D., Sacconi, S., Puppo, F., Duguez, S., Rameau, P., Caruso, N., Delezoide, A.L., et al. (2015). Correlation between low FAT1 expression and early affected muscle in facioscapulohumeral muscular dystrophy. Annals of neurology 78, 387-400.

96. Thijssen, P.E., Balog, J., Yao, Z., Pham, T.P., Tawil, R., Tapscott, S.J., and Van der Maarel, S.M. (2014). DUX4 promotes transcription of FRG2 by directly activating its promoter in facioscapulohumeral muscular dystrophy. Skeletal muscle 4, 19.

97. Deak, K.L., Lemmers, R.J., Stajich, J.M., Klooster, R., Tawil, R., Frants, R.R., Speer, M.C., van der Maarel, S.M., and Gilbert, J.R. (2007). Genotype-phenotype study in an FSHD family with a proximal deletion encompassing p13E-11 and D4Z4. Neurology 68, 578-582.

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

99. Lin, F., Wang, Z.Q., Lin, M.T., Murong, S.X., and Wang, N. (2015). New Insights into Genotype-phenotype Correlations in Chinese Facioscapulohumeral Muscular Dystrophy: A Retrospective Analysis of 178 Patients. Chin Med J (Engl) 128, 1707-1713.

100. Park, H.J., Hong, J.M., Lee, J.H., Lee, H.S., Shin, H.Y., Kim, S.M., Ki, C.S., Lee, J.H., and Choi, Y.C. (2015). Low D4Z4 copy number and gender difference in Korean patients with facioscapulohumeral muscular dystrophy type 1. Neuromuscular disorders : NMD 25, 859-864.

101. Statland, J.M., Donlin-Smith, C.M., Tapscott, S.J., Lemmers, R.J., van der Maarel, S.M., and Tawil, R. (2015). Milder phenotype in facioscapulohumeral dystrophy with 7-10 residual D4Z4 repeats. Neurology 85, 2147-2150.

102. Statland, J.M., Shah, B., Henderson, D., Van Der Maarel, S., Tapscott, S.J., and Tawil, R. (2015). Muscle pathology grade for facioscapulohumeral muscular dystrophy biopsies. Muscle Nerve 52, 521-526.

103. Nikolic, A., Ricci, G., Sera, F., Bucci, E., Govi, M., Mele, F., Rossi, M., Ruggiero, L., Vercelli, L., Ravaglia, S., et al. (2016). 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.

104. Gaillard, M.C., Roche, S., Dion, C., Tasmadjian, A., Bouget, G., Salort-Campana, E., Vovan, C., Chaix, C., Broucqsault, N., Morere, J., et al. (2014). Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. Neurology 83, 733-742.

105. Jones, T.I., King, O.D., Himeda, C.L., Homma, S., Chen, J.C., Beermann, M.L., Yan, C., Emerson, C.P., Jr., Miller, J.B., Wagner, K.R., et al. (2015). Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. Clinical epigenetics 7, 37.

106. Calandra, P., Cascino, I., Lemmers, R.J., Galluzzi, G., Teveroni, E., Monforte, M., Tasca, G., Ricci, E., Moretti, F., van der Maarel, S.M., et al. (2016). Allele-specific DNA hypomethylation characterises FSHD1 and FSHD2. Journal of medical genetics 53, 348-355.

107. Hartweck, L.M., Anderson, L.J., Lemmers, R.J., Dandapat, A., Toso, E.A., Dalton, J.C., Tawil, R., Day, J.W., van der Maarel, S.M., and Kyba, M. (2013). A focal domain of extreme demethylation within D4Z4 in FSHD2. Neurology 80, 392-399.

108. Hulten, M. (1978). Selective Somatic Pairing and Fragility at 1q12 in a Boy with Common Variable Immuno Deficiency. Clinical Genetics 14, 294.

109. Tiepolo, L., Maraschio, P., Gimelli, G., Cuoco, C., Gargani, G.F., and Romano, C. (1979). Multibranched chromosomes 1, 9, and 16 in a patient with combined IgA and IgE deficiency. Hum Genet 51, 127-137.

110. Maraschio, P., Zuffardi, O., Dalla Fior, T., and Tiepolo, L. (1988). Immunodeficiency, centromeric heterochromatin instability of chromosomes 1, 9, and 16, and facial anomalies: the ICF syndrome. Journal of medical genetics 25, 173-180.

111. Blanco-Betancourt, C.E., Moncla, A., Milili, M., Jiang, Y.L., Viegas-Pequignot, E.M., Roquelaure, B., Thuret, I., and Schiff, C. (2004). Defective B-cell-negative selection and terminal differentiation in the ICF syndrome. Blood 103, 2683-2690.

112. Gennery, A.R., Slatter, M.A., Bredius, R.G., Hagleitner, M.M., Weemaes, C., Cant, A.J., and Lankester, A.C. (2007). Hematopoietic stem cell transplantation corrects the immunologic abnormalities associated with immunodeficiency-centromeric instability-facial dysmorphism syndrome. Pediatrics 120, e1341-1344.

113. Harnisch, E., Buddingh, E.P., Thijssen, P.E., Brooks, A.S., Driessen, G.J., Kersseboom, R., and Lankester, A.C. (2016). Hematopoietic Stem Cell Transplantation in a Patient With ICF2 Syndrome Presenting With EBV-Induced Hemophagocytic Lymphohystiocytosis. Transplantation 100, e35-36.

114. Ehrlich, M., Jackson, K., and Weemaes, C. (2006). Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF). Orphanet J Rare Dis 1, 2.

115. Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A., and Viegas-Pequignot, E. (1993). An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. Human molecular genetics 2, 731-735.

116. Miniou, P., Jeanpierre, M., Blanquet, V., Sibella, V., Bonneau, D., Herbelin, C., Fischer, A., Niveleau, A., and Viegas-Pequignot, E. (1994). Abnormal methylation pattern in constitutive and facultative (X inactive chromosome) heterochromatin of ICF patients. Human molecular genetics 3, 2093-2102.

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

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

119. Jiang, Y.L., Rigolet, M., Bourc'his, D., Nigon, F., Bokesoy, I., Fryns, J.P., Hulten, M., Jonveaux, P., Maraschio, P., Megarbane, A., et al. (2005). DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. Hum Mutat 25, 56-63.

120. Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., et al. (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nat Commun 6, 7870.

121. Baubec, T., Colombo, D.F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A.R., Akalin, A., and Schubeler, D. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243-247.

122. Noh, K.M., Allis, C.D., and Li, H. (2015). Reading between the Lines: "ADD"-ing Histone and DNA Methylation Marks toward a New Epigenetic "Sum". ACS chemical biology.

123. Gowher, H., and Jeltsch, A. (2002). Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. J Biol Chem 277, 20409-20414.

124. Moarefi, A.H., and Chedin, F. (2011). ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. J Mol Biol 409, 758-772.

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

126. Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K., and Li, E. (2006). Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. Development 133, 1183-1192.

127. Velasco, G., Hube, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Pequignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proceedings of the National Academy of Sciences of the United States of America 107, 9281-9286.

128. Youngson, N.A., Epp, T., Roberts, A.R., Daxinger, L., Ashe, A., Huang, E., Lester, K.L., Harten, S.K., Kay, G.F., Cox, T., et al. (2013). No evidence for cumulative effects in a Dnmt3b hypomorph across multiple generations. Mammalian genome : official journal of the International Mammalian Genome Society 24, 206-217.

129. Hansen, R.S. (2003). X inactivation-specific methylation of LINE-1 elements by DNMT3B: implications for the Lyon repeat hypothesis. Human molecular genetics 12, 2559-2567.

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

131. Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K., Wang, J., den Hamer, B., Aytekin, C., van der Maarel, S.M., and Daxinger, L. (2016). Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. Human molecular genetics.

132. Liang, J., Yan, R., Chen, G., Feng, J., Wu, W.W., Ren, W., Zhu, C., Zhao, Y., Gao, X.M., and Wang, J. (2016). Downregulation of ZBTB24 hampers the G0/1- to S-phase cell-cycle transition via upregulating the expression of IRF-4 in human B cells. Genes Immun.

133. Nitta, H., Unoki, M., Ichiyanagi, K., Kosho, T., Shigemura, T., Takahashi, H., Velasco, G., Francastel, C., Picard, C., Kubota, T., et al. (2013). Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. J Hum Genet 58, 455-460.

134. Prescott, J.E., Osthus, R.C., Lee, L.A., Lewis, B.C., Shim, H., Barrett, J.F., Guo, Q., Hawkins, A.L., Griffin, C.A., and Dang, C.V. (2001). A novel c-Myc-responsive gene, JPO1, participates in neoplastic transformation. J Biol Chem 276, 48276-48284.

135. Gill, R.M., Gabor, T.V., Couzens, A.L., and Scheid, M.P. (2013). The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation. Molecular and cellular biology 33, 498-513.

136. Guiu, J., Bergen, D.J., De Pater, E., Islam, A.B., Ayllon, V., Gama-Norton, L., Ruiz-Herguido, C., Gonzalez, J., Lopez-Bigas, N., Menendez, P., et al. (2014). Identification of Cdca7 as a novel Notch transcriptional target involved in hematopoietic stem cell emergence. J Exp Med 211, 2411-2423.

137. Geiman, T.M., and Muegge, K. (2000). Lsh, an SNF2/helicase family member, is required for proliferation of mature T lymphocytes. Proceedings of the National Academy of Sciences of the United States of America 97, 4772-4777.

138. Myant, K., and Stancheva, I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. Molecular and cellular biology 28, 215-226.

139. Ren, J., Briones, V., Barbour, S., Yu, W., Han, Y., Terashima, M., and Muegge, K. (2015). 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 research 43, 1444-1455.

140. Zhu, H., Geiman, T.M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E., and Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. The EMBO journal 25, 335-345.

141. Simo-Riudalbas, L., Diaz-Lagares, A., Gatto, S., Gagliardi, M., Crujeiras, A.B., Matarazzo, M.R., Esteller, M., and Sandoval, J. (2015). Genome-Wide DNA Methylation Analysis Identifies Novel Hypomethylated Non-Pericentromeric Genes with Potential Clinical Implications in ICF Syndrome. PloS one 10, e0132517.

142. Jin, B., Tao, Q., Peng, J., Soo, H.M., Wu, W., Ying, J., Fields, C.R., Delmas, A.L., Liu, X., Qiu, J., et al. (2008). DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. Human molecular genetics 17, 690-709.

143. Velasco, G., Walton, E.L., Sterlin, D., Hedouin, S., Nitta, H., Ito, Y., Fouyssac, F., Megarbane, A., Sasaki, H., Picard, C., et al. (2014). 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.

144. Huang, K., Wu, Z., Liu, Z., Hu, G., Yu, J., Chang, K.H., Kim, K.P., Le, T., Faull, K.F., Rao, N., et al. (2014). Selective demethylation and altered gene expression are associated with ICF syndrome in human-induced pluripotent stem cells and mesenchymal stem cells. Human molecular genetics 23, 6448-6457.

145. Heyn, H., Vidal, E., Sayols, S., Sanchez-Mut, J.V., Moran, S., Medina, I., Sandoval, J., Simo-Riudalbas, L., Szczesna, K., Huertas, D., et al. (2012). Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. Epigenetics : official journal of the DNA Methylation Society 7, 542-550.

146. Yehezkel, S., Segev, Y., Viegas-Pequignot, E., Skorecki, K., and Selig, S. (2008). Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. Human molecular genetics 17, 2776-2789.

147. Deng, Z., Campbell, A.E., and Lieberman, P.M. (2010). TERRA, CpG methylation and telomere heterochromatin: lessons from ICF syndrome cells. Cell Cycle 9, 69-74.

148. Sagie, S., Ellran, E., Katzir, H., Shaked, R., Yehezkel, S., Laevsky, I., Ghanayim, A., Geiger, D., Tzukerman, M., and Selig, S. (2014). Induced pluripotent stem cells as a model for telomeric abnormalities in ICF type I syndrome. Human molecular genetics 23, 3629-3640.

149. Jefferson, A., Colella, S., Moralli, D., Wilson, N., Yusuf, M., Gimelli, G., Ragoussis, J., and Volpi, E.V. (2010). Altered intra-nuclear organisation of heterochromatin and genes in ICF syndrome. PloS one 5, e11364.

150. Matarazzo, M.R., Boyle, S., D'Esposito, M., and Bickmore, W.A. (2007). Chromosome territory reorganization in a human disease with altered DNA methylation. Proceedings of the National Academy of Sciences of the United States of America 104, 16546-16551.

151. Hassan, K.M., Norwood, T., Gimelli, G., Gartler, S.M., and Hansen, R.S. (2001). Satellite 2 methylation patterns in normal and ICF syndrome cells and association of hypomethylation with advanced replication. Hum Genet 109, 452-462.

152. Hansen, R.S., Stoger, R., Wijmenga, C., Stanek, A.M., Canfield, T.K., Luo, P., Matarazzo, M.R., D'Esposito, M., Feil, R., Gimelli, G., et al. (2000). Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Human molecular genetics 9, 2575-2587.

153. Lana, E., Megarbane, A., Tourriere, H., Sarda, P., Lefranc, G., Claustres, M., and De Sario, A. (2012). DNA replication is altered in Immunodeficiency Centromeric instability Facial anomalies (ICF) cells carrying DNMT3B mutations. European journal of human genetics : EJHG 20, 1044-1050.

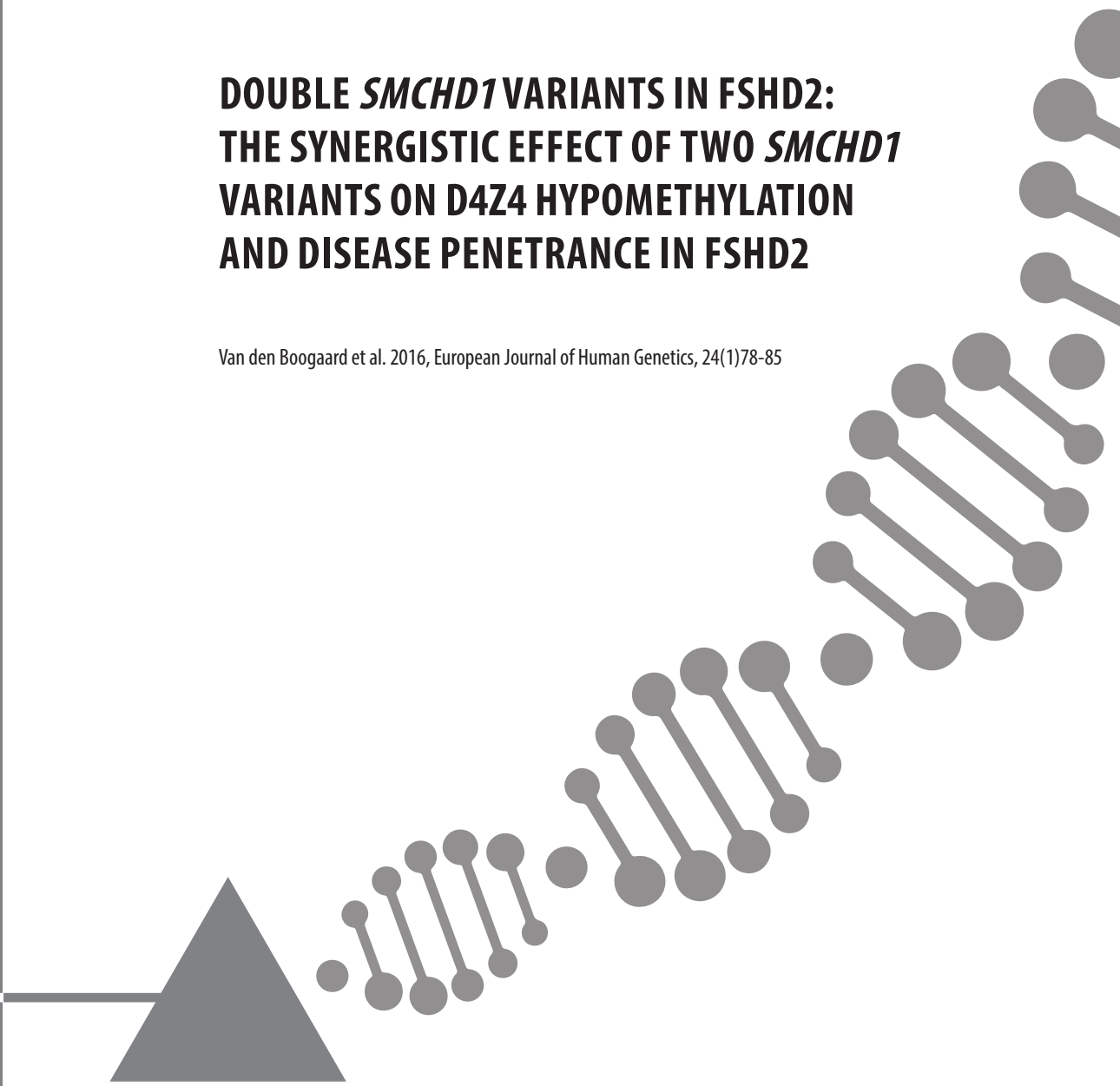
Marlinde L van den Boogaard[1], Richard JFL Lemmers[1], Pilar Camaño[2], Patrick J van der Vliet[1], Nicol Voermans[3], Baziel GM van Engelen[3], Adolfo López de Munain[2], Stephen J Tapscott[4], Nienke van der Stoep[5], Rabi Tawil[6] and Silvère van der Maarel[1]

1 Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
2 Neurosciences, BioDonostia Health Research Institute, Hospital Donostia, San Sebastián, Spain
3 Department of Neurology, Radboud University Medical Centre, Nijmegen, The Netherlands
4 Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, USA
5 Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
6 Department of Neurology, University of Rochester Medical Center, Rochester, USA

2

DOUBLE *SMCHD1* VARIANTS IN FSHD2: THE SYNERGISTIC EFFECT OF TWO *SMCHD1* VARIANTS ON D4Z4 HYPOMETHYLATION AND DISEASE PENETRANCE IN FSHD2

Van den Boogaard et al. 2016, European Journal of Human Genetics, 24(1)78-85



ABSTRACT

Facioscapulohumeral muscular dystrophy (FSHD) predominantly affects the muscles in the face, trunk and upper extremities and is marked by large clinical variability in disease onset and progression. FSHD is associated with partial chromatin relaxation of the D4Z4 repeat array on chromosome 4 and the somatic expression of the D4Z4 encoded *DUX4* gene. The most common form, FSHD1, is caused by a contraction of the D4Z4 repeat array on chromosome 4 to a size of 1-10 units. FSHD2, the less common form of FSHD, is most often caused by heterozygous variants in the chromatin modifier *SMCHD1*, which is involved in the maintenance of D4Z4 methylation. We identified three families in which the proband carries two potentially damaging *SMCHD1* variants. We investigated whether these variants were located *in cis* or *in trans* and determined their functional consequences by detailed clinical information and D4Z4 methylation studies. In the first family both variants *in trans* were shown to act synergistically on D4Z4 hypomethylation and disease penetrance, in the second family both *SMCHD1* function-affecting variants were located *in cis* while in the third family one of the two variants did not affect function. This study demonstrates that having two *SMCHD1* missense variants that affect function is compatible with life in males and females, which is remarkable considering its role in X inactivation in mice. The study also highlights the variability in *SMCHD1* variants underlying FSHD2 and the predictive value of D4Z4 methylation analysis in determining the functional consequences of *SMCHD1* variants of unknown significance.

INTRODUCTION

Facioscapulohumeral muscular dystrophy (FSHD; OMIM 158900) is a common myopathy in adults, with a recently reported prevalence of ~1:8 000[1]. FSHD is clinically characterized by weakness of the facial, shoulder girdle, trunk and upper arm muscles, which can be asymmetric, and progresses to involve humeral, anterior lower leg muscles and pelvic girdle muscles[2]. The onset of the disease is typically in the second decade of life, but the disease progression and severity are highly variable[3].

The genetic forms identified thus far, FSHD1 and FSHD2, are clinically indistinguishable[4]. Both forms are associated with partial chromatin relaxation of the D4Z4 macrosatellite repeat array on the subtelomere of the long arm of chromosome 4 and transcriptional derepression of the D4Z4 unit-encoded *DUX4* retrogene in skeletal muscle[5-9]. DUX4 is a germ line transcription factor that is normally repressed in somatic cells[7]. Its expression in skeletal muscle activates genes involved in germ line and early stem cell development, as well as specific classes of repeat elements, and overexpression of DUX4 in somatic cells causes cell death[10-12].

To cause FSHD, D4Z4 chromatin relaxation must occur on a specific genetic background of chromosome 4 (most often 4A161) that facilitates the production of stable DUX4 mRNA due to the presence of a polymorphic DUX4 polyadenylation signal distal to the D4Z4 repeat array[8; 13]. D4Z4 chromatin relaxation on non-permissive chromosomes lacking a *DUX4* polyadenylation signal, do not cause FSHD in the absence of detectable levels of DUX4 mRNA[8; 14].

Autosomal dominant FSHD1 is the most common form of FSHD (>95%), in which D4Z4 chromatin relaxation and *DUX4* expression are caused by a contraction of the D4Z4 repeat array to a size of 1-10 units[15; 16]. In the uncommon form of FSHD (FSHD2), D4Z4 chromatin relaxation occurs in the absence of D4Z4 repeat array contraction[5]. In FSHD1 chromatin relaxation and CpG hypomethylation are restricted to the contracted allele, whereas in FSHD2 chromatin relaxation and CpG hypomethylation occur at the D4Z4 repeat arrays of both copies of chromosome 4, and in the highly homologous repeat arrays on chromosome 10[4; 9].

Heterozygous variants in the *structural maintenance of chromosomes flexible hinge domain containing 1* (*SMCHD1*) gene on chromosome 18 account for the majority of FSHD2 cases[14]. SMCHD1 is an atypical member of the *SMC* gene superfamily, a family of proteins which is involved in chromosome condensation and cohesion, genome maintenance and gene regulation[17-19]. Chromatin immunoprecipitation (ChIP) studies showed the presence of SMCHD1 on the D4Z4 array, and its reduced binding to D4Z4 in *SMCHD1* mutation carriers[14]. *SMCHD1* variants can also modify the disease severity in FSHD1 families, explaining some of the clinical variability seen in FSHD[20].

As a measure of D4Z4 chromatin relaxation, often D4Z4 methylation is used. We have established a reliable and informative measure of D4Z4 methylation by measuring the methylation of all D4Z4 arrays simultaneously at a unique methylation-sensitive restriction site (*Fse*I) in the D4Z4 unit[14]. The methylation level at this site is significantly lower in FSHD2 compared to both FSHD1 and controls, and a threshold of 25% was established for FSHD2[14]. Recently we showed that D4Z4 methylation level at this site is repeat array size-dependent[21]. We introduced a new methylation parameter, Delta1, which represents the difference between the experimentally observed methylation and the predicted methylation level based on repeat size in controls. In *SMCHD1* mutation carriers the average Delta1 score is highly negative ranging between -20 and -45 suggesting a strong contribution of the variant to D4Z4 hypomethylation[21]. Accordingly, a second model was then fitted to predict the methylation in *SMCHD1* variant carriers, which resulted in the Delta2 score. For *SMCHD1* variants that preserve the open reading frame (ORF) a mean Delta2 score of -1.8% was found, which is significantly lower than the mean Delta2 of ORF-disrupting variants (mean 2.7%). This suggests that ORF-preserving variants are more deleterious for the maintenance of a repressed D4Z4 chromatin state in somatic cells[21].

Several studies have so far identified disease causing variants in *SMCHD1* in approximately 70 FSHD2 families. Heterozygous disease causing *SMCHD1* variants are distributed over the entire *SMCHD1* locus and include splice site, insertion-deletion, missense and nonsense variants[14; 20-24]. We have shown that a combination of the size of the permissive D4Z4 array and the type of *SMCHD1* variant together determine the epigenetic susceptibility to disease presentation[21]. Until now, only heterozygous *SMCHD1* variants, which are dominant in combination with a permissive haplotype, have been reported. Here we describe three families in which two *SMCHD1* variants which potentially affect function were identified in each proband. For each family we investigated whether the variants were located *in cis* or *in trans* and analyzed if both variants were contributing to D4Z4 hypomethylation independently and what the effect was on the FSHD phenotype.

MATERIAL AND METHODS

Subjects

Three families (Fig. 1) were studied after informed consent and the study protocol was approved by the relevant institutional review boards. Clinical assessment of disease severity was performed using the 10 point (0: unaffected – 10: wheelchair bound) standardized Clinical Severity Score (CSS)[25].

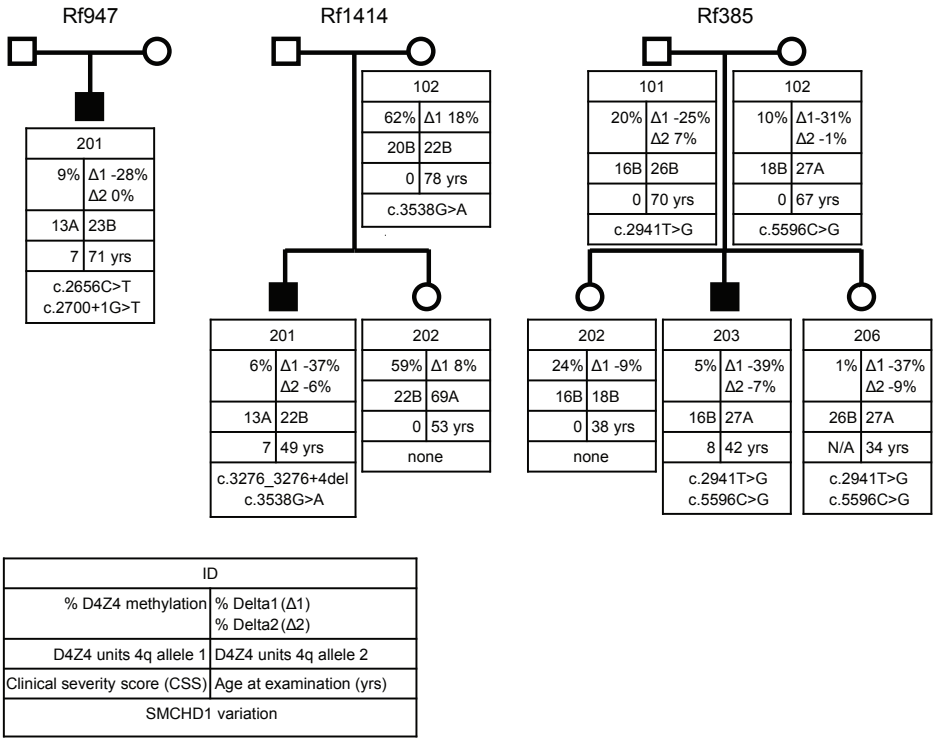


FIGURE 1. Pedigrees of the three FSHD2 families presented in this study. Shown are families with information about D4Z4 methylation, Delta scores, sizes of 4q-linked D4Z4 repeats, clinical severity scores at age of examination, and *SMCHD1* variants. Key is shown below. N/A: not available.

The first family (Rf947) consists of a single patient living in the United States of America. We were not able to get additional family information from this individual. The second family (Rf1414) is a Dutch family, with one individual diagnosed with FSHD. After the identification of the two *SMCHD1* variants in the proband, additional clinical information and blood samples from two relatives could be collected. The third family (Rf385) is a Spanish family, with one individual diagnosed with FSHD and four relatives.

D4Z4 repeat sizing, haplotype analysis and methylation analysis

For genotyping high quality genomic DNA was isolated from peripheral blood mononucleated cells (PBMCs). The sizing of the D4Z4 repeats on chromosomes 4 and 10 was done by pulsed field gel electrophoresis (PFGE) as described previously[8]. Haplotype analysis was done by hybridization of PFGE blots with probes A and B in combination with SSLP analysis according to previously described protocols[8].

For D4Z4 methylation analysis genomic DNA was double digested with *Eco*RI (Thermo Fisher Scientific inc., Waltham, MA, US) and *Bgl*II (Fermentas, Thermo Fisher Scientific) overnight at 37 °C, and cleaved DNA was purified using PCR extraction columns (NucleoSpin® Gel and PCR Clean-up, Machery-Nagel/BIOKÉ, Leiden, The Netherlands). Purified DNA was then digested with *Fse*I (New England Biolabs/BIOKÉ) for ≥4 h, separated by size on 0.8% agarose gels, transferred to a nylon membrane (Hybond XL, Amersham, GE Healthcare, Diegem, Belgium) by Southern blotting and probed using the p13E-11 ^{32}P labeled probe. Probe signals were quantified using the Storm 820 Phosphorimager (Amersham) and ImageQuant™ TL software (Amersham). The signal from the 4,061-bp fragment (methylated) was divided by the total amount of hybridizing fragments at 4,061 bp and 3,387 bp (unmethylated) to yield the average percentage of methylated *Fse*I sites within the most proximal D4Z4 unit on all four D4Z4 arrays. The Delta1 and Delta2 scores were calculated as described in Lemmers et al. 2014.

SMCHD1 variant analysis

For the index cases *SMCHD1* variant analysis was performed by Sanger sequencing after PCR amplification of all coding exons using intronic primers at a position of at least 50 nucleotides from the splice donor or acceptor site. The *SMCHD1* genomic sequence was obtained from Ensemble (build 37) [GRCh37:18:2655286:2805615] (Genomic Refseq: NG_031972.1, Transcript Refseq: NM_015295.2). Exons were numbered like in NG_031972.1 and primers were published previously[21]. The functional consequences of variants were predicted using Alamut Visual version 2.4 (Interactive Biosoftware, Rouen, France). Identified variants are submitted to the Leiden Open Variation Database (http://databases.lovd.nl/shared/individuals/SMCHD1: submission IDs 00028967- 00028973)

For individual 385-203 Sanger sequencing of exons 24 and 45 in relatives was used to identify if the two *SMCHD1* variants were located *in cis* or *in trans*. To identify if the two *SMCHD1* variants found in individual 947-201 and individual 1414-201 were present on different alleles, both alleles were PCR amplified and then cloned in a TOPO vector (Zero Blunt® TOPO® PCR Cloning Kit, Invitrogen by Life technologies, Bleiswijk, The Netherlands).

For individual 947-201 *SMCHD1* exon 21 was PCR amplified and cloned in pCR™-Blunt II-TOPO® vector and transformed in DH5α heat shock competent cells (Subcloning Efficiency™ DH5α™ Competent Cells, Invitrogen™, Life Technologies). Multiple clones were analyzed, by sequencing their insert, to find clones containing either of the alleles.

For individual 1414-201 a long range PCR (Phusion® High-Fidelity DNA Polymerase, Phusion® GC Buffer Pack, New England Biolabs) from *SMCHD1* exon 25 to exon 28 was performed (primers 25F + 28R, product 8708 bp), and product ends were ligated using T4 DNA Ligase (Thermo Scientific). An additional PCR was performed using the *SMCHD1*

exon 28F and exon 25R primers to amplify the artificially fused exons 25 and 28, this PCR product was cloned in the pCR™-Blunt II-TOPO® vector and transformed in DH5α heat shock competent cells. Multiple clones were analyzed, by sequencing their insert, to find clones containing either of the amplified alleles.

RNA analysis

From individual 947-201 RNA was isolated from a PAXgene Blood RNA Tube using the PAXgene Blood RNA Kit (PreAnalytiX, GmbH, Hombrechtikon, Switzerland). cDNA was synthesized using the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) using random hexamer primers. Reverse transcriptase PCR was performed using primers 2481F (5'-CATGGAGGAAAATGGCCTTA-3') and 2981R (5'-TTCAGTCGACGAGGGTGAC-3') located in exon 18 and exon 23, respectively. Subsequently, PCR products were separated by size on 2% agarose gels and PCR products were gel purified and analyzed by Sanger sequencing.

RESULTS

Two *SMCHD1* variants *in cis* in Rf947

Individual 201 of Rf947 was suspected of FSHD based on physical examination with a clinical severity score (CSS) of 7 at the age of 71 (Fig. 1)[25]. Genetic analysis showed that he carries a permissive allele of 13 D4Z4 units and the D4Z4 methylation analysis revealed a methylation level of 9% (Delta1 score -28%), which is consistent with the diagnosis of FSHD2. *SMCHD1* Sanger sequencing identified two *SMCHD1* variants in exon 21 in individual 947-201 (Table 1). The first variant (c.2656C>T p.Arg886*) is a nonsense substitution that is predicted to result in a premature stop codon. The second variant (c.2700+1G>T) is a splice donor site variant, which is predicted to result in the skip of exon 21 by multiple splicing predictors (MaxEnt, NNSPLICE and HSF). Both variants have not been reported previously.

To investigate if the two variants in individual 947-201 were located *in cis* or *in trans*, genomic DNA of *SMCHD1* exon 21 was PCR amplified and cloned. Sanger sequencing of individual exon 21 PCR clones showed that both variants were located on the same allele (Fig. 2A). cDNA analysis by an RT-PCR from *SMCHD1* exon 19 to exon 23, followed by Sanger sequencing showed the presence of two PCR products representing wild type transcript, and the mutant transcript, which lacks exon 21 (Fig. 2B,C). This confirms that the splice donor site variant c.2700+1G>T results in a skip of exon 21 (r.2604_2700del). Skipping of exon 21 neutralizes the p.Arg886* nonsense variant but will result in the disruption of the ORF by a premature stop codon in exon 22. The intensity of the PCR

products suggests that wild type and mutant allele are equally expressed. In contrast to previous findings where disrupting ORF variants seem to result in haploinsufficiency[21], this mutant allele might not be subject to nonsense mediated mRNA decay, and possibly a truncated protein will be produced.

TABLE 1. *SMCHD1* variants identified

Family	Mutation type	Position	Chromosome position (GRCh37.p5)	Transcript position (NM_015295.2) (NG_031972.1)	Protein position (NP_056110.2)	RNA analysis
Rf947	Nonsense	Exon 21	g.2724949C>T	c.2656C>T	p.Arg886*	-
Rf947	5' splice site	Exon 21	g.2724994G>T	c.2700+1G>T	-	r.2604_2700del
Rf1414	5' splice site	Exon 25	g.2732490_2732494del	c.3276_3276+4del	-	-
Rf1414	Missense	Exon 28	g.2740724G>A	c.3538G>A	p.Gly1180Arg	-
Rf385	Missense	Exon 24	g.2729300T>G	c.2941T>G	p.Tyr981Asp	-
Rf385	Missense	Exon 45	g.2784496C>G	c.5596C>G	p.Arg1866Gly	-

Two *SMCHD1* variants *in trans* in Rf1414 of which one affects function

Individual 201 of Rf1414 was suspected of FSHD based on physical examination with a CSS of 7 at the age of 49 (Fig. 1). Genetic analysis showed that he carries a permissive allele of 13 units and the D4Z4 methylation analysis revealed a methylation level of 6% (Delta1 score -37%), supportive of FSHD2.

SMCHD1 Sanger sequencing identified two *SMCHD1* variants in individual 1414-201, one located in exon 25 (c.3276_3276+4del) and the other in exon 28 (c.3538G>A p.Gly1180Arg) (Table 1). Deletions in the 5' splice site of exon 25 were previously reported in eight other FSHD2 families suggesting it to be a mutation hotspot[21; 24]. Previous RNA analysis of an independent FSHD2 family with a c.3276_3276+1del variant showed that this variant results in both cryptic splicing and in complete skipping of exon 25, both with retention of the ORF[21]. The same splice effect is expected for the c.3276_3276+4del variant in individual 1414-201. The variant in exon 28 has not been reported previously. The missense predictions of SIFT and MutationTaster defined the variant in exon 28 as deleterious and disease causing, respectively, whereas Align GVGD gives a score of C15 which indicates that is not very likely that the variant interferes with protein function (Table 2).

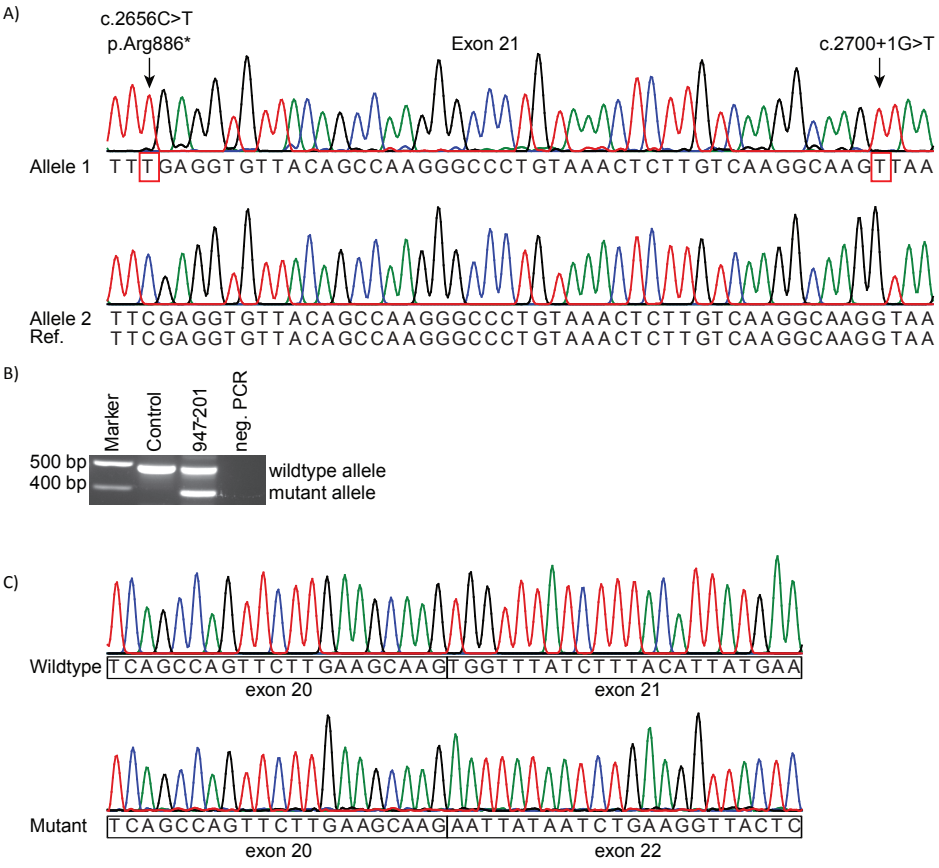


FIGURE 2. Two *SMCHD1* variants in cis in Rf947. (a) Sanger sequence tracks from two clones showing the different *SMCHD1* alleles of individual 947-201. Allele 1 contains the two variants in exon 21 at positions c.2656C>T and c.2700+1G>T, indicated with the arrows. Allele 2 contains the reference nucleotides at these positions. (b) RT-PCR of *SMCHD1* with primers in exon 18 and exon 23 for a control and individual 947-201. The expected PCR product size for wild type transcript is 491 bp, additionally there is a smaller PCR product for 947-201, indicating an alternatively spliced mutant transcript. (c) Sanger sequence tracks of the wild type and mutant RT-PCR products, showing skipping of exon 21 in the mutant transcript.

To investigate if the two variants were located on the same allele, a long range PCR of *SMCHD1* exons 25-28 was performed. This PCR product was circularized and an additional PCR was performed to amplify the fused exons 25 and 28. The PCR product was cloned in a TOPO vector and individual clones were sequenced demonstrating that the two variants are located on different alleles (Fig. 3).

TABLE 2. Variant predictions for identified *SMCHD1* missense variants

	c.3538G>A, p.Gly1180Arg	**c.2941T>G, p.Tyr981Asp**	**c.5596C>G, p.Arg1866Gly**
Conservation (PhyloP)	Moderate (4.16)	Moderate (2.38)	Weak (1,09)
Grantham distance	Moderate (125)	Large (160)	Moderate (125)
Align GVGD	C15 (GV: 206.04 - GD: 124.98)	C65 (GV: 0.00 - GD: 159.94)	C65 (GV: 0.00 - GD: 125.13)
SIFT	Deleterious (Score:0)	Deleterious (Score:0)	Deleterious (Score:0)
Mutation taster	Disease causing (p-value: 0.999)	Disease causing (p-value: 0.689)	Disease causing (p-value: 0.882)

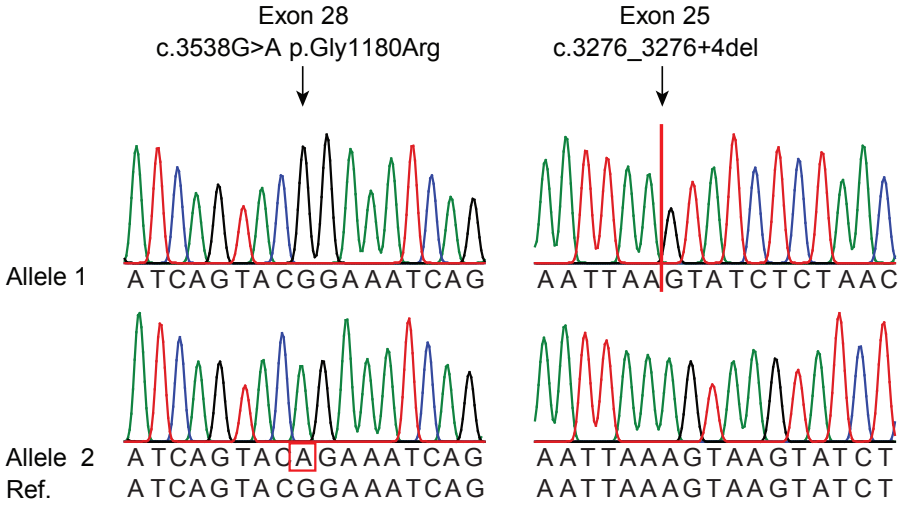


FIGURE 3. Two *SMCHD1* variants in trans in Rf1414. Sanger sequence tracks from two clones showing the different *SMCHD1* alleles of individual 1414-201. Allele 1 contains the variant in exon 25 (c.3276_3276+4del) and allele 2 contains the variant in exon 28 (c.3538G>A), both indicated with an arrow.

To establish if both variants affect function, two additional family members were clinically evaluated and genetic analysis was performed. Sanger sequencing showed that the mother (1414-102) of individual 1414-201 carries the variant in exon 28. She shows no FSHD phenotype (CSS of 0 at age of 78) and has no FSHD permissive chromosome. D4Z4 methylation analysis indicated that her D4Z4 methylation level (62%, Delta1 score 18%) is within the normal range (Fig. 1). The sister of the index case (1414-202) does

not carry either of the *SMCHD1* variants, no D4Z4 hypomethylation (59%, Delta1 score 8%) and had no clinical signs of FSHD (CSS of 0 at age of 53). Information about the father (1414-101) of individual 1414-201 was not available. This suggests that although individual 1414-201 carries two *SMCHD1* variants, only the variant in exon 25 causes D4Z4 hypomethylation and is causal to FSHD.

Two *SMCHD1* variants *in trans* in Rf385 with additive effect

Individual 203 of Rf385 was suspected of FSHD based on physical examination with a CSS of 8 at age 42 (Fig. 1). Genetic analysis showed that he carries a permissive allele of 27 D4Z4 units and the D4Z4 methylation analysis revealed a methylation level of 5% (Delta1 score -39%), which is consistent with the diagnosis of FSHD2.

One of the sisters (385-206) of the index case suffered from a meningitis in childhood and she is mentally disabled. Furthermore she broke both her arms and a leg in the past and the neurologist was unable to conclusively establish the clinical diagnosis of FSHD. However, she does carry one permissive allele and D4Z4 methylation analysis revealed a methylation level of 1% (Delta1 score -37%), which is suggestive of FSHD2.

The other sister (385-202) does not show symptoms of FSHD with a CSS of 0 at age 38. She does carry two non-permissive alleles and D4Z4 methylation analysis revealed a methylation level of 24% (Delta1 score -9%), which is higher than the Delta1 range between -20% and -45% observed in carriers of an *SMCHD1* variant that affects function[21]. The father (385-101) and mother (385-102) of 385-203 do not show symptoms of FSHD, having a CSS of 0 at age 70 and CSS of 0 at age 67, respectively. The mother (385-102) carries one permissive allele and D4Z4 methylation analysis revealed a methylation level of 10% (Delta1 score -31%), suggestive of FSHD2. The father (385-101) carries two non-permissive alleles and D4Z4 methylation analysis revealed a methylation level of 20% (Delta1 score -25%).

SMCHD1 Sanger sequencing identified two *SMCHD1* variants in 385-203 and his sister (385-206) (Fig. 4). The first variant is a missense variant in exon 24 (c.2941T>G, p.Tyr981Asp), the second variant is a missense variant in exon 45 (c.5596C>G p.Arg1866Gly) (Table 1). Both variants have not been reported previously. The father (385-101) carries the *SMCHD1* variant in exon 24, and the mother (385-102) carries the *SMCHD1* variant in exon 45 (Fig. 4). This shows that both variants are located on different alleles. The other sister (385-202) does not carry either of the *SMCHD1* variants.

The missense predictions of SIFT and MutationTaster predict both variants as deleterious and disease causing, respectively. The Align GVGD scores of C65 also indicate that the variants are both likely to interfere with protein function (Table 2). Additionally, both variants are individually associated with D4Z4 methylation levels and Delta1 scores in the FSHD2 range. However, D4Z4 methylation levels and Delta1 scores

are more strongly reduced in the two family members with both variants than in the single variant carriers. This indicates that the effect of the variants is additive and that the combination of both variants is more deleterious than the corresponding single variants.

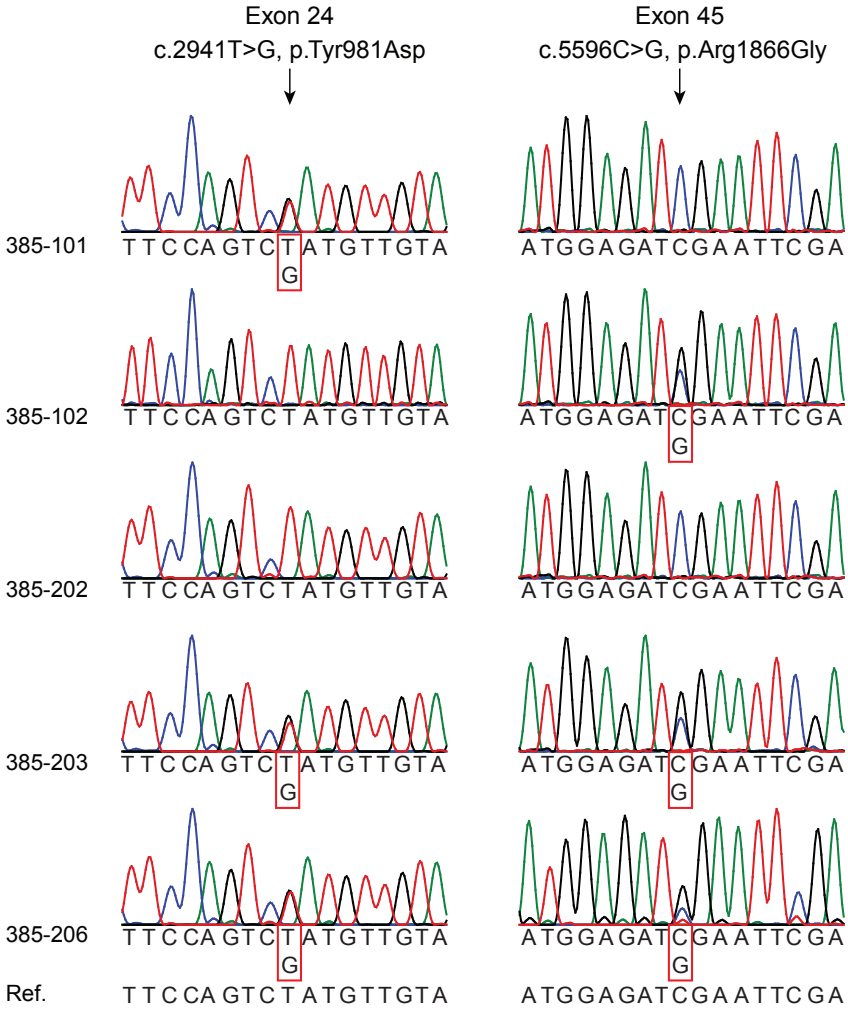


FIGURE 4. Two *SMCHD1* variants in trans in Rf385. Sanger sequence tracks from Rf385 family members showing that the two variants are located *in trans*. Individuals 385-203 and 385-206 carry both *SMCHD1* variants (c.2941T>G and c.5596C>G), both indicated with an arrow. Individual 385-101 carries only the *SMCHD1* variant in exon 24. Individual 385-102 carries only the *SMCHD1* variant in exon 45. Individual 385-202 carries the reference nucleotide for both variants.

DISCUSSION

In this study we analyzed three FSHD2 families in which two *SMCHD1* variants that potentially affect function were found in each proband. Of the six *SMCHD1* variants described here, one was reported previously, one we demonstrated to be neutral based on segregation analysis, and the other four variants are novel *SMCHD1* variants that affect function.

In individual 947-201 two variants in exon 21, separated by only 45 nucleotides, were identified on the same allele. Multiple variants in close proximity are seen more often and might result from chronocoordinate events due to transient error-prone conditions[26-29]. In this case, both variants could individually be expected to affect function. The first variant (c.2656C>T) is a nonsense substitution, which will cause reading frame interruption by a premature stop codon. The second variant (c.2700+1G>T) is located in the splice donor site of exon 21 and this variant is predicted to cause skipping of exon 21. mRNA analysis revealed that the splice donor site variant c.2700+1G>T indeed results in a skip of exon 21. The skipping of exon 21 will disturb the ORF, resulting in a premature stop codon in exon 22. The intensity of the mutant transcript on gel indicates that this may not lead to nonsense mediated mRNA decay but we were unable to study the variant at the protein level. The mean Delta2 score for ORF disrupting variants is 2.7%, while it is -1.8% for ORF preserving variants[21]. The Delta2 score of 947-201 is 0%, which in this case is inconclusive. With the skipping of exon 21, the nonsense substitution c.2656C>T has no effect on the transcript. However, if this variant would be present independently, it would also disrupt the ORF, probably resulting in hypomethylation and an FSHD phenotype.

In individual 1414-201 two variants on different alleles were identified. Although the missense substitution in exon 28 was predicted to affect function by SIFT and MutationTaster (but not by Align GVGD), segregation analysis showed that this variant does not independently cause D4Z4 hypomethylation, and is therefore likely a neutral variant, even on a FSHD-permissive background. The second variant found in individual 1414-201 is a splice donor site variant in exon 25. Splice donor site variants in exon 25 have already been reported in eight other FSHD2 families to segregate with D4Z4 hypomethylation and disease presentation, confirming the functional consequences of this variant[21; 24].

In individual 385-203 two variants on different alleles were detected. Both variants are predicted to affect function by SIFT, MutationTaster and Align GVGD. In accordance, both variants cause D4Z4 hypomethylation independently, as observed in the parents of 385-203. In this family, the combination of both variants further decreases the Delta1 score at D4Z4 from -25% in the father (385-101) and -31% in the mother (385-102), which both carry one *SMCHD1* variant, to -39% in the patient (385-203) and -37% in his

sister (385-206), both double *SMCHD1* variant carriers (Fig. 1). This suggests a synergistic effect of both variants. The other sister (202) has a Delta1 score of -9%, indicating that she has moderate D4Z4 hypomethylation, independent of an *SMCHD1* variant. Therefore, the data are consistent with an additive effect of having one or two *SMCHD1* missense variants on D4Z4 methylation. This additive effect is also shown by the Delta2 scores, which are reduced in 385-203 (-7%) and 385-206 (-9%), which carry both missense variants, compared to the father (7%) and mother (-1%).

The synergistic effect of both variants also explains the very low D4Z4 methylation levels and severe FSHD phenotype of individual 385-203. It is unfortunate that it is not possible to establish whether the sister 385-206 is affected with FSHD. The father (385-101) carries two non-permissive D4Z4 alleles, which explains why he is not affected with FSHD. In regards to the mother (385-102), the size of her permissive allele (27 units) might explain why she is not affected with FSHD, despite her *SMCHD1* variant and D4Z4 hypomethylation. In 2010, de Greef *et al.* showed that the average size of the shortest permissive allele in FSHD2 is 16 units, which is much shorter than the average of 28 units found in control individuals, which was confirmed in a later study[4; 21]. The permissive allele in the proband (385-203) and his mother (385-102) is 27 units, relatively long for FSHD2. This might partially explain why the mother does not show an FSHD phenotype and why two *SMCHD1* variants are necessary in this family to present FSHD symptoms. However, it is also known that there are FSHD2 patients carrying a single *SMCHD1* variant and a single permissive allele of as much as 40 units[21].

Furthermore, variability in clinical representation is a hallmark of FSHD, which is seen both within and between families, and more often in females than in males[30-32]. Some, but not all, of this variability can be explained by the size of the permissive D4Z4 repeat and presence and type of an *SMCHD1* variant[21]. Both variants in this family are ORF preserving variants, which have in general a more profound effect on D4Z4 methylation level than ORF disrupting variants[21].

In the mouse, Smchd1 has a role in the establishment and maintenance of CpG methylation of a subset of genes on the inactive X-chromosome and in the expression of several autosomal gene clusters which are monoallelicly expressed[33-37]. Female homozygous MommeD1 mice, which completely lack Smchd1 protein, die at midgestation because of a failure in X inactivation[38]. Family Rf385 shows that having two *SMCHD1* copies with a missense variant is viable, both in males and in females. This suggests that either SMCHD1 is not essential in human or, more likely, that the variants in family Rf385 only partially abrogate SMCHD1 protein function. This may also be true for the other variants found in FSHD2 patients, because so far no comorbidities for FSHD2, with regard to SMCHD1 function, have been identified.

This report describes four new *SMCHD1* variants which affect function and a neutral variant, in addition to the approximately 70 variants that already have been identified since the discovery of *SMCHD1* as the most common FSHD2 gene in 2012. This is the first report of families with individuals carrying two *SMCHD1* variants and this study highlights the usefulness of D4Z4 methylation analysis to determine the functional consequences of *SMCHD1* variants. These families highlight the variability in *SMCHD1* variants underlying FSHD2. In one of the three families both variants contribute separately to the disease, as is reflected by the Delta1 and Delta2 scores in this family and the penetrance of the disease. Importantly, this family shows that a combination of two ORF-preserving variants in *SMCHD1* is compatible with life.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank the families for participating in our studies. This study was supported with 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 Muscular Dystrophy Association (MDA; 217596), the Fields Center for FSHD Research, the Geraldi Norton and Eklund family foundation, the FSH Society, The Prinses Beatrix Spierfonds (W.OR12-20), Spieren voor Spieren, The Friends of FSH Research, and European Union Framework Programme 7 agreement 2012-305121 (NEUROMICS).

REFERENCES

1. Deenen, J.C., Arnts, H., van der Maarel, S.M., Padberg, G.W., Verschuuren, J.J., Bakker, E., Weinreich, S.S., Verbeek, A.L., and van Engelen, B.G. (2014). Population-based incidence and prevalence of facioscapulohumeral dystrophy. Neurology 83, 1056-1059.
2. Padberg, G.W., Lunt, P.W., Koch, M., and Fardeau, M. (1991). Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscular disorders : NMD 1, 231-234.
3. Statland, J.M., and Tawil, R. (2011). Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions. Current opinion in neurology 24, 423-428.
4. de Greef, J.C., Lemmers, R.J., Camano, P., Day, J.W., Sacconi, S., Dunand, M., van Engelen, B.G., Kiuru-Enari, S., Padberg, G.W., Rosa, A.L., et al. (2010). Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology 75, 1548-1554.
5. van Overveld, P.G., Lemmers, R.J., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nature genetics 35, 315-317.
6. Zeng, W., de Greef, J.C., Chen, Y.Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A., et al. (2009). Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS genetics 5, e1000559.
7. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., et al. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics 6, e1001181.
8. Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329, 1650-1653.
9. Balog, J., Thijssen, P.E., de Greef, J.C., Shah, B., van Engelen, B.G., Yokomori, K., Tapscott, S.J., Tawil, R., and van der Maarel, S.M. (2012). Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics : official journal of the DNA Methylation Society 7, 579-584.
10. Bosnakovski, D., Xu, Z., Gang, E.J., Galindo, C.L., Liu, M., Simsek, T., Garner, H.R., Agha-Mohammadi, S., Tassin, A., Coppee, F., et al. (2008). An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. The EMBO journal 27, 2766-2779.
11. Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O., et al. (2007). The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscular disorders : NMD 17, 611-623.

12. Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., et al. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Developmental cell 22, 38-51.
13. Lemmers, R.J., de Kievit, P., Sandkuijl, L., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2002). Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nature genetics 32, 235-236.
14. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics 44, 1370-1374.
15. van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., and Frants, R.R. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Human molecular genetics 2, 2037-2042.
16. Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nature genetics 2, 26-30.
17. Hirano, T. (2005). SMC proteins and chromosome mechanics: from bacteria to humans. Philosophical transactions of the Royal Society of London Series B, Biological sciences 360, 507-514.
18. Hirano, T. (2006). At the heart of the chromosome: SMC proteins in action. Nature reviews Molecular cell biology 7, 311-322.
19. Ball, A.R., Jr., Chen, Y.Y., and Yokomori, K. (2014). Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies. Biochimica et biophysica acta 1839, 191-202.
20. Sacconi, S., Lemmers, R.J., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. American journal of human genetics 93, 744-751.
21. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2014). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics.
22. Mitsuhashi, S., Boyden, S.E., Estrella, E.A., Jones, T.I., Rahimov, F., Yu, T.W., Darras, B.T., Amato, A.A., Folkerth, R.D., Jones, P.L., et al. (2013). Exome sequencing identifies a novel SMCHD1 mutation in facioscapulohumeral muscular dystrophy 2. Neuromuscular disorders : NMD 23, 975-980.

23. Winston, J., Duerden, L., Mort, M., Frayling, I.M., Rogers, M.T., and Upadhyaya, M. (2014). Identification of two novel SMCHD1 sequence variants in families with FSHD-like muscular dystrophy. European journal of human genetics : EJHG.

24. Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., and Muller, C.R. (2014). Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. European journal of human genetics : EJHG.

25. Ricci, E., Galluzzi, G., Deidda, G., Cacurri, S., Colantoni, L., Merico, B., Piazzo, N., Servidei, S., Vigneti, E., Pasceri, V. (1999). Progress in the molecular diagnosis of facioscapulohumeral dystrophy and correlation between the number of KpnI repeat at the 4q35 locus and clinical phenotype. Ann Neurol 45, 751-757.

26. Hill, K.A., Wang, J., Farwell, K.D., Scaringe, W.A., and Sommer, S.S. (2004). Spontaneous multiple mutations show both proximal spacing consistent with chronocoordinate events and alterations with p53-deficiency. Mutation research 554, 223-240.

27. Colgin, L.M., Hackmann, A.F., Emond, M.J., and Monnat, R.J., Jr. (2002). The unexpected landscape of in vivo somatic mutation in a human epithelial cell lineage. Proceedings of the National Academy of Sciences of the United States of America 99, 1437-1442.

28. Chen, Z., Feng, J., Buzin, C.H., and Sommer, S.S. (2008). Epidemiology of doublet/multiplet mutations in lung cancers: evidence that a subset arises by chronocoordinate events. PloS one 3, e3714.

29. Buettner, V.L., Hill, K.A., Scaringe, W.A., and Sommer, S.S. (2000). Evidence that proximal multiple mutations in Big Blue transgenic mice are dependent events. Mutation research 452, 219-229.

30. Zatz, M., Marie, S.K., Cerqueira, A., Vainzof, M., Pavanello, R.C., and Passos-Bueno, M.R. (1998). The facioscapulohumeral muscular dystrophy (FSHD1) gene affects males more severely and more frequently than females. American journal of medical genetics 77, 155-161.

31. van der Maarel, S.M., Deidda, G., Lemmers, R.J., van Overveld, P.G., van der Wielen, M., Hewitt, J.E., Sandkuijl, L., Bakker, B., van Ommen, G.J., Padberg, G.W., et al. (2000). De novo facioscapulohumeral muscular dystrophy: frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. American journal of human genetics 66, 26-35.

32. Tonini, M.M., Passos-Bueno, M.R., Cerqueira, A., Matioli, S.R., Pavanello, R., and Zatz, M. (2004). Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). Neuromuscular disorders : NMD 14, 33-38.

33. Blewitt, M.E., Gendrel, A.V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., et al. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nature genetics 40, 663-669.

34. Gendrel, A.V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Developmental cell 23, 265-279.

35. Gendrel, A.V., Tang, Y.A., Suzuki, M., Godwin, J., Nesterova, T.B., Greally, J.M., Heard, E., and Brockdorff, N. (2013). Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Molecular and cellular biology 33, 3150-3165.

36. Nozawa, R.S., Nagao, K., Igami, K.T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nature structural & molecular biology 20, 566-573.

37. Mould, A.W., Pang, Z., Pakusch, M., Tonks, I.D., Stark, M., Carrie, D., Mukhopadhyay, P., Seidel, A., Ellis, J.J., Deakin, J., et al. (2013). Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. Epigenetics & chromatin 6, 19.

38. Blewitt, M.E., Vickaryous, N.K., Hemley, S.J., Ashe, A., Bruxner, T.J., Preis, J.I., Arkell, R., and Whitelaw, E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proceedings of the National Academy of Sciences of the United States of America 102, 7629-7634.

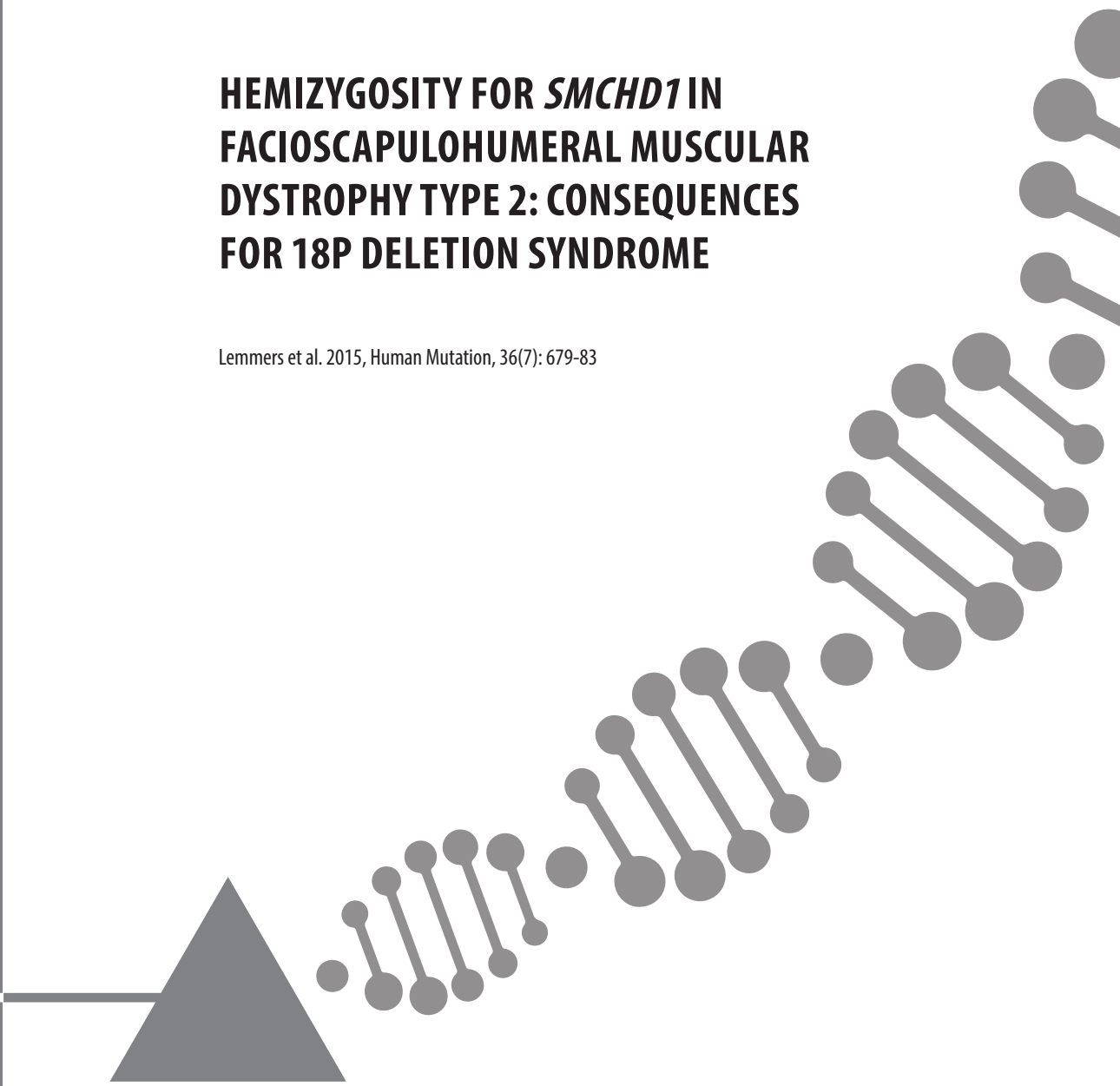
Richard JFL Lemmers[1], Marlinde L van den Boogaard[1], Patrick J van der Vliet[1], Colleen M Donlin-Smith[2], Sharon P Nations[3], Claudia AL Ruivenkamp[4], Patricia Heard[5], Bert Bakker[4], Stephen J Tapscott[6], Jannine D Cody[5], Rabi Tawil[2], Silvère van der Maarel[1]

1 Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
2 Neuromuscular Disease Unit, Department of Neurology, University of Rochester Medical Center, Rochester, USA
3 Department of Neurology, University of Texas Southwestern Medical Center, Dallas, USA
4 Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
5 University of Texas Health Science Center at San Antonio, Chromosome 18 Research Center, San Antonio, USA
6 Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, USA

3

HEMIZYGOSITY FOR *SMCHD1* IN FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY TYPE 2: CONSEQUENCES FOR 18P DELETION SYNDROME

Lemmers et al. 2015, Human Mutation, 36(7): 679-83



ABSTRACT

Facioscapulohumeral muscular dystrophy (FSHD) is most often associated with variegated expression in somatic cells of the normally repressed *DUX4* gene within the D4Z4 repeat array. The most common form, FSHD1, is caused by a D4Z4 repeat array contraction to a size of 1-10 units (normal range 10-100 units). The less common form, FSHD2, is characterized by D4Z4 CpG hypomethylation and is most often caused by loss of function mutations in the structural maintenance of chromosomes hinge domain 1 (*SMCHD1*) gene on chromosome 18p. The chromatin modifier SMCHD1 is necessary to maintain a repressed D4Z4 chromatin state. Here we describe two FSHD2 families with a 1.2 Mb deletion encompassing the *SMCHD1* gene. Numerical aberrations of chromosome 18 are relatively common and the majority of 18p deletion syndrome (18p-) cases have, like these FSHD2 families, only one copy of *SMCHD1*. Our finding therefore raises the possibility that 18p- cases are at risk of developing FSHD. To address this possibility, we combined genome wide array analysis data with D4Z4 CpG methylation and repeat array sizes in individuals with 18p- and conclude that approximately 1:8 18p- cases might be at risk of developing FSHD.

Individuals with facioscapulohumeral muscular dystrophy (FSHD; MIM 158900 [FSHD1] and MIM 158901 [FSHD2]) have progressive and often asymmetric muscle weakness affecting the face, shoulder, and upper arms, followed by involvement of the distal and proximal lower extremities[1; 2]. These symptoms are associated with a higher probability of skeletal muscle expression of the *DUX4* retrogene (OMIM 606009) which is normally repressed in somatic tissue[3]. There are at least two genetically distinct causes of FSHD that are clinically indistinguishable. The most common form, FSHD1, is caused by a variant of the polymorphic D4Z4 macrosatellite repeat array on chromosome 4 that is between 1 and 10 units[4]. In unaffected individuals this polymorphic repeat array ranges between 11-100 D4Z4 units and is heavily CpG methylated in somatic tissue thereby repressing expression of the region[5]. The shorter array in individuals with FSHD1 is insufficient to maintain the repressed chromatin state necessary for sustaining repression of the *DUX4* retrogene expression. The less common form, FSHD2, is also characterized by a partial loss of CpG methylation at D4Z4 but with repeat array sizes > 10 units and is usually caused by mutations in the structural maintenance of chromosomes hinge domain 1 (*SMCHD1*) gene (OMIM 614982) on chromosome 18p[5-7]. SMCHD1 is a chromatin modifier that binds to the D4Z4 repeat array and is necessary to maintain a repressed D4Z4 chromatin state in somatic cells. Each of the 3,3kb large D4Z4 units contains a copy of the double homeobox 4 (*DUX4*) retrogene[8]. In FSHD1, with the short D4Z4 repeat array, or in FSHD2 with heterozygous *SMCHD1* mutations, either cause leads to a partial derepression of the D4Z4 repeat array in somatic cells. This derepression is marked by D4Z4 hypomethylation, a partial loss of repressive chromatin marks and other chromatin changes and the sporadic expression of DUX4 in a small set of muscle nuclei[5; 9-11].

However, either mechanism; D4Z4 short repeat sizes, or *SMCHD1* mutations, leading to depression of D4Z4 by themselves are not sufficient to cause FSHD. The D4Z4-embedded *DUX4* retrogene requires a polyadenylation signal in the 3' untranscribed region to produce a stable transcript and hence DUX4 protein in somatic cells[12]. This *DUX4* polyadenylation signal can only be found immediately distal to the last *DUX4* copy, but its existence is polymorphic in the population and is only found on 4qA-type chromosomes. 4qB chromosomes, that are equally common, or 10q chromosomes that contain a highly homologous copy of the D4Z4 repeat array, do not have a *DUX4* polyadenylation signal[12-14]. Thus, digenic inheritance of a short D4Z4 repeat array *in cis* with a *DUX4* polyadenlylation signal in FSHD1, or a heterozygous *SMCHD1* mutation and a *DUX4* polyadenylation signal in FSHD2, are both mechanisms for the development of FSHD[6; 12].

We previously reported that the average repeat array size of the 4qA (i.e. DUX4 producing) chromosome in FSHD2 patients is most often between 11-16 units, which is at the lower end of the repeat size spectrum of 4qA chromosomes in control individuals

(Fig. S1)[15; 16]. We define FSHD2 patients by a methylation level of <25% measured by Southern blotting at the methylation-sensitive *Fse*I restriction site in D4Z4, a threshold that has been instrumental in identifying *SMCHD1* as FSHD2 disease gene[6; 16; 17]. In our recent *SMCHD1* mutation screen in affected individuals with *Fse*I methylation levels <25%, we showed that mutations that do not disrupt the open reading frame (ORF) are more deleterious than those that introduce a premature stop codon[16]. This conclusion was based on the difference in D4Z4 hypomethylation for the two types of mutations. We found a significant correlation between the cumulative number of all D4Z4 repeat units in the arrays on chromosomes 4 and 10 and the average methylation of the D4Z4 repeats in controls and FSHD patients. This correlation allowed us to define the Delta1 value, which is the observed CpG methylation level at this site in D4Z4 minus the predicted methylation based on control individuals. The average Delta1 in controls is about zero and ranges between -13% (5% percentile) and 17% (95% percentile), while in carriers of an *SMCHD1* mutation this value is about -30% and ranges between -42% (5% percentile) and -22% (95% percentile)[16]. Therefore, the Delta1 value represents the methylation defect corrected for the size of the D4Z4 repeat array, emphasizing the strong effect of *SMCHD1* mutations on D4Z4 methylation. We also introduced the Delta2 value, which is the observed methylation minus the predicted methylation based on *SMCHD1* mutation carriers and reveals the significance of the mutation[16]. *SMCHD1* mutations that preserve the ORF most probably result in a dominant negative type mechanism and lead to the lowest Delta1 value, and a negative Delta2 value in somatic cells. ORF disrupting *SMCHD1* mutations, seem to be associated with nonsense mediated RNA decay suggesting a haploinsufficiency type mechanism and are associated with lesser degrees of D4Z4 CpG hypomethylation resulting in positive Delta2 values. The observation that both the nature of the *SMCHD1* mutation and the size of the D4Z4 repeat array determines the D4Z4 methylation and the disease severity in FSHD2, led to a model in which haploinsufficiency-causing mutations require medium short D4Z4 arrays (most often in the range of 11-16 units) to cause FSHD2, while dominant negative SMCHD1 mutations can also cause FSHD2 with larger 4qA array sizes (Fig. S1)[16]. Thus, while establishing FseI methylation and Delta values are not standard operating procedures in diagnostic settings, these values have contributed to our understanding of FSHD pathogenesis.

In approximately 15% of the FSHD2 families, identified by phenotype and D4Z4 hypomethylation, we could not identify a mutation in *SMCHD1*[6; 16]. In six out of ten of these families, Sanger sequencing of *SMCHD1* in the index case did not reveal a causal variant but showed absence of heterozygous SNPs in all of the 48 exons and flanking sequences. These families were subjected to whole exome sequencing and genome wide array analysis.

The first family (Rf898) reported in this study came to our attention because of classical features of FSHD, while FSHD1 was excluded by D4Z4 repeat array sizing (Fig. 1A and Table S1). The index case (898.205) was diagnosed at age 49 when he presented with difficulty lifting his arms overhead. In retrospect, he recalls having shoulder difficulties when he was 12 years of age. His physical examination showed a combination of facial weakness, scapular winging and mild foot dorsiflexor weakness as well as mild abdominal muscle weakness. Two additional family members (898.202, 898.301) were moderately affected with similar clinical features. Subject 898.202 was asymptomatic but had definite signs of FSHD with scapular weakness, foot dorsiflexion weakness and mild abdominal muscle weakness when examined at age 65. Her son (898.301) reports symptom onset at age 13 with clear signs of FSHD when examined at age 36 (Supp. Table S1). All carry an 11 units D4Z4 repeat array on a 4qA chromosome containing a *DUX4* polyadenylation site. Methylation analysis provided evidence for reduced D4Z4 methylation levels and Delta1 values consistent with FSHD2 in all three affected individuals, but also in two unaffected individuals. Subject 898.303 carries the same 11 unit 4qA allele as his affected father, reports shoulder pain and weakness since the age of 24 but shows no clear signs of weakness on examination at age 36. An EMG study, however, showed myopathic units in his supraspinatus and cervical paraspinal muscles. Individual 898.204 is unaffected at the age of 59, but carries a 35 unit repeat array on a permissive 4qA allele, which appears to be too long to be pathogenic based on previous findings in carriers of a haploinsufficiency-causing *SMCHD1* mutation[16]. *SMCHD1* mutation analysis by Sanger sequencing failed to identify potentially damaging variants. Whole exome sequencing on 898.301 and 898.303 failed to identify the causative gene, but extended the region of homozygosity into several *SMCHD1* neighboring genes, suggestive of a heterozygous deletion of the *SMCHD1* locus. Subsequently, we performed a genome wide array analysis in these subjects and identified a 1.2Mb deletion that spans the *METTL4*, *NDC80*, *SMCHD1*, *EMILIN2*, *LPIN2* genes (Fig. 1B). This deletion was confirmed in subjects 898.202, 898.204 and 898.205 and this variant was submitted to the Leiden Open Variation Database (http://databases.lovd.nl/shared/variants/SMCHD1).

In the second family (Rf929), we identified two FSHD individuals carrying a 13 D4Z4 unit repeat array on 4qA and CpG hypomethylation at D4Z4 suggesting FSHD2 (Fig. 1C; Table S1). The index case's (929.201) initial complaints were difficulty with his proximal shoulders in his late twenties but was not fully evaluated until age 60 when he presented with facial weakness, shoulder and upper extremity weakness and foot dorsiflexor weakness. One sister (929.204) is also similarly affected whereas his other sister (929.206) is asymptomatic and clinically unaffected. By array analysis in individual

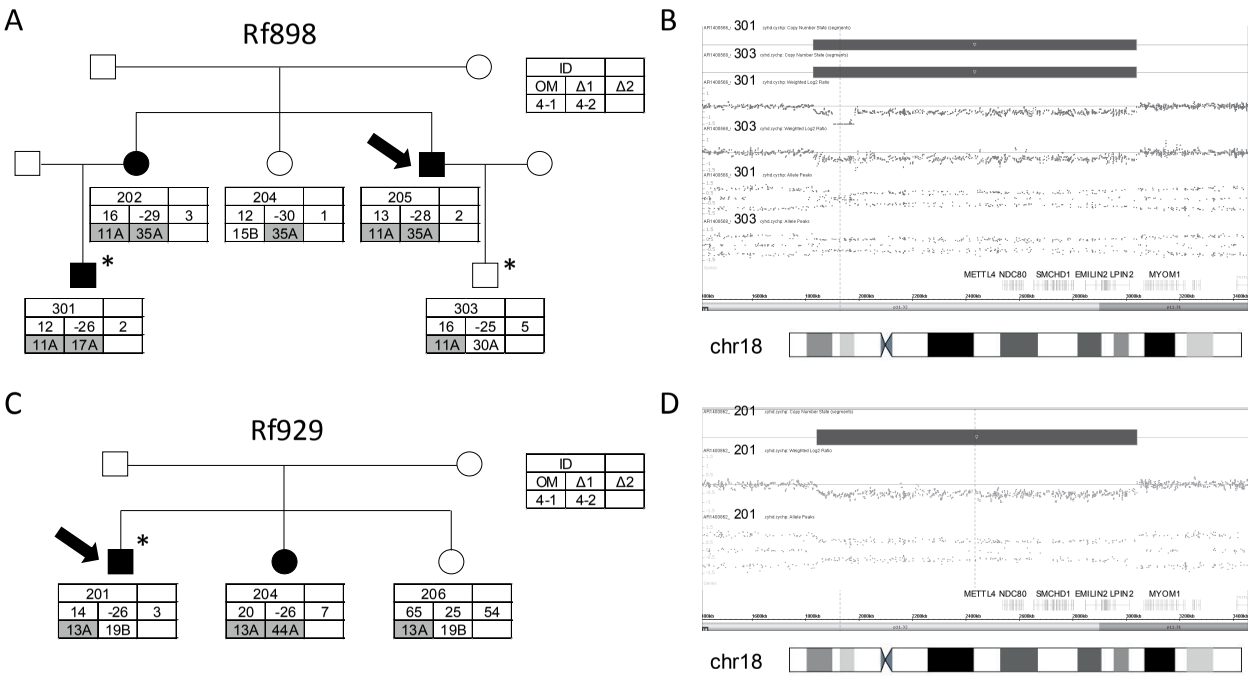


FIGURE 1. (a) Pedigree of family Rf898, showing 3 affected individuals (898.202, 898.205 and 898.301) all carrying an 11 unit D4Z4 repeat array on chromosome 4qA (11A) and a partial loss of D4Z4 CpG methylation. The index case (arrow) and the individuals that have been analyzed on the SNP array (asterisk) are indicated. The key indicates the individual ID, the different methylation values (Delta1; Δ1, Delta2; Δ2 and the measured methylation; OM) and the D4Z4 repeat array type and sizes in units on both chromosomes 4 (4-1 and 4-2) **(b)** Array plots of chromosome 18. A 1.2 Mb deletion in 18p11.32p11.31 was detected in patients 898.301 and 898.303 (arr[hg19] 18p11.32p11.31(1,829,674-3,039,186)x1). The deletion is marked by four panels showing from top to bottom: the Copy Number State, the Weighted Log2 Ratio, the Allele Peaks and the genes. **(c)** Pedigree of family Rf929, showing 2 affected individuals (929.201 and 929.204) both carrying a 13 unit D4Z4 repeat array on chromosome 4qA and a loss of D4Z4 CpG methylation. **(d)** Array analysis of patient 929.201 presenting a 1.2 Mb deletion in 18p11.32p11.31 (arr[hg19] 18p11.32p11.31(1,844,188-3,039,186)x1)).

929.201, we identified the same interstitial chromosome 18p deletion including *SMCHD1* as was found in family Rf898 (Fig. 1D). This 1.2 Mb deletion was confirmed in affected individual 929.204.

Both families reside in the same state in the US and although these two families are not known to be related, the observation that they carry the same interstitial 18p deletion suggest otherwise. Detailed SNP analysis on chromosome 18 showed that the SNP difference between nephews 898.301 and 898.303 is 1.8% and the difference between 898.301 (or 898.303) and 929.201 is 4.9%, while the average difference at chromosome 18 between unrelated individuals is usually >15% (CAL Ruivenkamp, personal communication), strongly suggesting that both families are related. Array analysis in the remaining four FSHD2 families that showed absence of heterozygous SNPs in all 48 *SMCHD1* exons and flanking sequences did not reveal evidence for a deletion in the *SMCHD1* locus.

Although one third of all identified *SMCHD1* mutations in FSHD2 most probably lead to haploinsufficiency, *SMCHD1* was not known as a dosage sensitive gene. Our current study emphasizes that monosomy of *SMCHD1* can cause familial FSHD2, although we cannot rule out that genetic variants in other (currently unknown) epigenetic modifiers of D4Z4 are shared between both families. We identified two families with five affected family members who all carry the same 1.2Mb 18p microdeletion in combination with a D4Z4 repeat array with 11 or 13 units on a 4qA chromosome. This finding also corroborates our previous study in which we show that haploinsufficiency-causing *SMCHD1* mutations in general result in positive Delta2 values, suggestive of a less deleterious outcome[16]. The Delta2 value in the 7 carriers of the *SMCHD1* deletion varied between 1 and 7%.

None of the other genes within the microdeletion have been reported to be sensitive to hemizygosity. Homozygous mutations in lipin 2 *(LPIN2)* have been associated with Majeed syndrome, characterized by chronic recurrent multifocal osteomyelitis and congenital dyserythropoietic anaemia[18]. *METTL4* (methyltransferase-like 4) encodes for a putative methyltransferase of unknown function while *NDC80* encodes a central component of the kinetochore protein complex necessary for proper chromosome segregation and spindle checkpoint signaling[19].

Structural and numerical aberrations of the short arm of chromosome 18 are relatively common amongst the chromosomal rearrangement syndromes. The incidence of monosomy 18p is estimated at about 1:50;000 newborns worldwide, and although the dysmorphic features in 18p deletion syndrome (18p-) individuals is not very uniform, muscular dystrophy has not been reported[20]. Given the association of *SMCHD1* mutations with FSHD and the identification of two families carrying a deletion of *SMCHD1*, it raises the question whether individuals with 18p- are at increased risk of developing FSHD

when also carrying a chromosome 4q that contains a *DUX4* polyadenylation signal. To address this possibility, analysis of CGH array data from 83 individuals with 18p-, showed that *SMCHD1* was hemizygous in all but one case[21]. Subsequent CpG methylation analysis of the cases with *SMCHD1* deletion confirmed a reduced methylation at D4Z4 in 72/82 cases (Fig. 2A and B), suggesting that individuals with 18p- have a similar degree of D4Z4 chromatin relaxation as individuals with FSHD2. Previously, we showed that 4qA chromosomes (containing a *DUX4* polyadenylation signal) and 4qB chromosomes (without *DUX4* polyadenylation signal) are almost equally common in the Caucasian and Asian populations, while 4qA is the most common chromosome 4 variant in the African population[22]. This suggests that about 25% of 18p- individuals carry two 4qB chromosomes and are therefore not at risk for developing FSHD. In two recent studies, we analyzed the D4Z4 repeat array size and haplotype distribution in control individuals[16; 23]. These results showed that 20 out of 159 independent 4A161 chromosomes, the most common risk allele for FSHD in the Caucasian population accounting for 79% of all 4qA alleles[22], had repeat arrays of ≤16 units, which gives an estimated incidence of 12.6% (Fig. 2C). Because half of the chromosomes 4 are of the 4qB variant, about 12.6% (1 in 8) of individuals in the Caucasian population with 18p- carry a medium-short 4qA array and might therefore be at risk for FSHD. To investigate this possibility, we performed accurate D4Z4 repeat array sizing in nineteen individuals with 18p- that were selected on the basis that they carried a 4A161 allele and had D4Z4 hypomethylation and we found that two individuals carried an array ≤16 units and may have an increased risk of developing FSHD (Fig S2 and Table S2). Individual 18p-14 carries a 13 unit repeat array and is 27 years old, while 18p-37 carries a 15 unit repeat and is 24 years. At this time, we do not have information about these individuals clinical status. Contractions on other 4qA alleles, such as 4A159, 4A168 and 4A166H, may also confer a risk to FSHD[12]. However, these alleles are much less common than the 4A161 allele[22], and therefore only increase the FSHD risk in the 18p- patient population by approximately 1.5%.

CONCLUSION

We report heterozygous deletions of *SMCHD1* in two distantly related FSHD2 families and therefore 18p- deletions may constitute a risk factor for FSHD. While other genetic and epigenetic risk factors are likely to contribute to disease penetrance, in order to be at an increased risk of developing FSHD2, two other factors also must be present. First, the presence of a chromosome 4qA containing a *DUX4* polyadenylation signal and second, the size of the D4Z4 repeat array with array sizes at the lower end of the normal repeat size spectrum being more sensitive to *SMCHD1* hemizygosity than longer arrays.

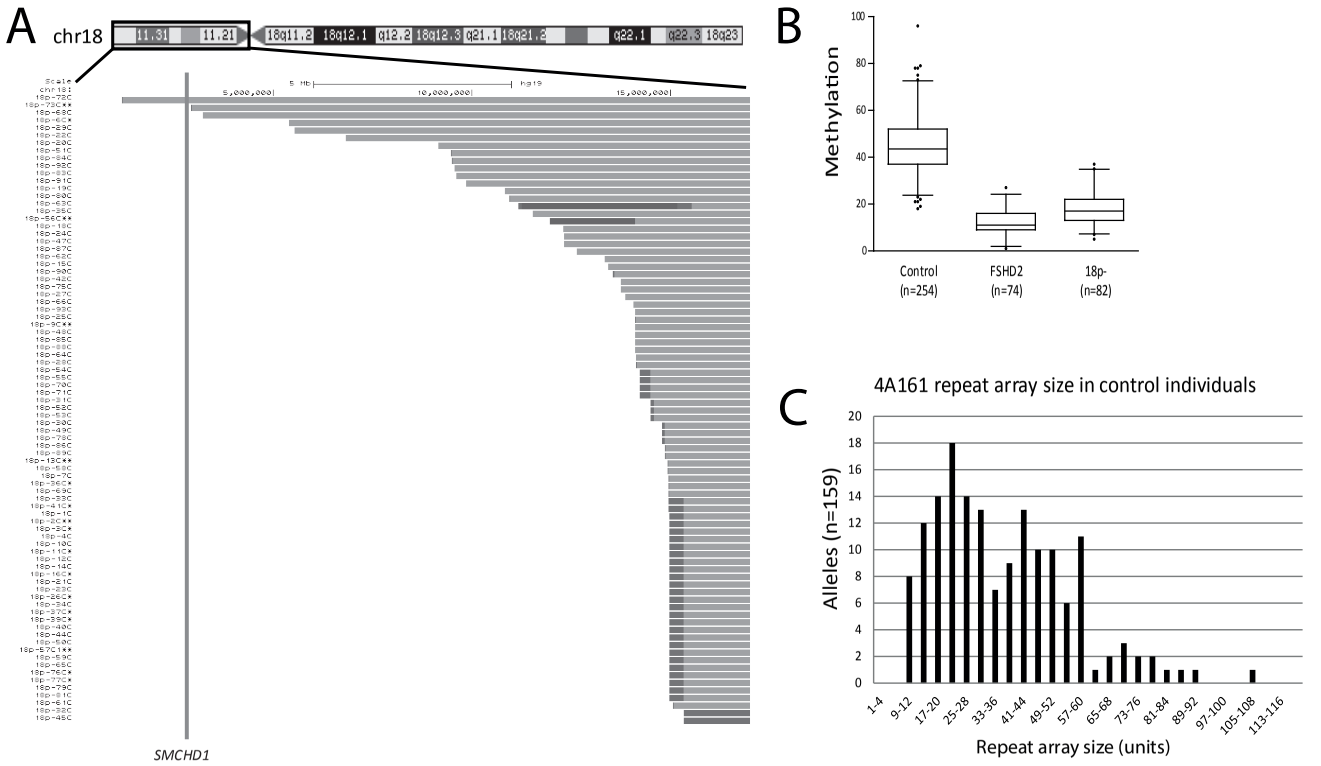


FIGURE 2. (a) Chromosome content of individuals with 18p-. The black box around the chromosome ideogram across the top of the figure indicates the region of the chromosome shown below. The study number of each participant is aligned to the left of their data. The intact region of the chromosome is shown as individual grey bars. The breakpoint region is shown in darker grey and regions of duplication are shown as black. The vertical line indicates the location of the *SMCHD1* gene. **(b)** The methylation value measured at D4Z4 for control individuals (disomic 18p), individuals that carry an *SMCHD1* mutation (FSHD2) and 18p- individuals that have a deletion of the *SMCHD1* gene (18p-). **(c)** Distribution of D4Z4 repeat array sizes on 159 independent 4A161 chromosomes in European control individuals[16; 23].

Acknowledgement

The authors thank all individuals with FSHD and 18p- as well as their family members for participating in this study.

Funding

The research leading to these results has received funding from the 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), The Prinses Beatrix Spierfonds (W.OR12-20), Spieren voor Spieren, NIH NIAMS AR066248, NIH NCRR UL1RR024160, the FSHD Global Research Foundation and the Chromosome 18 Registry and Research Society.

REFERENCES

1. Padberg, G.W., Lunt, P.W., Koch, M., and Fardeau, M. (1991). Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscular disorders : NMD 1, 231-234.
2. Tawil, R., and Van Der Maarel, S.M. (2006). Facioscapulohumeral muscular dystrophy. Muscle Nerve 34, 1-15.
3. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., et al. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics 6, e1001181.
4. Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nature genetics 2, 26-30.
5. van Overveld, P.G., Lemmers, R.J., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nature genetics 35, 315-317.
6. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics 44, 1370-1374.
7. Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., and Muller, C.R. (2014). Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. European journal of human genetics : EJHG.
8. Gabriels, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D., et al. (1999). Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene 236, 25-32.
9. Balog, J., Thijssen, P.E., de Greef, J.C., Shah, B., van Engelen, B.G., Yokomori, K., Tapscott, S.J., Tawil, R., and van der Maarel, S.M. (2012). Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics : official journal of the DNA Methylation Society 7, 579-584.
10. Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., et al. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Developmental cell 22, 38-51.

11. Zeng, W., de Greef, J.C., Chen, Y.Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A., et al. (2009). Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS genetics 5, e1000559.

12. Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329, 1650-1653.

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

14. Lemmers, R.J., Wohlgemuth, M., Frants, R.R., Padberg, G.W., Morava, E., and van der Maarel, S.M. (2004). Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. American journal of human genetics 75, 1124-1130.

15. de Greef, J.C., Lemmers, R.J., Camano, P., Day, J.W., Sacconi, S., Dunand, M., van Engelen, B.G., Kiuru-Enari, S., Padberg, G.W., Rosa, A.L., et al. (2010). Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology 75, 1548-1554.

16. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2014). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics.

17. Sacconi, S., Lemmers, R.J., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. American journal of human genetics 93, 744-751.

18. Ferguson, P.J., Chen, S., Tayeh, M.K., Ochoa, L., Leal, S.M., Pelet, A., Munnich, A., Lyonnet, S., Majeed, H.A., and El-Shanti, H. (2005). Homozygous mutations in LPIN2 are responsible for the syndrome of chronic recurrent multifocal osteomyelitis and congenital dyserythropoietic anaemia (Majeed syndrome). Journal of medical genetics 42, 551-557.

19. Martin-Lluesma, S., Stucke, V.M., and Nigg, E.A. (2002). Role of Hec1 in spindle checkpoint signaling and kinetochore recruitment of Mad1/Mad2. Science 297, 2267-2270.

20. Turleau, C. (2008). Monosomy 18p. Orphanet J Rare Dis 3, 4.

21. Sebold, C., Soileau, B., Heard, P., Carter, E., O'Donnell, L., Hale, D.E., and Cody, J.D. (2015). Whole arm deletions of 18p: medical and developmental effects. Am J Med Genet A 167A, 313-323.

22. Lemmers, R.J., van der Vliet, P.J., van der Gaag, K.J., Zuniga, S., Frants, R.R., de Knijff, P., and van der Maarel, S.M. (2010). Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete interchromosomal sequence transfers in human evolution. American journal of human genetics 86, 364-377.

23. Schaap, M., Lemmers, R.J., Maassen, R., van der Vliet, P.J., Hoogerheide, L.F., van Dijk, H.K., Basturk, N., de Knijff, P., and van der Maarel, S.M. (2013). 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.

SUPPLEMENTARY INFORMATION

Supplementary materials and methods

Subjects

For FSHD2 families, this study was approved by the institutional review board of the University of Rochester Medical Center. Families Rf898 and Rf929 were studied after informed consent. Clinical assessment was performed with a standardized clinical form available at the website of the Fields Center for FSHD Research (www.urmc.rochester.edu/fields-center/). The clinical severity score was based on the Ricci score and varied between 0 and 10 (0: unaffected – 10: wheelchair bound)[1]. Genetic diagnosis for FSHD2 was based on previously established genetic, and epigenetic criteria[2].

For the 18p- individuals, the study was approved by the University of Texas Health Science Center at San Antonio Institutional Review Board. Participants were eligible for the study if they had a clinically suspected or diagnosed abnormality of chromosome 18. Informed consent was appropriately documented at all stages of the process.

Genotyping and methylation analysis FSHD locus

High quality DNA was isolated from peripheral blood mononuclear cell. The detailed genotyping of the FSHD locus was based on the haplotype (SSLP variation), the D4Z4 repeat array size and the distal variation A and B by using pulsed field gel electrophoresis[3]. Methylation of the D4Z4 repeat was measured at the proximal unit of the D4Z4 arrays and the Delta1 and Delta2 values were calculated as described previously[4]. Detailed protocols can be downloaded from the Fields Center website (www.urmc.rochester.edu/fields-center/). D4Z4 repeat sizes on 4A161 alleles in the Caucasian population were derived from Schaap *et al.* and Lemmers *et al*[4; 5].

Mutation Analysis SMCHD1

Mutation analysis for *SMCHD1* was performed by Sanger sequencing for all 48 coding exons using intronic primers described previously[4]. Variant analysis was done using Sequence Pilot (JSI Medical Systems GmbH). The *SMCHD1* genomic sequence was obtained from Ensemble (build 37) [GRCh37:18:2655286:2805615] (Genomic Refseq: NG_031972.1, Transcript Refseq: NM_015295.2).

Whole exome sequencing analysis

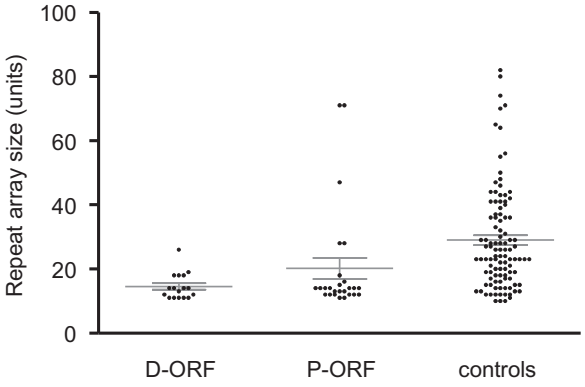
Whole exome sequencing was performed in the Neuromics project by deCODE Genetics (Reykjavik – Iceland). The WES data were analyzed using deCODE Clinical Sequence Miner. The report builder LOHZ (with standard settings) was used to identify regions with loss of heterozygosity.

SNP Arrays

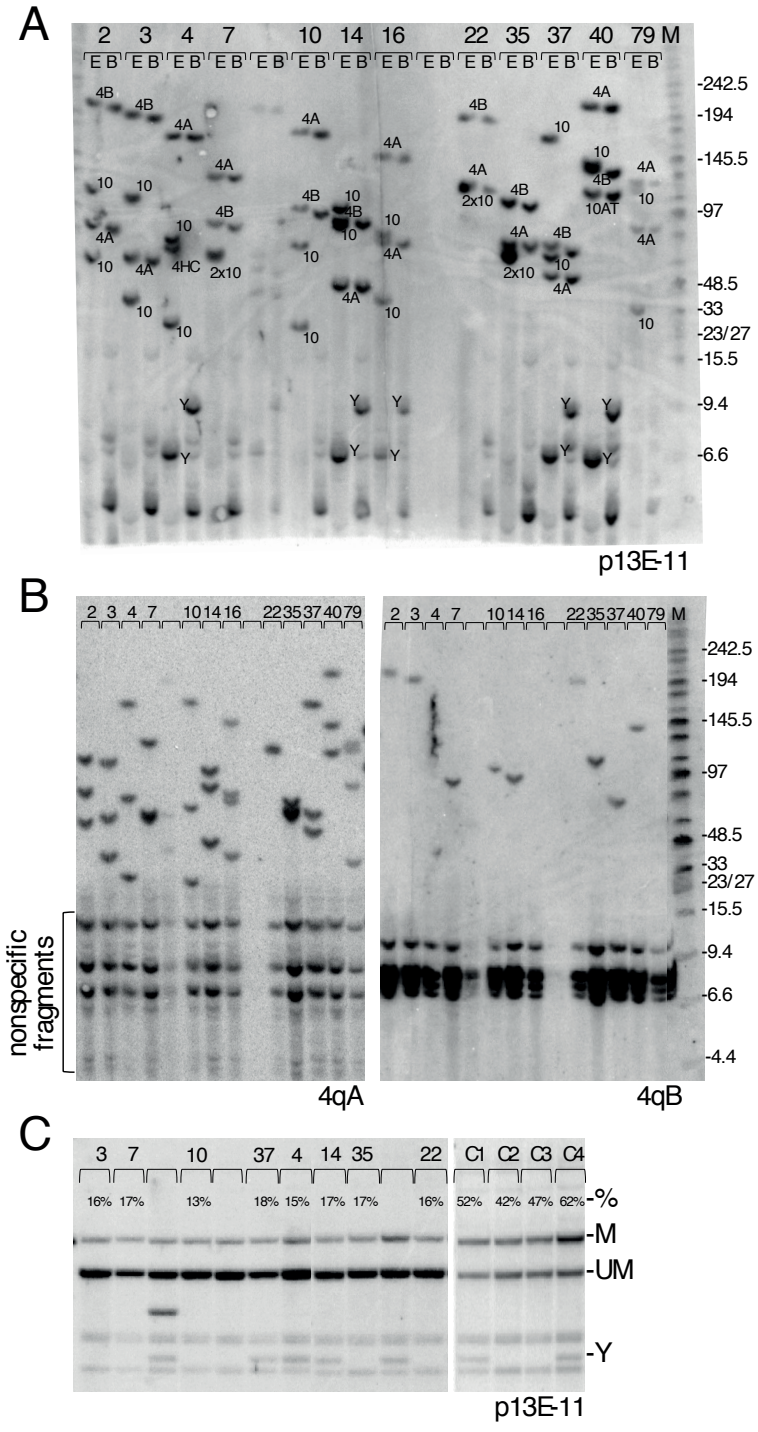
Genome wide array analysis on families Rf898 and Rf929 was performed on DNA using the Affymetrix Cytoscan HD Array platform (Affymetrix, Inc., Santa Clara, CA, USA) following the manufacturer's protocols. Copy number was assessed using the Affymetrix Chromosome Analysis Suite (ChAS) Software.

Comparative Genomic Hybridization (CGH) of the 18p deletion DNA samples was performed as previously described (Heard et al., 2009)[6]. The oligonucleotide arrays were produced by Agilent Technologies (Santa Clara, CA) and scanned using the Agilent laser scanner. The scan data were extracted using Agilent Feature Extraction (version 8.1.1) and then analyzed using the CGH Analytics 3.4.27 software.

Supplementary figures



SUPPLEMENTARY FIGURE S1. Shortest repeat array size on permissive 4qA allele in FSHD2 and controls. Distribution of the shortest D4Z4 repeat array size on permissive 4qA chromosomes in FSHD2 patients and controls. Previously, we categorized SMCHD1 mutations into disrupting ORF (D-ORF) mutations, that likely result in haploinsufficiency through nonsense mediated decay, and preserving ORF (P-ORF) mutations. The shortest repeat arrays were identified in controls (n=105) and FSHD2 patients with either D-ORF (n=17) or P-ORF (n=26) mutations[4]. The controls were randomly chosen from independent families that were not related to FSHD2 patients[4].


A
2 3 4 7 10 14 16 22 35 37 40 79 M
E B E B E B E B E B E B E B E B E B E B E B E B
-242.5
-194
-145.5
-97
-48.5
-33
-23/27
-15.5
-9.4
-6.6
p13E-11
B
2 3 4 7 10 14 16 22 35 37 40 79
2 3 4 7 10 14 16 22 35 37 40 79 M
nonspecific fragments
-242.5
-194
-145.5
-97
-48.5
-33
-23/27
-15.5
-9.4
-6.6
-4.4
4qA
4qB
C
3 7 10 37 4 14 35 22 C1 C2 C3 C4
16% 17% 13% 18% 15% 17% 17% 16% 52% 42% 47% 62%
-%
-M
-UM
-Y
p13E-11

SUPPLEMENTARY FIGURE S2. Example of genetic and methylation analysis in some 18p- individuals. **(a)** Representative example of D4Z4 repeat array size analysis in twelve 18p- individuals (18p-2, 18p-3, 18p-4, 18p-7, 18p-10, 18p-14, 18p-16, 18p-22, 18p-35, 18p-37, 18p-40 and 18p-79) by pulsed field gel electrophoresis followed by Southern blotting and hybridization with probe p13E-11[3]. Each DNA sample is double digested with *Eco*RI and *Hin*dIII (lane E) or *Eco*RI and *Bln*I (lane B). Lane E reveals all four repeat arrays on chromosomes 4 and 10, while lane B only show the chromosome 10-type fragments. Cross hybridizing chromosome Y fragments are indicated (Y). The molecular weight marker (in kb) is indicated at the right of the blot. **(b)** 4qA (left) and 4qB (right) hybridizations of *Hin*dIII digested DNA from the same individuals as in A[3]. Nonspecific fragments in the region between 15 kb and 7 kb (4qA) and 12 kb and 7kb (4qB) are Indicated. **(c)** Representative example of methylation analysis in eight 18p- individuals (18p-3, 18p-7, 18p-10, 18p-37, 18p-4, 18p-14, 18p-35 and 18p-22) and four controls (C1-C4) according to previous study [2]. Methylated (M) and unmethylated (UM) D4Z4 fragments are indicated. Below each lane the methylation value is indicated in %. Y indicates cross hybridizing chromosome Y fragment.

Supplementary tables

SUPPLEMENTARY TABLE S1. Clinical Data of the FSHD Families

Rf	Nr	Sex	AAE	Age at onset	CSS	Diagnosis
Rf898	301	M	36	13	5	FSHD2
	202	F	65	asymptomatic	5	FSHD2
	204	F	59	asymptomatic	0	unaffected
	205	M	55	10	5	FSHD2
	303*	F	36	24	0	FSHD2
Rf929	201	M	73	30	8	FSHD2
	204	F	70	5	6	FSHD2
	206	F	67	asymptomatic	0	unaffected

CSS, clinical severity score (ranging from 0-10) based on the Ricci score multiplied by two[1]; AAE, age at the last examination
* 898.303 reported muscle pain at age 24, EMG shows myopathic units at age 36

SUPPLEMENTARY TABLE S2. D4Z4 repeats array size and methylation in 19 individuals with 18p-

ID	Age	Sex	D4Z4 units	OM	Delta1	Delta2
18p-2	26	F	24	13	-39	-2
18p-3	25	F	18	16	-31	2
18p-4	21	M	49	15	-29	3
18p-7	21	F	36	17	-29	4
18p-10	29	F	51	13	-33	0
18p-14	25	M	13	17	-27	5
18p-16	26	M	22	13	-33	0
18p-20	17	M	57	14	-38	-2
18p-22	21	F	35	16	-40	-1
18p-25	22	F	29	35	-19	19
18p-30	20	F	30	23	-31	7
18p-35	25	F	21	17	-28	4
18p-37	23	M	15	18	-26	6
18p-40	59	M	33	24	-33	6
18p-41	16	F	60	14	-37	-1
18p-48	13	M	34	28	-27	11
18p-49	28	F	22	18	-23	7
18p-79	28	F	25	19	-27	6

D4Z4 units is the shortest repeat array size on the 4A161 chromosome
Methylation values (Delta1; Δ1, Delta2; Δ2 and the measured methylation; OM)
Individuals that carry a repeat array size ≤16 units on a 4A161 allele are highlighted in grey.

REFERENCES

1. Ricci, E., Galluzzi, G., Deidda, G., Cacurri, S., Colantoni, L., Merico, B., Piazzo, N., Servidei, S., Vigneti, E., Pasceri, V. (1999). Progress in the molecular diagnosis of facioscapulohumeral dystrophy and correlation between the number of KpnI repeat at the 4q35 locus and clinical phenotype. Ann Neurol 45, 751-757.
2. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics 44, 1370-1374.
3. Lemmers, R.J., van der Vliet, P.J., van der Gaag, K.J., Zuniga, S., Frants, R.R., de Knijff, P., and van der Maarel, S.M. (2010). Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete interchromosomal sequence transfers in human evolution. American journal of human genetics 86, 364-377.
4. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2014). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics.
5. Schaap, M., Lemmers, R.J., Maassen, R., van der Vliet, P.J., Hoogerheide, L.F., van Dijk, H.K., Basturk, N., de Knijff, P., and van der Maarel, S.M. (2013). 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.
6. Heard, P.L., Carter, E.M., Crandall, A.C., Sebold, C., Hale, D.E., and Cody, J.D. (2009). High resolution genomic analysis of 18q- using oligo-microarray comparative genomic hybridization (aCGH). Am J Med Genet A 149A, 1431-1437.

Marlinde L van den Boogaard[1], Richard JFL Lemmers[1], Judit Balog[1], Patrick J van der Vliet[1], Nienke van der Stoep[2], Stephen J Tapscott[3], Sabrina Sacconi[4,5], Rabi Tawil[6], Silvère van der Maarel[1]

1 Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
2 Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
3 Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
4 Neuromuscular Diseases Centre, University Hospital of Nice, Nice, France
5 CNRS UMR7277, IBV, Faculty of Medicine, University of Nice, Nice, France
6 Department of Neurology, University of Rochester Medical Center, Rochester, USA

4

INTRONIC *SMCHD1* VARIANTS IN FSHD

In preparation

ABSTRACT

Fascioscapulohumeral dystrophy is associated with partial chromatin relaxation of the *DUX4* encoding D4Z4 macrosatellite repeat located on chromosome 4, and transcriptional derepression of *DUX4* in skeletal muscle. The most common form, FSHD1, is caused by a D4Z4 repeat array contraction to 1-10 units (normal range 8-100 units). The less common form, FSHD2, is most often caused by heterozygous variants in *SMCHD1*, which encodes a chromatin modifier which binds to D4Z4 to maintain a repressed chromatin state. In this study we identified intronic variants in *SMCHD1* in two FSHD families. In the first family we identified a variant 15 nucleotides proximal from the 3' splice site of exon 14. This *SMCHD1* variant creates a 3' splice site, which results in partial intron retention with inclusion of the distal 14 nucleotides of intron 13 into the transcript. In the second family we identified a deep intronic variant in intron 34. This *SMCHD1* variant creates a 3' splice site in intron 34, which results in exonisation of 53 nucleotides of intron 34. In this family the deep intronic variant acts as a modifier of disease severity. In both families the aberrant transcripts are predicted to lead to a premature stop codon. The identification of these intronic variants further expands the *SMCHD1* mutation spectrum in FSHD2 and emphasizes the importance of screening for intronic variants in *SMCHD1*.

INTRODUCTION

Fascioscapulohumeral dystrophy (FSHD, [OMIM 158900 and 158901]) is a common muscular dystrophy in adults (prevalence ~1:8.000) and is clinically mainly characterized by progressive weakness and wasting of the facial, shoulder girdle, trunk and upper arm muscles[1; 2]. Most often the onset of the disease occurs during the second decade of life. However, both within and between families there is a large variability in disease onset and progression[3]. Two genetic forms of FSHD have been identified, FSHD1 and FSHD2, which are clinically almost indistinguishable[4], but seem to represent opposite extremes of a disease spectrum[5]. Both forms are associated with partial chromatin relaxation of the D4Z4 macrosatellite repeat array on chromosome 4 in somatic tissue, characterized by reduced CpG methylation and loss of repressive histone marks[6-8]. This chromatin relaxation results in transcriptional derepression of the D4Z4 encoded *DUX4* gene in skeletal muscle[9]. DUX4 is a transcription factor normally expressed in the germ line and cleavage stage embryos, that is normally suppressed in most other somatic tissues[9-12]. DUX4 causes cell death when overexpressed in somatic cell lines or endogenously but inappropriately expressed in FSHD myotubes[13; 14]. The D4Z4 chromatin relaxation in FSHD must occur on a permissive chromosome 4 (4qA haplotype), which contains a polymorphic *DUX4* polyadenylation signal distal to the D4Z4 repeat array[15]. This polyadenylation signal is required for the production of stable DUX4 mRNA in somatic cells. Chromatin relaxation on the homologous D4Z4 repeats on non-permissive 4qB or 10q chromosomes do not cause FSHD since these chromosomal backgrounds lack a somatic *DUX4* polyadenylation signal[15].

FSHD1, the most common form of FSHD (>95%), is caused by contraction of the D4Z4 repeat array to 1-10 units on a 4qA chromosome[16]. FSHD2 is most often caused by heterozygous variants in *structural maintenance of chromosomes flexible hinge domain containing 1* (*SMCHD1*) in combination with a smaller, but normal-sized permissive D4Z4 repeat array (8-20) on a 4qA chromosome[5; 17]. SMCHD1 is an atypical member of the SMC gene superfamily and originally identified as regulator of epigenetic silencing in an ENU mutagenesis screen in mice[18; 19]. SMCHD1 is normally binding to the D4Z4 repeat, thereby repressing DUX4 expression in somatic cells[17]. FSHD2 patients with an *SMCHD1* variant show reduced binding of SMCHD1 to the D4Z4 repeat, which leads to D4Z4 chromatin relaxation and *DUX4* expression in skeletal muscle[17]. Additionally, SMCHD1 is also a modifier of disease severity and progression in FSHD1 since *SMCHD1* variants have been identified in some unusually severely affected members of FSHD1 families who carry both a contracted repeat array and an *SMCHD1* variant[20]. For some FSHD2 patients, however, we found D4Z4 hypomethylation but could not identify an (exonic) *SMCHD1* variant. Two of these families are now explained by *SMCHD1* hemizygosity[21]. In two other unexplained FSHD2 families heterozygous variants in *DNA*

methyltransferase 3B (*DNMT3B*) have recently been identified[22]. DNMT3B is one of the *de novo* DNA methyltransferases[23], and is likely important for establishing a repressed D4Z4 chromatin structure in somatic cells.

Since the discovery of *SMCHD1* as the most common FSHD2 gene, disease causing variants in *SMCHD1* have been identified in approximately 80 FSHD2 families. The mutation spectrum of *SMCHD1* in FSHD2 includes missense, nonsense, and splice site variants, insertions and deletions[5; 17; 20; 21; 24-28]. In this study we describe two families with an intronic variant in *SMCHD1* which results in aberrant SMCHD1 transcripts. In the first family an intronic variant in *SMCHD1* was identified which alters splicing and results in partial intron retention. In the second family a deep intronic variant in *SMCHD1* was identified, resulting in exonisation of 53 nucleotides of intron 34. These variants further expand the *SMCHD1* mutation spectrum in FSHD2.

MATERIAL METHODS

Subjects

A French family (Rf744, fig. 1A) and an American family (Rf1034, fig. 1B) were studied after informed consent and the study protocol was approved by the relevant institutional review boards. Clinical assessment of disease severity was performed using the 11 point (0: unaffected – 10: wheelchair bound) standardized Clinical Severity Score (CSS)[29].

D4Z4 repeat sizing, haplotype analysis and methylation analysis

For genotyping high quality genomic DNA was isolated from peripheral blood mononuclear cells (PBMCs). The sizing of the D4Z4 repeats on chromosomes 4 and 10 was done by pulsed field gel electrophoresis (PFGE) as described previously[15]. Haplotype analysis was done by hybridization of PFGE blots with probes specific for the 4qA and 4qB haplotype in combination with PCR-based SSLP analysis according to previously described protocols[15]. D4Z4 methylation analysis was measured using the FseI restriction site in the most proximal unit of the D4Z4 arrays on chromosomes 4 and 10 as published previously[17]. The Delta1 value of D4Z4 methylation was calculated as described in Lemmers et al. 2014[5].

Genomic SMCHD1 variant analysis

For the index cases *SMCHD1* variant analysis of all coding exons and splice regions was performed by Sanger sequencing after PCR amplification. The intronic primers were located at a position of at least 50 nucleotides from the splice donor or acceptor site and were previously published[5]. For Rf1034 a PCR was performed in intron 34 to identify a deep

intronic variant using primers intron_34fwd (5'-TTGAAATACAAAACTGTCGCTTAGA-3') and intron_34rev (5'-AGGGGGAAGGAATTCAAAGA-3'). The PCR product was analysed by Sanger sequencing.

The *SMCHD1* genomic sequence was obtained from Ensembl human assembly GRCh37 [GRCh37:18:2655286:2805615] (Genomic Refseq: NG_031972.1, Transcript Refseq: NM_015295.2), exons were numbered like in NG_031972.1. The functional consequences of *SMCHD1* variants were predicted using Alamut Visual version 2.6 (Interactive Biosoftware, Rouen, France).

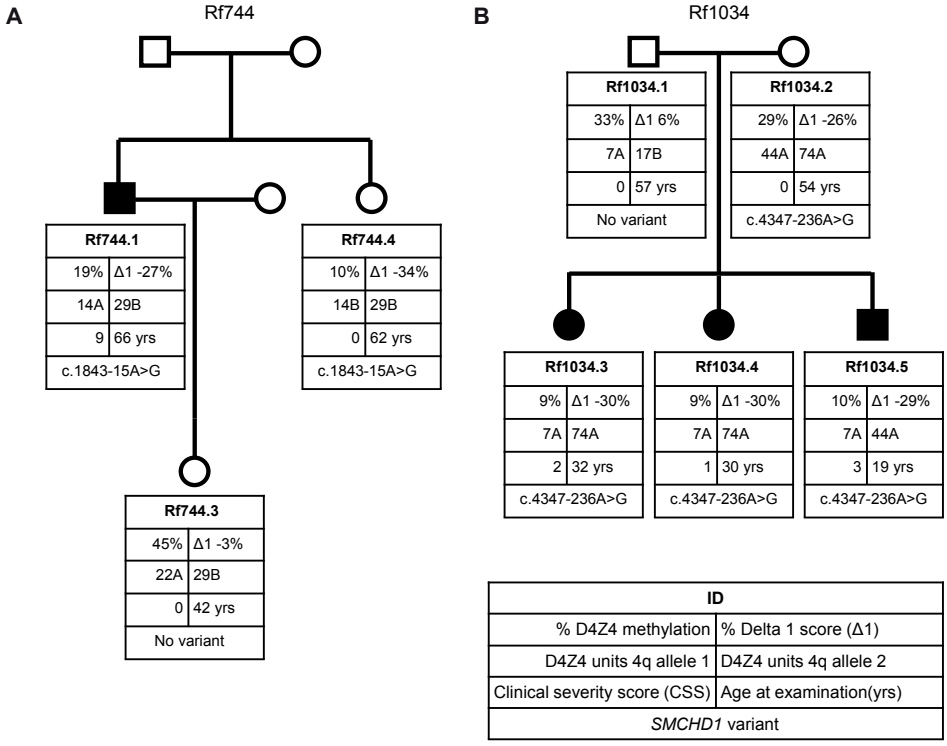


FIGURE 1. Pedigrees of families Rf744 (A) and Rf1034 (B). Clinically affected individuals are indicated in black. The following information is provided: the family identifier, D4Z4 methylation, Delta1 score, the size and type (A permissive, B non-permissive) of 4q-linked D4Z4 repeats, the clinical severity score, the age of examination and the *SMCHD1* variant. Key is shown below.

RNA analysis

RNA was isolated from PAXgene Blood RNA Tubes using the PAXgene Blood RNA Kit (PreAnalytiX, a Qiagen/BD company). cDNA was synthesized with 800 ng to 2000 ng of RNA using the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific) using random hexamer primers. A reverse transcriptase polymerase chain reaction (RT-PCR) for *SMCHD1* exon 12 to 16 was performed using primers 1482F (5'-TCCTAAGAAGAGAGGGCTTGC-3') and 2105R (5'-TCATCTCCTTCAGGCCAAGT-3'). A RT-PCR for *SMCHD1* exon 32 to 35 was performed using primers 4098F (5'-AAAACCCGTTCGTCTCAATG-3') and 4406R (5'-TCCATCATAAAACCAAACTGGA-3'). RT-PCRs were performed in 30 µl reactions using 0.5 units DreamTaq DNA polymerase (5U/µl Thermo Fisher Scientific), 1x DreamTaq buffer (Thermo Fisher Scientific), 3 µl of dNTPs (2mM of each nucleotide) and 25 pmol of each primer. The following RT-PCR protocol was used: 95°C 5 min, 35 cycles of: 95°C 30 sec, 60°C 30 sec, 72°C 30 sec, then 72°C 10 min. RT-PCR products were separated by size on 2% agarose gels after which PCR products were gel purified (NucleoSpin® Gel and PCR Clean-up, Machery Nagel). Purified PCR products were cloned into a pCR™4-TOPO vector (Invitrogen, Life Technologies) and transformed in DH5α heat-shock competent cells (Subcloning Efficiency DH5α Competent Cells, Invitrogen, Life Technologies). Multiple clones were analysed by sequencing their insert to find the sequence of the altered transcript.

RESULTS

D4Z4 length and methylation analysis in Rf744 and Rf1034 individuals

Index case Rf744.1 was suspected of FSHD based on physical examination with a CSS of 9 at age 66. His physical examination showed asymmetric scapular winging, right foot drop, asymmetric distribution of facial weakness, symmetric weakness of fixator shoulder girdle muscles, weakness of the pelvic girdle muscles, humeral weakness involving both biceps and triceps brachii, abdominal weakness with positive Beevor's sign and tibialis anterior weakness. Rf744.1 also has a benign myelodysplastic syndrome. D4Z4 repeat length and haplotype analysis showed that the shortest permissive D4Z4 allele of Rf744.1 consists of 14 units (Fig. 1A). D4Z4 methylation was measured and the Delta1 value was calculated. The Delta 1 value ranges between -42% and -22% (5th and 95th percentile, respectively) in carriers of an *SMCHD1* variant which affects functions[5]. D4Z4 methylation analysis in Rf744.1 revealed a FseI methylation level of 19% (Delta1 value -27%), indicative of FSHD2. The unaffected sister of the proband (Rf744.4) also

shows D4Z4 hypomethylation but she does not carry a permissive allele. The daughter of the proband (Rf744.3) does not show D4Z4 hypomethylation and she is unaffected (Fig. 1A).

Index case Rf1034.5 was suspected of FSHD based on physical examination with a CSS of 3 at age 19 [29]. His physical examination showed a combination of pectus excavatum, progressive weakness of the right arm, bilateral scapular winging, facial weakness, and Beevor's sign. D4Z4 repeat length and haplotype analysis showed that Rf1034.5 carries a 7 units D4Z4 repeat on a permissive chromosome and D4Z4 hypomethylation (Delta1 score -29%), suggestive for both FSHD1 and FSHD2 (Fig. 1B). Additional family-member material was obtained and D4Z4 repeat sizes, haplotypes and D4Z4 methylation levels were determined and a physical examination was performed for the four additional family members of Rf1034[22] (Fig. 1B). The father (Rf1034.1) of the proband carries a 7 unit D4Z4 repeat array, but he is unaffected. The unaffected mother (Rf1034.2) of the proband shows D4Z4 hypomethylation and she carries two permissive 4qA alleles of 44 and 74 units. The two sisters (Rf1034.3 and Rf1034.4) of the proband both carry the 7 unit D4Z4 repeat array as well as D4Z4 hypomethylation and they are also affected. The physical examination of Rf1034.3 showed a combination of weakness of the scapular stabilizers and weakness of the right arm. The physical examination of Rf1034.4 showed only weakness of the facial muscles. This family information strengthened the suggestion that there is a combination of FSHD1 and FSHD2 in this family.

TABLE 1. Splice site predictions in *SMCHD1*

	3'splice site c.4347-236A>G	5'splice site c.4347-183	3' splice site c.1843-15A>G
SpliceSiteFinder-like (0-100)	87.4	94.7	89.9
MaxEntScan (0-16)	8.9	10.8	7.4
NNSPLICE (0-1)	0.9	1	1
GeneSplicer (0-15)	5.7	0.54	5.1
Human Splicing Finder (0-100)	89.4	97.7	86.1

Identification of an intronic variant in *SMCHD1* in Rf744

SMCHD1 variant analysis of all coding exons and splice regions in Rf744 identified an intronic *SMCHD1* variant in Rf744.1. This variant (c.1843-15A>G) is located 15 base pairs proximal to exon 14 and various splicing prediction tools predict that this variant creates a 3' splice site (Fig. 2A, Table 1). The variant was also identified in Rf744.4, which also shows D4Z4 hypomethylation, but not in Rf744.3 without D4Z4 hypomethylation (Fig.

S1A). To investigate whether this variant alters the transcript, an RT-PCR from *SMCHD1* exon 12 to 16 was performed and analysed by gel electrophoresis. Besides the normal PCR product of the expected size, two longer PCR products were identified (Fig. 2B). Sanger sequencing of TOPO clones of those two additional PCR products identified that the altered transcript contains the sequence from c.1843-14 to c.1843-1, confirming that c.1843-15A>G creates a 3' splice site (Fig. 2C, S1B). The inclusion of these 14 nucleotides is predicted to disrupt the open reading frame with a premature stop codon in exon 14. No other sequences were identified, suggesting that the highest band in the gel is a heteroduplex of the normal and altered transcript. No RNA was available from Rf744.3 and Rf744.4.

Identification of a deep intronic variant in *SMCHD1* in Rf1034

SMCHD1 variant analysis in all *SMCHD1* exons and splice regions in the proband did not identify any putative *SMCHD1* variants that affect function[5]. Therefore, whole exome sequencing (WES) was performed in Rf1034 but this did not identify a causative variant. By serendipity, an RT-PCR targeting *SMCHD1* exon 32 to 35 followed by agarose gel electrophoresis revealed two PCR products for Rf1034.3, the normal PCR product with the expected size and a PCR product that was larger than expected (Fig. 3A). This larger PCR product was also identified with an RT-PCR for Rf1034.2, Rf1034.4 (Fig. 3A) and Rf1034.5 (not shown), while it was absent in Rf1034.1 (Fig. 3A). This additional PCR product contained a sequence corresponding to 53 nucleotides of intron 34, from c.-235 to c.-183 proximal to exon 35 (Fig S2A). These 53 nucleotides are included in the transcript as a new exon and are predicted to disrupt the open reading frame and lead to a premature stop codon in exon 35 (Fig. S2A). Subsequently, an intronic PCR was performed, followed by Sanger sequencing, to identify the variant which is responsible for this new exon. A heterozygous deep intronic variant (c.4347-236A>G, g.2760414A>G) in *SMCHD1* was identified in individuals Rf1034.2, Rf1034.3, Rf1034.4, and Rf1034.5, which was absent in Rf1034.1 (Fig S2B). Various splicing prediction tools predict that this variant creates a 3' splice site, while a cryptic 5' splice site is already predicted in the reference sequence at position c.4347-183 (Table 1). In this family this deep intronic variant in *SMCHD1* segregates with D4Z4 hypomethylation and modifies disease severity.

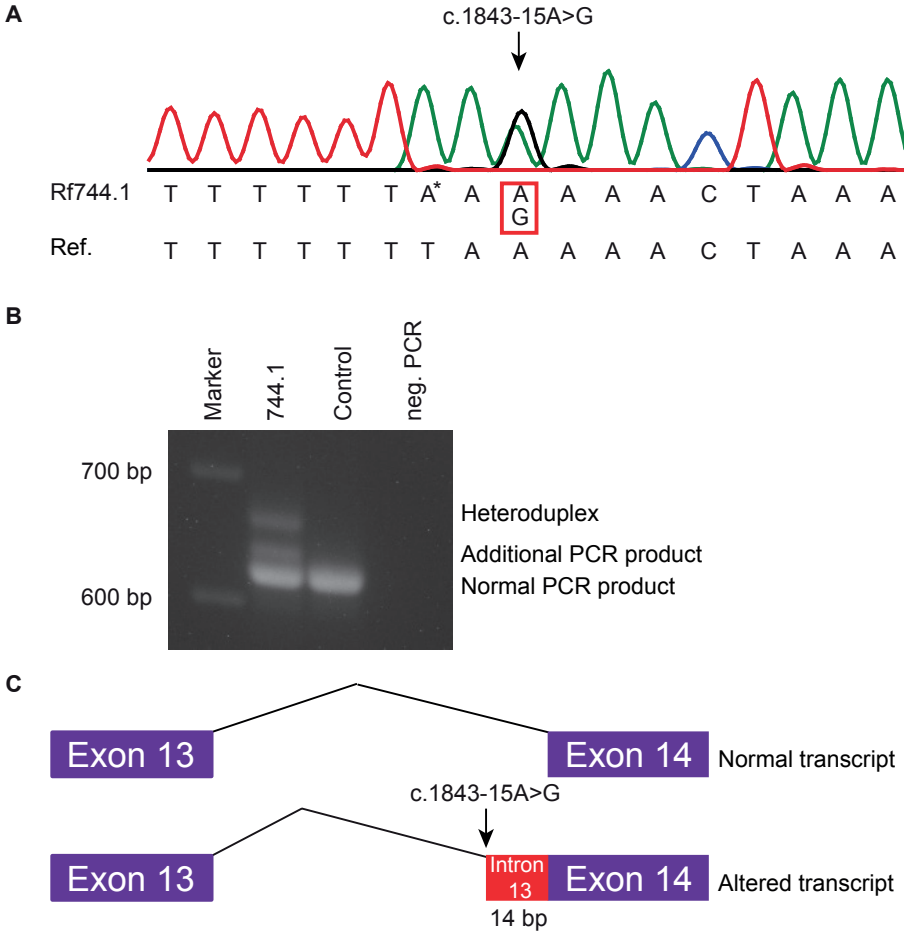


FIGURE 2. Identification of intronic variant in *SMCHD1* in Rf744. A) Sanger sequence track from Rf744.1 showing the intronic variant in *SMCHD1* at position c.1843-15, highlighted with a red rectangle. * indicates common SNP rs8090988 (T/A, ancestral T, minor allele frequency 0.33 (A)) B) Gel of RT-PCR of *SMCHD1* exon 12 to 16 in Rf744.1, a control and a negative PCR (no DNA). C) Schematic representation of splicing of the normal transcript and the altered transcript containing the intronic variant

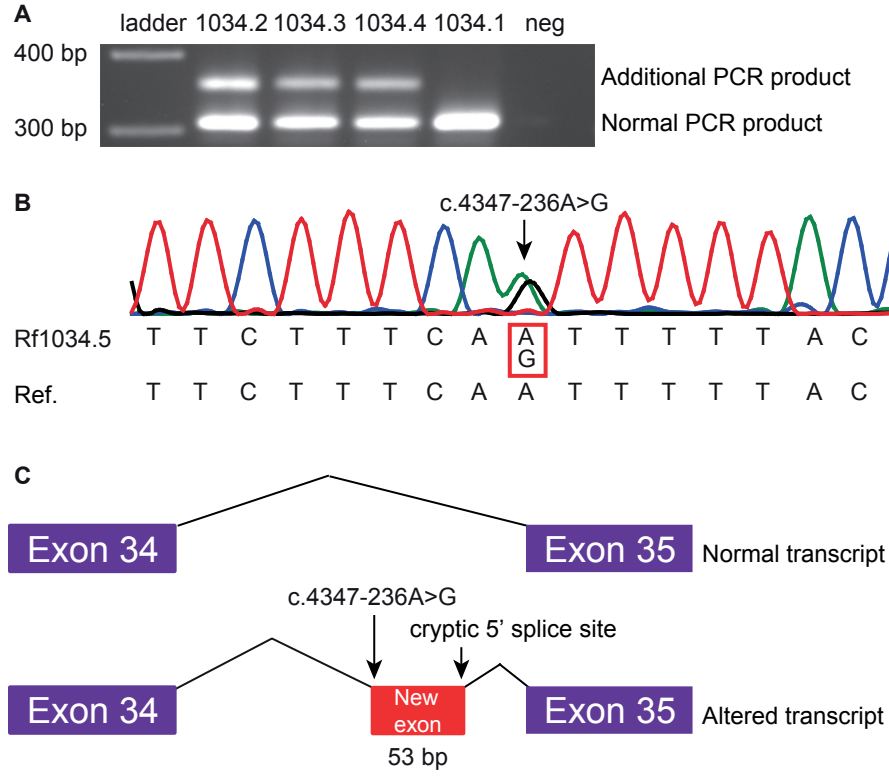


FIGURE 3. Identification of deep intronic variant in *SMCHD1* in Rf1034. A) Gel of RT-PCR of *SMCHD1* exon 32 to 35 in four members of family Rf1034 and a negative PCR (no DNA), B) Sanger sequence track showing the deep intronic variant in *SMCHD1* at position c.4347-236 in Rf1034.5, highlighted with a red rectangle C) Schematic representation of splicing of the normal transcript and the altered transcript containing the deep intronic variant and showing the exonisation of 53 basepairs in red.

DISCUSSION

In this study we identified a deep intronic variant in *SMCHD1* to act as a modifier for disease severity in an FSHD1 family. Furthermore, we identified an intronic variant in *SMCHD1* in an FSHD2 family.

In family Rf744 we identified an intronic variant located at 15 base pairs proximal to exon 14, which creates a 3' splice site. This *SMCHD1* variant results in the inclusion of the distal 14 nucleotides of intron 13 (c.1843-14 to c.1843-1) into the transcript, which is predicted to disrupt the open reading frame with a premature stop codon in exon 14.

The intronic variant and D4Z4 hypomethylation were also detected in the unaffected sister of the proband. She carries two non-permissive alleles, which explains why she remained unaffected. The unaffected daughter of the proband does not carry the variant and shows no D4Z4 hypomethylation.

In family Rf1034, exonic *SMCHD1* variant analysis by *SMCHD1* Sanger sequencing and WES did not identify any variants that affect function in *SMCHD1* or elsewhere in the genome[5]. However, in this study, by serendipity, a deep intronic variant was identified, which segregates with D4Z4 hypomethylation. This *SMCHD1* variant creates a 3' splice site in intron 34 and this results in exonisation of 53 nucleotides of intron 34. Inclusion of these 53 nucleotides in the transcript is predicted to disrupt the open reading frame and to result in a premature stop codon in exon 35. In family Rf1034 this *SMCHD1* variant acts as a modifier for disease severity. The proband and his two sisters all carry both a permissive D4Z4 repeat array of 7 units and the deep intronic variant in *SMCHD1* and present an FSHD phenotype. The proband is more severely affected than his sisters, indicating clinical variability, which is common in FSHD[5]. The mother (Rf1034.2) carries the deep intronic variant in *SMCHD1* and two permissive 4qA alleles of 44 and 74 units, while the median repeat size in controls is 23 units. The length of the D4Z4 repeats of the mother is much longer than the median length of the shortest permissive allele in FSHD2 patients, which is only 13 units[5]. Probably, the permissive allele of the mother is too long to develop FSHD2. This has also been shown in other FSHD2 families, where carriers of an *SMCHD1* variant are most often only affected with FSHD when they also carry a relatively short but normal sized permissive D4Z4 repeat of 11-20 units[5]. The father (Rf1034.1) carries an FSHD1 sized allele of 7 units and is unaffected. Non-penetrance and mild phenotypes are seen more often in carriers of a 7-10 units FSHD1 size allele[30]. Additionally, in 1-3% of the control population D4Z4 repeats of 7–10 units on disease permissive 4qA chromosomes are found, indicating the reduced penetrance of these alleles[31; 32]. In conclusion, in Rf1034 only the combination of a permissive D4Z4 repeat array of 7 units with the deep intronic variant in *SMCHD1* causes an FSHD phenotype, illustrating this *SMCHD1* variant modifies disease severity. This modifying role of *SMCHD1* variants has been described in multiple FSHD1 families with upper sized FSHD1 repeat arrays, which explains clinical variability in these families[20; 24].

The variants identified in this study affect splicing by introducing new 3' splice sites in *SMCHD1* outside the consensus sequence. Previously, an intronic *SMCHD1* variant with a similar effect as the variant in Rf744 was identified in another FSHD2 patient (Rf1352 in Lemmers et al. 2014)[5]. The variant c.3634-19A>G creates a 3' splice site, which results in the inclusion of the distal 18 nucleotides of intron 28 into the transcript and introduces a premature stop codon immediately proximal to exon 29[5]. In total, we have identified approximately 100 variants in *SMCHD1* which affect function (ref. 5,17,20,21,22 and

unpublished results), including 3 intronic variants outside the consensus sequence that introduce a 3' splice site. This indicates that the frequency of intronic variants in *SMCHD1* that introduce a new splice site is approximately 3% and that this type of variants might explain FSHD in patients in which no variant was identified in the exonic SMCHD1 region or in the splice site consensus.

Therefore, it would be useful to perform whole genome sequencing in FSHD2 patients without exonic *SMCHD1* variants, in combination with RT-PCR to identify alternative splicing. However, the products of such variants might be masked by efficient nonsense mediate decay (NMD), which would make it difficult to study the splicing effect of intronic variants. One way to address this issue would be to culture cells from blood of FSHD patients and controls in the presence of cycloheximide to block NMD. Alternatively, intronic variants might influence expression levels of *SMCHD1*. Furthermore, variants in the promoter or regulatory regions of *SMCHD1* might cause FSHD2. The functional consequences of these types of variants will be difficult to predict since the information on regulatory regions of *SMCHD1* is limited. Recently, in two FSHD families a variant in a putative regulatory region of *SMCHD1* was identified, however segregation with D4Z4 hypomethylation was inconclusive[33]. The functional effects of variants in regulatory regions could be studied with reporter assays in combination with segregation analysis of D4Z4 hypomethylation.

In summary, this report expands the *SMCHD1* mutation spectrum in FSHD2 with two intronic variants in *SMCHD1*. Both variants lead to aberrant splicing and the altered SMCHD1 transcripts are predicted to lead to a premature stop codon. Our study also highlights the importance of the additional variant screening in FSHD2 patients negative for exonic *SMCHD1* variants.

REFERENCES

1. Deenen, J.C., Arnts, H., van der Maarel, S.M., Padberg, G.W., Verschuuren, J.J., Bakker, E., Weinreich, S.S., Verbeek, A.L., and van Engelen, B.G. (2014). Population-based incidence and prevalence of facioscapulohumeral dystrophy. Neurology 83, 1056-1059.
2. Mul, K., Lassche, S., Voermans, N.C., Padberg, G.W., Horlings, C.G., and van Engelen, B.G. (2016). What's in a name? The clinical features of facioscapulohumeral muscular dystrophy. Pract Neurol 16, 201-207.
3. Statland, J.M., and Tawil, R. (2011). Facioscapulohumeral muscular dystrophy: molecular pathological advances and future directions. Current opinion in neurology 24, 423-428.
4. de Greef, J.C., Lemmers, R.J., Camano, P., Day, J.W., Sacconi, S., Dunand, M., van Engelen, B.G., Kiuru-Enari, S., Padberg, G.W., Rosa, A.L., et al. (2010). Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology 75, 1548-1554.
5. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2014). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics.
6. Balog, J., Thijssen, P.E., de Greef, J.C., Shah, B., van Engelen, B.G., Yokomori, K., Tapscott, S.J., Tawil, R., and van der Maarel, S.M. (2012). Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics : official journal of the DNA Methylation Society 7, 579-584.
7. van Overveld, P.G., Lemmers, R.J., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nature genetics 35, 315-317.
8. Zeng, W., de Greef, J.C., Chen, Y.Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A., et al. (2009). Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS genetics 5, e1000559.
9. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., et al. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics 6, e1001181.
10. Hendrickson, P.G., Dorais, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nature genetics.
11. Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. Nature genetics.

12. De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nature genetics.

13. Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O., et al. (2007). The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscular disorders : NMD 17, 611-623.

14. Rickard, A.M., Petek, L.M., and Miller, D.G. (2015). Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Human molecular genetics 24, 5901-5914.

15. Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329, 1650-1653.

16. Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nature genetics 2, 26-30.

17. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics 44, 1370-1374.

18. Hirano, T. (2005). SMC proteins and chromosome mechanics: from bacteria to humans. Philosophical transactions of the Royal Society of London Series B, Biological sciences 360, 507-514.

19. Blewitt, M.E., Vickaryous, N.K., Hemley, S.J., Ashe, A., Bruxner, T.J., Preis, J.I., Arkell, R., and Whitelaw, E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proceedings of the National Academy of Sciences of the United States of America 102, 7629-7634.

20. Sacconi, S., Lemmers, R.J., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. American journal of human genetics 93, 744-751.

21. Lemmers, R.J., van den Boogaard, M.L., van der Vliet, P.J., Donlin-Smith, C.M., Nations, S.P., Ruivenkamp, C.A., Heard, P., Bakker, B., Tapscott, S., Cody, J.D., et al. (2015). Hemizygosity for SMCHD1 in Facioscapulohumeral Muscular Dystrophy Type 2: Consequences for 18p Deletion Syndrome. Hum Mutat 36, 679-683.

22. van den Boogaard, M.L., Lemmers, R.J., Balog, J., Wohlgemuth, M., Auranen, M., Mitsuhashi, S., van der Vliet, P.J., Straasheijm, K.R., van den Akker, R.F., Kriek, M., et al. (2016). Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. American journal of human genetics 98, 1020-1029.

23. Jeltsch, A., and Jurkowska, R.Z. (2016). Allosteric control of mammalian DNA methyltransferases - a new regulatory paradigm. Nucleic acids research 44, 8556-8575.

24. Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., and Muller, C.R. (2014). Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. European journal of human genetics : EJHG.

25. Mitsuhashi, S., Boyden, S.E., Estrella, E.A., Jones, T.I., Rahimov, F., Yu, T.W., Darras, B.T., Amato, A.A., Folkerth, R.D., Jones, P.L., et al. (2013). Exome sequencing identifies a novel SMCHD1 mutation in facioscapulohumeral muscular dystrophy 2. Neuromuscular disorders : NMD 23, 975-980.

26. Winston, J., Duerden, L., Mort, M., Frayling, I.M., Rogers, M.T., and Upadhyaya, M. (2014). Identification of two novel SMCHD1 sequence variants in families with FSHD-like muscular dystrophy. European journal of human genetics : EJHG.

27. van den Boogaard, M.L., Jfl Lemmers, R., Camano, P., van der Vliet, P.J., Voermans, N., van Engelen, B.G., Lopez de Munain, A., Tapscott, S.J., van der Stoep, N., Tawil, R., et al. (2016). Double SMCHD1 variants in FSHD2: the synergistic effect of two SMCHD1 variants on D4Z4 hypomethylation and disease penetrance in FSHD2. European journal of human genetics : EJHG 24, 78-85.

28. Hamanaka, K., Goto, K., Arai, M., Nagao, K., Obuse, C., Noguchi, S., Hayashi, Y.K., Mitsuhashi, S., and Nishino, I. (2016). Clinical, muscle pathological, and genetic features of Japanese facioscapulohumeral muscular dystrophy 2 (FSHD2) patients with SMCHD1 mutations. Neuromuscular disorders : NMD 26, 300-308.

29. Ricci, E., Galluzzi, G., Deidda, G., Cacurri, S., Colantoni, L., Merico, B., Piazzo, N., Servidei, S., Vigneti, E., Pasceri, V. (1999). Progress in the molecular diagnosis of facioscapulohumeral dystrophy and correlation between the number of KpnI repeat at the 4q35 locus and clinical phenotype. Ann Neurol 45, 751-757.

30. Statland, J.M., Donlin-Smith, C.M., Tapscott, S.J., Lemmers, R.J., van der Maarel, S.M., and Tawil, R. (2015). Milder phenotype in facioscapulohumeral dystrophy with 7-10 residual D4Z4 repeats. Neurology 85, 2147-2150.

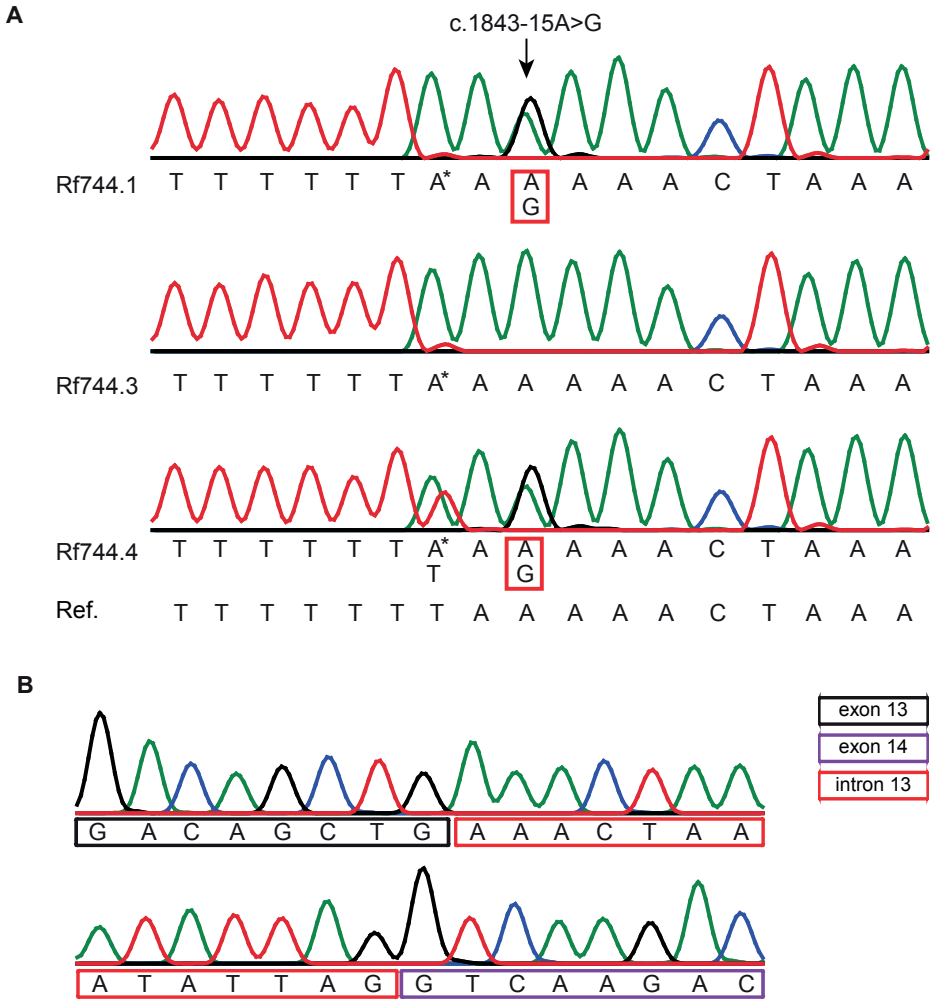
31. Lemmers, R.J., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.M., de Knijff, P., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2007). Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. American journal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

32. Scionti, I., Fabbri, G., Fiorillo, C., Ricci, G., Greco, F., D'Amico, R., Termanini, A., Vercelli, L., Tomelleri, G., Cao, M., et al. (2012). Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. Journal of medical genetics 49, 171-178.

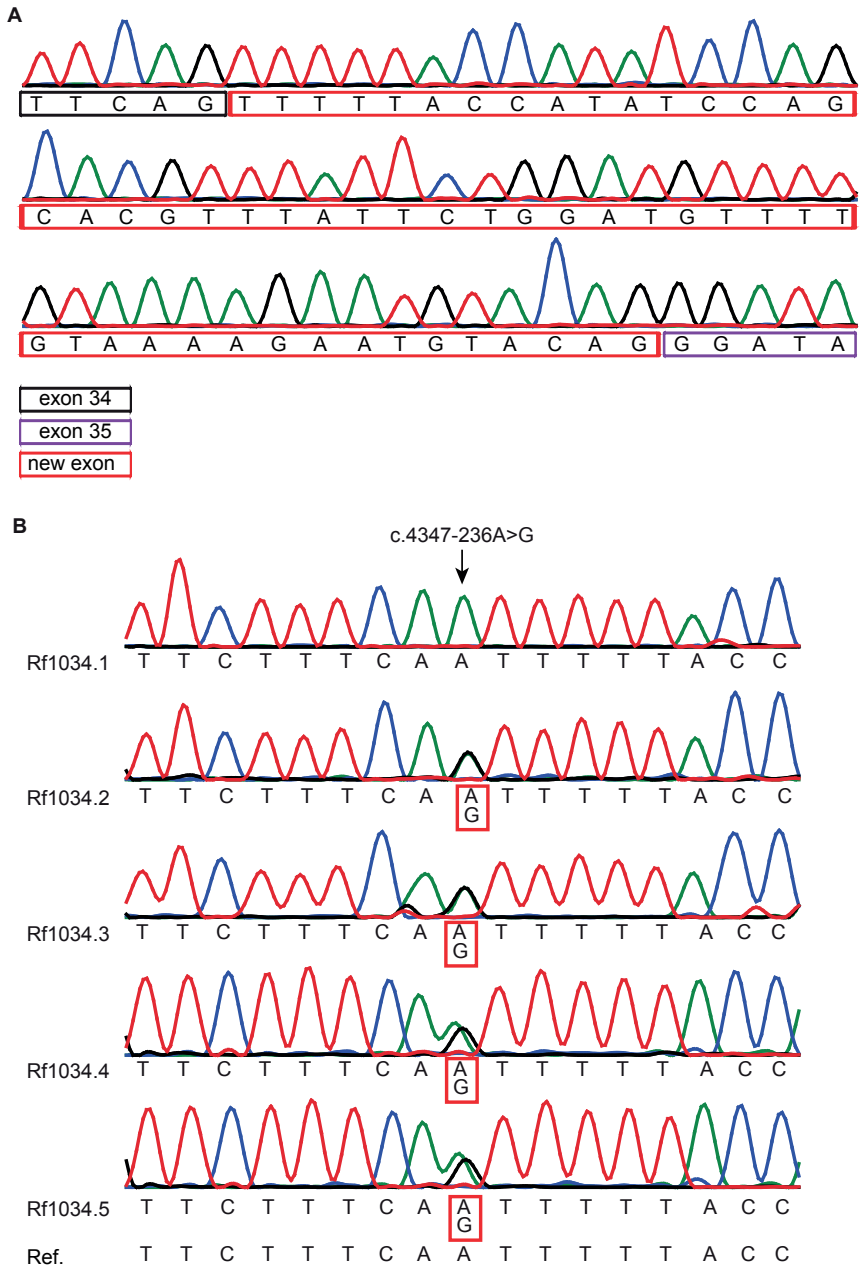
33. Mayes, M.B., Morgan, T., Winston, J., Buxton, D.S., Kamat, M.A., Smith, D., Williams, M., Martin, R.L., Kleinjan, D.A., Cooper, D.N., et al. (2015). Remotely acting SMCHD1 gene regulatory elements: in silico prediction and identification of potential regulatory variants in patients with FSHD. Hum Genomics 9, 25.

SUPPLEMENTARY INFORMATION

Supplementary figures



SUPPLEMENTARY FIGURE 1. Intronic variant in *SMCHD1* in Rf744. A) Sanger sequence track from Rf744.1, Rf744.3 and Rf744.4 showing the intronic variant in *SMCHD1* at position c.1843-15 in Rf744.1 and Rf744.4, highlighted with a red rectangle. * indicates common SNP rs8090988 (T/A, ancestral T, minor allele frequency 0.33 (A)). B) Sanger sequence track of the altered *SMCHD1* transcript in Rf744.1 shows the inclusion of the last 14 nucleotides of intron 13 between exon 13 and exon 14.



SUPPLEMENTARY FIGURE 2. Deep intronic variant in *SMCHD1* in Rf1034. A) Sanger sequence track of the altered *SMCHD1* transcript in Rf1034 shows exonisation of 53 nucleotides of intron 34 between exon 34 and exon 35. B) Sanger sequence track from family members of Rf1034, showing the deep intronic variant in *SMCHD1* at position c.4347-236 in Rf1034.2, Rf1034.3, Rf1034.3 and Rf1034.5, highlighted with a red rectangle.

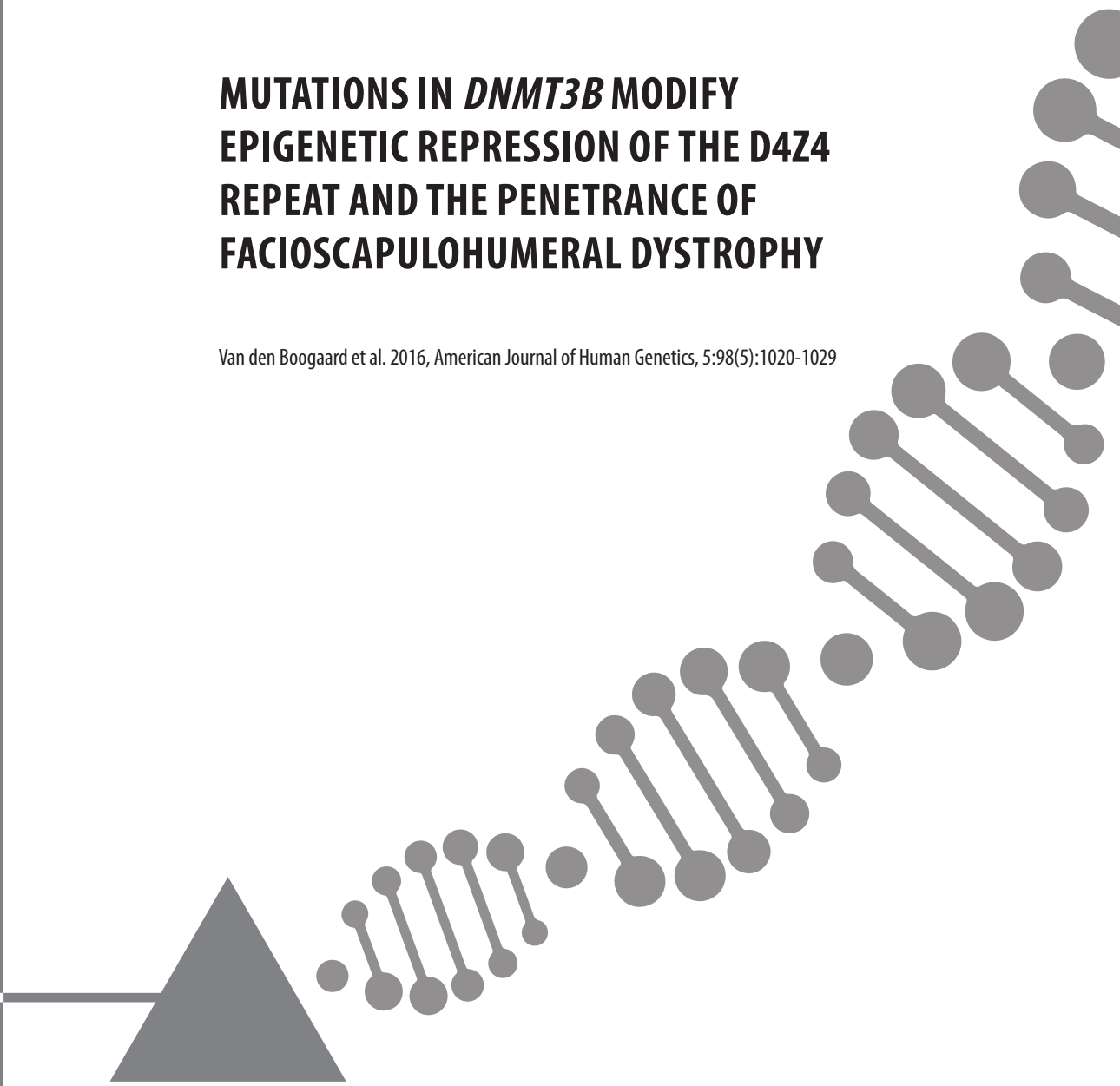
Marlinde L van den Boogaard[1,+], Richard JFL Lemmers[1,+], Judit Balog[1,+], Mariëlle Wohlgemuth[2], Mari Auranen[3], Satomi Mitshuhashi[4], Patrick J van der Vliet[1], Kirsten R Straasheijm[1], Rob FP van den Akker[1], Marjolein Kriek[1,5], Marlies EY Laurense-Bik[5], Vered Raz[1], Monique M van Ostaijen-ten Dam[6], Kerstin BM Hansson[5], Elly L van der Kooi[7], Sari Kiuru-Enari[3], Bjarne Udd[8], Maarten JD van Tol[6], Ichizo Nishino[4], Rabi Tawil[9], Stephen J Tapscott[10], Baziel GM van Engelen[2], Silvère van der Maarel[1]

1 Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands
2 Department of Neurology, Radboud University Medical Center, Nijmegen, The Netherlands
3 Clinical Neurosciences, Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
4 Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan
5 Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
6 Department of Pediatrics, Leiden University Medical Center, Leiden, The Netherlands
7 Medisch Centrum Leeuwarden, Leeuwarden, The Netherlands
8 Neuromuscular Research Center, Department of Neurology, Tampere University Hospital and University of Tampere, Tampere, Finland
9 Department of Neurology, University of Rochester Medical Center, Rochester, NY, USA
10 Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA, USA
+ These authors contributed equally to this work

5

MUTATIONS IN *DNMT3B* MODIFY EPIGENETIC REPRESSION OF THE D4Z4 REPEAT AND THE PENETRANCE OF FACIOSCAPULOHUMERAL DYSTROPHY

Van den Boogaard et al. 2016, American Journal of Human Genetics, 5:98(5):1020-1029



ABSTRACT

Facioscapulohumeral dystrophy (FSHD) is associated with somatic chromatin relaxation of the D4Z4 repeat array and derepression of the D4Z4 encoded *DUX4* retrogene coding for a germline transcription factor. Somatic *DUX4* derepression is either caused by repeat array contraction to a size of 1-10 units (FSHD1), or by mutations in *SMCHD1*, which encodes a chromatin repressor that binds to D4Z4 (FSHD2). We here show that heterozygous mutations in the DNA methyltransferase 3B (*DNMT3B*) gene are a likely cause of D4Z4 derepression associated with low levels of *DUX4* expression from the D4Z4 repeat and increased penetrance of FSHD. Recessive mutations in *DNMT3B* were previously shown to cause Immunodeficiency, Centromeric instability, and Facial anomalies (ICF) syndrome. This study suggests that transcription of *DUX4* in somatic cells is modified by variations in its epigenetic state and provides a basis for understanding the reduced penetrance of FSHD within families.

Facioscapulohumeral dystrophy (FSHD; OMIM 158900 and 158901) is a common muscular dystrophy typically presenting in the second decade and characterized by progressive weakness and atrophy of the facial and upper extremity muscles. With disease progression, also other muscles become affected[1]. A clinical hallmark of the disease is the variability in onset and progression with 20% of mutation carriers eventually becoming wheelchair dependent, and a similar proportion of mutation carriers remaining asymptomatic[2].

The common form of the disease, FSHD1, is associated with a contraction of the polymorphic D4Z4 macrosatellite repeat array on chromosome 4q to a size of 1-10 units (Fig. 1A)[3,4]. In the healthy control population, this array varies between 8-100 units with 1-3% of individuals carrying an FSHD-sized allele of 8-10 units[5,6]. Each unit of the repeat array contains a copy of the double homeobox 4 (*DUX4*) retrogene, which is normally expressed in testis, and silenced in somatic tissue[7]. In FSHD1 the epigenetic repression of *DUX4* is incomplete in somatic cells, leading to sporadic *DUX4* expression in myonuclei[7,8]. Stable *DUX4* transcripts are only produced in combination with a polymorphic polyadenylation signal (PAS) immediately distal to the D4Z4 repeat array present on 4qA chromosomes, of which two major variants exist (4qA-S and 4qA-L) (Fig. 1A)[9]. Contractions of the highly homologous repeat arrays on chromosomes 4qB or 10 are non-pathogenic due to the absence of a *DUX4*-PAS[9].

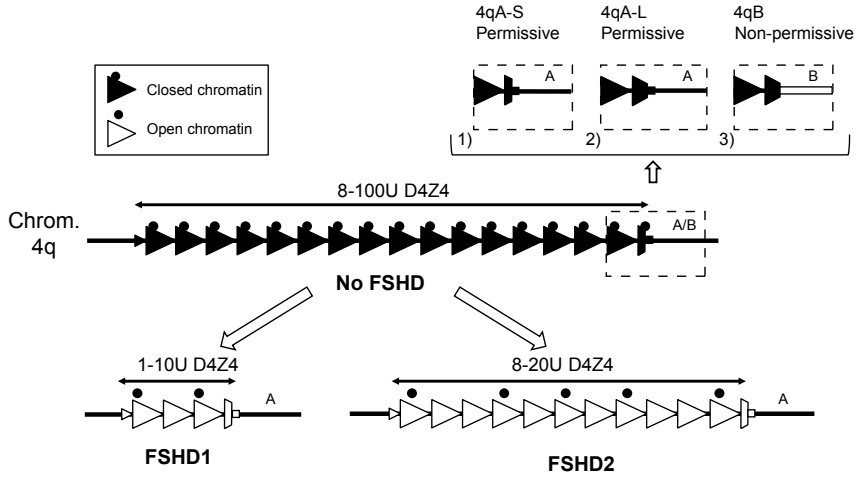
Somatic repression of *DUX4* requires a combination of epigenetic mechanisms with D4Z4 hypomethylation consistently being reported as an aberrant epigenetic feature in FSHD[10-13]. In FSHD1 D4Z4 hypomethylation is restricted to the contracted allele. In the rare FSHD2 type of the disease, D4Z4 hypomethylation is observed on all D4Z4 repeat arrays in the absence of D4Z4 repeat array contractions (Fig. 1A)[14,15]. D4Z4 methylation linearly correlates with the size of the D4Z4 array in controls and FSHD[16]. FSHD2 individuals often carry smaller but normal-sized D4Z4 repeat arrays (8-20 units) given that this renders them more susceptible to further D4Z4 hypomethylation[14,16]. Dominant segregation of D4Z4 hypomethylation in FSHD2 families was instrumental in identifying mutations in the structural maintenance of chromosomes flexible hinge domain-containing protein 1 gene (*SMCHD1*; OMIM 614982) in >85% of FSHD2 families[17]. SMCHD1 is a chromatin repressor involved in the establishment and/or maintenance of CpG methylation at specific loci and binds directly to D4Z4[17-19]. Therefore, the disease presentation in FSHD2 depends on a combination of repeat length and damaging potential of the *SMCHD1* mutation[16]. Mutations in *SMCHD1* have also been reported as modifiers of disease severity in FSHD1 families with FSHD1 alleles of 8-10 D4Z4 units[20,21]. Thus, D4Z4 methylation is dependent on repeat array size and on the activity of the partially characterized D4Z4-repressive mechanisms. Deviations of the expected

D4Z4 methylation, expressed as the Delta1 factor, can be diagnostic for the presence of damaging variants in D4Z4-chromatin modifiers. Indeed, Delta1 factors of ≤ -22% are generally associated with mutations in *SMCHD1*[16].

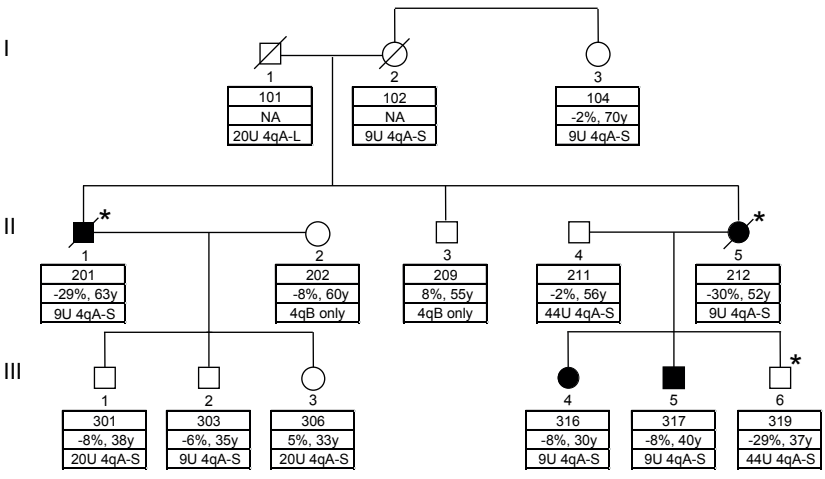
Since not all FSHD2 families can be explained by *SMCHD1* mutations, we applied exome sequencing in eight families in which we found D4Z4 hypomethylation without evidence for an exonic *SMCHD1* mutation (Fig. 1B, 1C and Fig. S1). All samples were obtained in an anonymized manner, and all families gave consent. The study was approved by the Medical Ethical Committees of the Leiden University Medical Center and of the Radboud University Medical Center Nijmegen. Whole exome sequencing (WES) was performed by deCODE Genetics (Reykjavik – Iceland) in the context of the EU Neuromics project. To identify variants the WES data were analyzed using deCODE Clinical Sequence Miner. Dominant analysis for multiple cases and controls and annotation of gene variants (with variant effect predictor consequences moderate to high) were used to identify possible dominant mutations. Under these conditions, in two families we identified a potentially damaging variant in *DNMT3B* (DNA methyltransferase 3B; OMIM 602900), encoding a known D4Z4-chromatin modifier. These variants have not been reported previously in dbSNP, the 1000 Genomes Project, the ESP Exome Variant Server, Exome Aggregation Consortium (ExAC), or in-house databases.

FIGURE 1. D4Z4 locus and FSHD2 families. **(a)** Schematic representation of the D4Z4 locus. In controls the D4Z4 repeat array ranges between 8-100 units and shows characteristics of a closed chromatin structure (black triangles) characterized by amongst others high CpG methylation. For both FSHD1 and FSHD2 the chromatin adopts a more open configuration (white triangles) marked by a loss of CpG methylation and other chromatin changes. FSHD1 is caused by a contraction of the D4Z4 repeat to 1-10 units while in FSHD2 there is chromatin relaxation due to mutations that affect a chromatin modifier (black dots), most often being SMCHD1. The chromatin relaxation must occur on a permissive 4qA (marked by 4qA-S in this figure) or 4qA-L chromosome to cause FSHD, 4qB chromosomes are non-permissive for FSHD (chromosome 4 variants are displayed in the dashed boxes)[9]. 4qA-S and 4qA-L differ by the length of the last partial D4Z4 unit, and protein studies have demonstrated the production of DUX4 protein from both 4qA variants. The 3' UTR region of *DUX4* is missing in 4qB chromosomal regions (white square in dashed box), which makes them non-permissive to DUX4 expression. (b,c) Pedigrees of families Rf210 **(b)** and Rf732 **(c)**. Clinically affected individuals are indicated in black. Key shows the family identifier (ID), the Delta1 score, age at examination (AAE), and the size of the smallest D4Z4 repeat array on a FSHD permissive allele (4qA-S and 4qA-L). Additionally, it is indicated when no permissive allele is present (4qB only). The cDNA position behind the family ID indicates the cDNA position of the *DNMT3B* mutation (NM_006892.3) present in this family. The asterisk indicates individuals carrying the *DNMT3B* mutation.

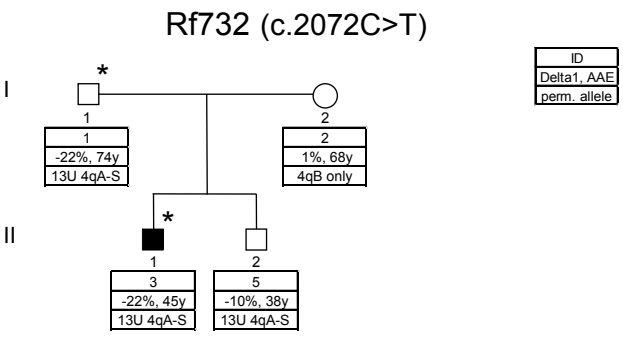
A


4qA-S
Permissive
4qA-L
Permissive
4qB
Non-permissive
Closed chromatin
Open chromatin
A
A
B
1)
2)
3)
8-100U D4Z4
Chrom.
4q
A/B
No FSHD
1-10U D4Z4
8-20U D4Z4
A
A
FSHD1
FSHD2

B

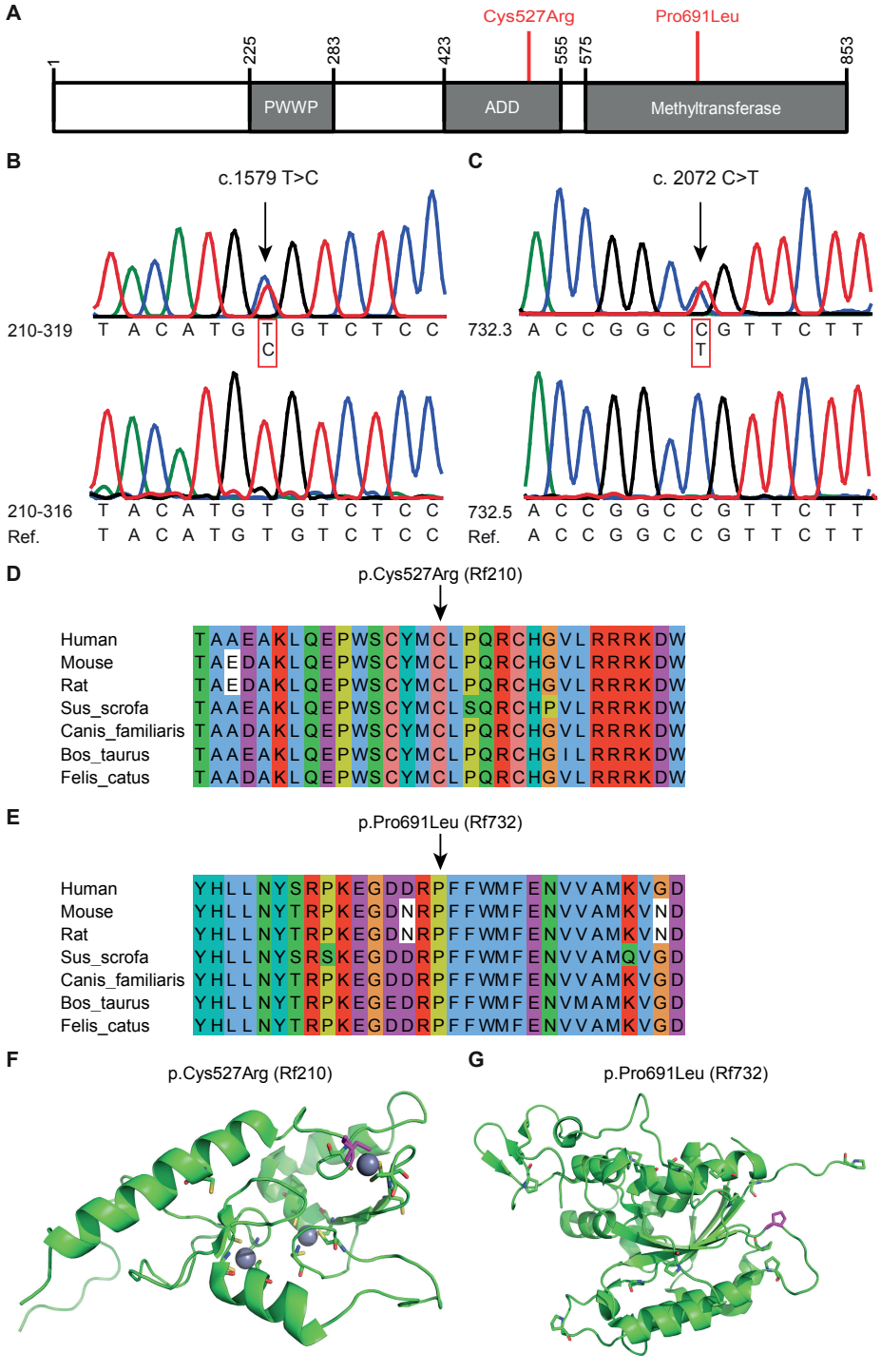

Rf210 (c.1579T>C)
I
1
101
NA
20U 4qA-L
2
102
NA
9U 4qA-S
3
104
-2%, 70y
9U 4qA-S
II
1
201
-29%, 63y
9U 4qA-S
2
202
-8%, 60y
4qB only
3
209
8%, 55y
4qB only
4
211
-2%, 56y
44U 4qA-S
5
212
-30%, 52y
9U 4qA-S
III
1
301
-8%, 38y
20U 4qA-S
2
303
-6%, 35y
9U 4qA-S
3
306
5%, 33y
20U 4qA-S
4
316
-8%, 30y
9U 4qA-S
5
317
-8%, 40y
9U 4qA-S
6
319
-29%, 37y
44U 4qA-S

C


Rf732 (c.2072C>T)
ID
Delta1, AAE
perm. allele
I
1
1
-22%, 74y
13U 4qA-S
2
2
1%, 68y
4qB only
II
1
3
-22%, 45y
13U 4qA-S
2
5
-10%, 38y
13U 4qA-S

Family Rf210 is a FSHD1 family with an array of 9 D4Z4 units on a permissive 4qA chromosome (Fig. 1B, Table S1). Despite the presence of this disease allele in 7 family members, only four of them are clinically affected, while one carrier (Rf210.102 (I-2)) could not be clinically examined. D4Z4 methylation at the FseI site was determined by Southern blotting and expressed as the Delta1 score, which is the observed methylation at the FseI site in D4Z4 corrected for the repeat array size[16]. In Rf210, D4Z4 methylation analysis identified robust D4Z4 hypomethylation in two severely affected individuals (Rf210.201 (II-1) and Rf210.212 (II-5)) and one clinically unaffected individual (Rf210.319 (III-6)) as evidenced by the strongly reduced Delta1 values. This reduced Delta1 value is indicative for the involvement of a defective D4Z4-chromatin modifier. Genetic studies excluded the involvement of the *SMCHD1* locus (Fig. S2) but exome sequencing identified a potentially damaging variant in *DNMT3B* co-segregating with D4Z4 hypomethylation (Fig. 2A, 2B and Table S1). This variant (NM_006892.3(DNMT3B):c.1579T>C, p.Cys527Arg) was confirmed by Sanger sequencing and disrupts the C2C2-type zinc finger motif in the ATRX-DNMT3-DNMT3L (ADD) domain, a highly conserved domain which can be found in several chromatin-associated proteins that play a role in establishing and/or maintaining a normal DNA methylation pattern (Fig. 2B, 2D and 2F)[22,23]. Like SMCHD1, DNMT3B was previously identified as suppressor of metastable epialleles in mice, alleles that display unusual variable expressivity in the absence of genetic heterogeneity but depending on their epigenetic state[18,24,25]. In these *Dnmt3b*-hypomorphic mice also the ADD domain seems to be primarily affected[26].

FIGURE 2. *DNMT3B* mutations in FSHD2. **(a)** Schematic representation of DNMT3B with the amino acid changes (NP_008823.1) found in FSHD2 families indicated in red. **(b,c)** Sanger sequence confirmation of *DNMT3B* variants (NM_006892.3) in Rf210 and Rf732, respectively. **(d,e)** Multiple sequence alignment (MSA) of DNMT3B protein across distinct species for DNMT3B variants in Rf210 and Rf732 respectively. MSA was performed with ClustalOmega, alignment was viewed in Jalview and coloured as ClustalX. **(f)** Ribbon representation of the NMR structure of the ADD domain of ATRX (PDB entry 2JM1)[22]. The cysteine residues are shown in sticks. Cys527 is shown in magenta. Zinc ions are represented as spheres. **(g)** Ribbon representation of the crystallography structure of the C terminal domain of DNMT3A (Chain A, PDB entry 2QRV). The proline residues are shown in sticks. Pro691 is shown in magenta.


A
Cys527Arg
Pro691Leu
1
225
283
423
555
575
853
PWWP
ADD
Methyltransferase
B
c.1579 T>C
210-319
T A C A T G T G T C T C C
C
210-316
T A C A T G T G T C T C C
Ref.
T A C A T G T G T C T C C
C
c. 2072 C>T
732.3
A C C G G C C G T T C T T
T
732.5
A C C G G C C G T T C T T
Ref.
A C C G G C C G T T C T T
D
p.Cys527Arg (Rf210)
Human T A A E A K L Q E P W S C Y M C L P Q R C H G V L R R R K D W
Mouse T A E D A K L Q E P W S C Y M C L P Q R C H G V L R R R K D W
Rat T A E D A K L Q E P W S C Y M C L P Q R C H G V L R R R K D W
Sus_scrofa T A A E A K L Q E P W S C Y M C L S Q R C H P V L R R R K D W
Canis_familiaris T A A D A K L Q E P W S C Y M C L P Q R C H G V L R R R K D W
Bos_taurus T A A E A K L Q E P W S C Y M C L P Q R C H G I L R R R K D W
Felis_catus T A A D A K L Q E P W S C Y M C L P Q R C H G V L R R R K D W
E
p.Pro691Leu (Rf732)
Human Y H L L N Y S R P K E G D D R P F F W M F E N V V A M K V G D
Mouse Y H L L N Y T R P K E G D N R P F F W M F E N V V A M K V N D
Rat Y H L L N Y T R P K E G D N R P F F W M F E N V V A M K V N D
Sus_scrofa Y H L L N Y S R S K E G D D R P F F W M F E N V V A M Q V G D
Canis_familiaris Y H L L N Y T R P K E G D D R P F F W M F E N V V A M K V G D
Bos_taurus Y H L L N Y T R P K E G E D R P F F W M F E N V M A M K V G D
Felis_catus Y H L L N Y T R P K E G D D R P F F W M F E N V V A M K V G D
F
p.Cys527Arg (Rf210)
G
p.Pro691Leu (Rf732)

In family Rf210 the *DNMT3B* variant perfectly segregates with D4Z4 hypomethylation, but not with disease presentation. *DNMT3B* mutation carrier Rf210.319 (III-6, Fig. 1B) may be protected from disease presentation because of the large size of the FSHD permissive D4Z4 repeat (44 units). This is reminiscent to the situation in *SMCHD1* mutation carriers, where individuals with smaller normal-sized D4Z4 repeat arrays (11-20 units) have a greater likelihood of developing FSHD than individuals with larger repeat arrays[16]. The two *DNMT3B* variant carriers with an array of 9 D4Z4 units, however, have an age corrected clinical severity score (ACCS) that is greater than the carriers of only a 9 D4Z4 units allele. This suggests that the *DNMT3B* variant acts as a modifier of disease severity in this FSHD1 family, similar to *SMCHD1* mutation carriers in FSHD1 families[20]. Of the four carriers of an array of 9 D4Z4 units without *DNMT3B* variant, two are clinically unaffected (Rf210.104 (I-3) and Rf210.303 (III-2)). This variability in severity is typical for this borderline FSHD1 repeat array size. Indeed, 1-3% of the control population carries an array of 8-10 units on a permissive allele, demonstrating the strongly reduced penetrance for these alleles[5,6]. Penetrance is, amongst others, dependent on age and the degree of D4Z4-chromatin relaxation in somatic tissue[12,16,27].

In family Rf732, the index case Rf732.3 (II-1) carries a D4Z4 repeat array of 13 units on a 4qA chromosome (Fig. 1C, Table S1), which is also present in his unaffected father and brother. Methylation analysis showed that the index case and his father (Rf732.3 (II-1) and Rf732.1 (I-1)) had severe D4Z4 hypomethylation on all four alleles with reduced Delta1 values. Exome sequencing identified a potentially damaging variant affecting a highly conserved residue in the enzymatic domain of DNMT3B (NM_006892.3(DNMT3B):c.2072C>T, p.Pro691Leu) in the index case and his father, which was confirmed by Sanger sequencing and absent in the son with normal D4Z4 methylation (Fig. 2A, 2C, 2E and 2G). Although Rf732.1 (I-1) and Rf732.3 (II-1) both carry this *DNMT3B* variant, have the same Delta1 value, and a 13 units FSHD-permissive D4Z4 allele, only Rf732.3 (II-1) is clinically affected. This family emphasizes the reduced penetrance that is typical for FSHD[16,27]. The Delta1 value in this family is low, but not as low as typically found in *SMCHD1* mutation carriers[16]. This suggests a lesser degree of D4Z4-chromatin relaxation in this family, which might explain why the father has remained unaffected.

Analysis of all coding exons of *DNMT3B* in 25 additional cases with a permissive D4Z4 allele and borderline to severely reduced D4Z4 methylation, but excluded for exonic *SMCHD1* mutations, did not identify additional mutations in *DNMT3B* (Table S2 and Table S4).

Biallelic *DNMT3B* mutations have been reported in autosomal recessive Immunodeficiency, Centromeric instability, and Facial anomalies syndrome type 1 (ICF1; OMIM 242860)[28,29]. This primary immunodeficiency syndrome is characterized

by hypo- or agammaglobulinemia with B cells, and by distinct facial appearance. There is a progressive decrease of B- and T-cells during childhood and adolescence[30,31]. The cytogenetic hallmark of ICF syndrome is the presence of chromosome abnormalities involving the juxtacentromeric domains of chromosomes 1, 9 and 16 in metaphase spreads of phytohemagglutinin (PHA) stimulated cells[30,32]. ICF1 patients show CpG hypomethylation of juxtacentromeric satellite repeats type II and III, and the macrosatellite repeats NBL2 and D4Z4[33,34]. ICF1 mutations most often affect the catalytic domain of DNMT3B and are believed to result in strongly reduced DNMT3B activity[31].

Since our data suggest that FSHD2 and ICF1 can both be caused by *DNMT3B* mutations, with dominant mutations in *DNMT3B* associated with FSHD2 and recessive mutations causing ICF syndrome, we analyzed six ICF1 patients belonging to five families (Rf285, Rf286, Rf614, Rf699, and Coriell Cell Repositories family 2081, here annotated as Rf1178) for D4Z4 repeat array sizes, presence of a *DUX4*-PAS, D4Z4 hypomethylation, and *DUX4* expression (Fig. 3). If possible, we also included unaffected relatives. Table S3 lists all ICF1 families with reference to their original description. Consistent with earlier reports[33], methylation analysis showed that all ICF1 patients tested had severe D4Z4 hypomethylation with Delta1 values varying between -35% and -46% (Fig. 3). However, depending on the mutation, some heterozygous carriers (parents of Rf699 and mother of Rf1178) also showed reduced Delta1 values similar to what we observed in our FSHD2 families (-19% to -26%). This not only suggests an additive effect of both *DNMT3B* mutations in the affected ICF1 children, but also puts ICF1-mutation carriers with a reduced Delta1 value at risk of stable *DUX4* expression and FSHD, if combined with a *DUX4*-PAS. Analysis of D4Z4-repeat sizes, however, showed that about half of the heterozygous *DNMT3B* carriers in our ICF1-affected families do not carry a FSHD-permissive chromosome. For those who do have D4Z4 repeat arrays on FSHD-permissive chromosomes (containing a *DUX4* PAS), the arrays are well beyond the size of what is typically found in FSHD2 individuals (Fig. 3). The smallest permissive D4Z4 repeat array found in these heterozygous *DNMT3B* carriers contained 32 units, suggesting that these individuals may be protected from somatic *DUX4* expression because of their long D4Z4 repeat arrays, since in FSHD2, we already demonstrated a D4Z4 repeat size-dependent penetrance for *SMCHD1* mutations[16]. In concordance, to our knowledge, muscle weakness has never been reported in ICF1 mutation carriers.

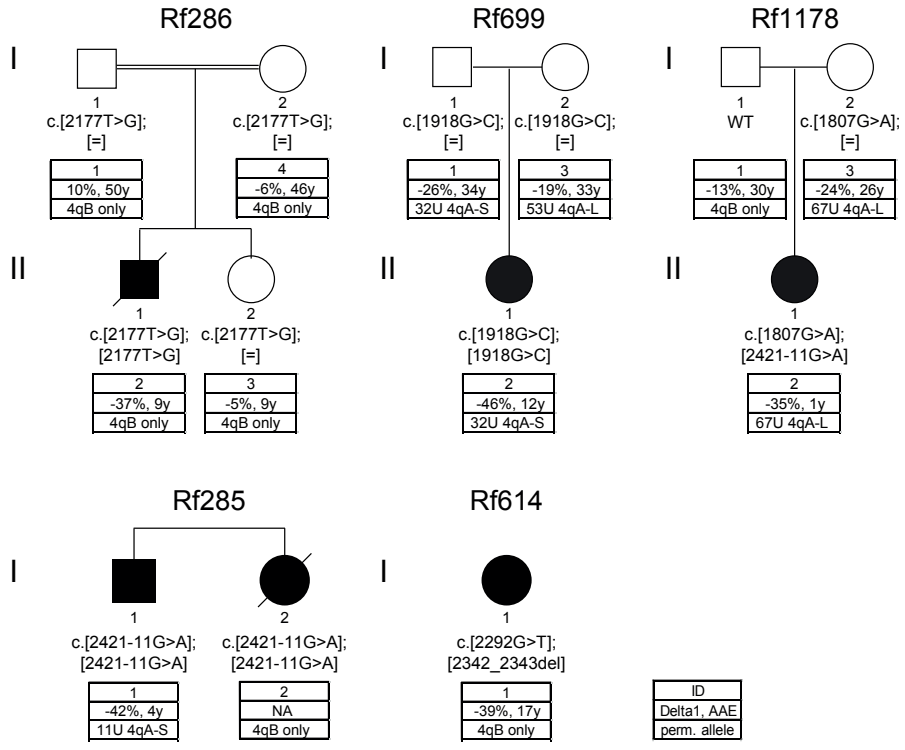


FIGURE 3. Pedigrees of autosomal recessive ICF1 families Rf286, Rf699, Rf1178, Rf285, and Rf614. Affected individuals are indicated in black and *DNMT3B* mutations (NM_006892.3) are shown below each individual. Their clinical phenotypes and *DNMT3B* mutations have been described before[28,30,31,43,44]. Key description identical to Fig. 1.

To address the possibility of *DUX4* expression in carriers of a single *DNMT3B* mutation, we trans-differentiated primary fibroblasts of controls, FSHD1 and FSHD2 patients and of non-affected and affected carriers of an FSHD2 mutation in *DNMT3B* (Rf210.319 (III-6), Fig. 1B and Rf732.3 (II-1), Fig. 1C) into myotubes by lentiviral MyoD expression. A lentivirus containing GFP or FLAG was used as a control. *MYOG* and *MYH3* expression levels were measured by Q-PCR to examine differentiation[35,36]. For almost all cell lines we observed *MYOG* and *MYH3* expression only in the fibroblasts transduced with MyoD, indicating that these cells were trans-differentiated into myogenic cells (Fig. 4A). In one FSHD2 cell line (FSHD2 (1)) *MYH3* expression was detected in the GFP transduced fibroblast as well, possibly because of a technical or biological artifact. We next analyzed the expression of *DUX4* and three DUX4 target genes (*LEUTX*, *TRIM43* and *PRAMEF2*) by Q-PCR and gel electrophoresis[37]. We found *DUX4* and DUX4 target gene expression in MyoD transduced fibroblasts of affected FSHD2 individual Rf732.3 (II-1)

with a D4Z4 repeat array of 13 units, but not in unaffected individual Rf210.319 (III-6) with a 44 units array on a 4qA chromosome (Fig. 4A and Fig. S3A). No *DUX4* or upregulation of DUX4 target gene expression was detected in GFP transduced fibroblasts and no fibroblasts were available from other FSHD2 family members. These data are consistent with the suggestion that heterozygous *DNMT3B* mutations, only when combined with smaller D4Z4 repeat arrays, can de-repress DUX4 in somatic cells and cause FSHD.

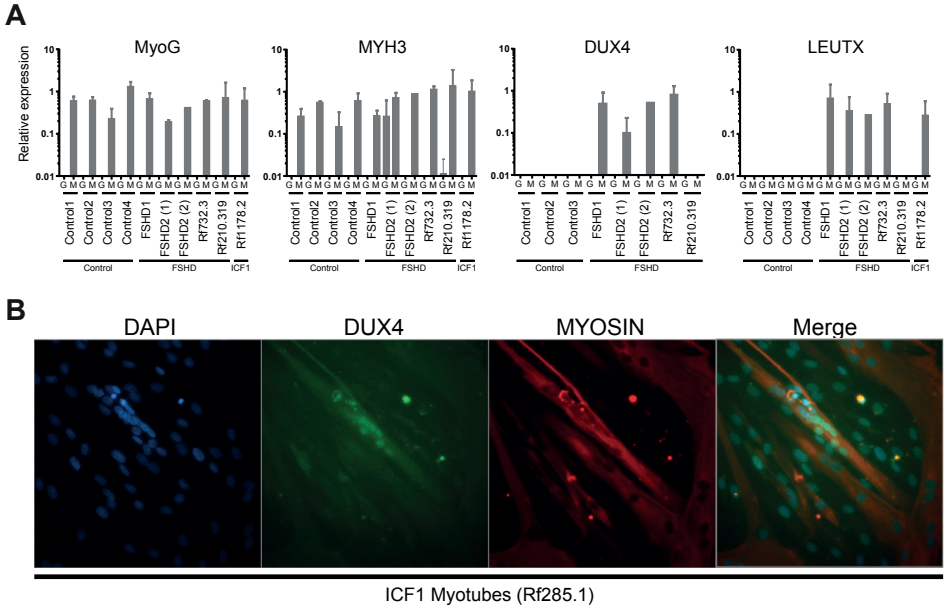


FIGURE 4. DUX4 presence in FSHD and ICF1. **(a)** Expression of *MyoG*, *MYH3*, *DUX4*, and *LEUTX* (DUX4 target) by Q-PCR in GFP (G)- or MyoD (M)-lentivirus-transduced fibroblasts from controls, FSHD1, FSHD2, Rf210.319, Rf732.3 and ICF1 (Rf1178.2) individuals. All transductions were performed twice for each cell line, except for control 4 (1x transduced with GFP, 2x transduced with MyoD) and FSHD2 (2) (transduced 1x with GFP and 1x with MyoD). Mean expression values with standard deviations are shown relative to the reference genes *GUSB* and *RPL27*. *DUX4* is measured with primers for the most common *DUX4*-4A-S variant, but the primers do not recognize *DUX4*-4A-L. The fibroblasts from control individual 4 and Rf1178.2 carry a 4qA-L allele, and are therefore excluded from analysis of *DUX4* expression. Primers are listed in table S5. **(b)** Immunofluorescent staining for DUX4 and Myosin in fixed ICF1 myotubes from Rf285.1 (Fig. 3) shows DUX4 immunoreactivity in a small percentage of myotubes.

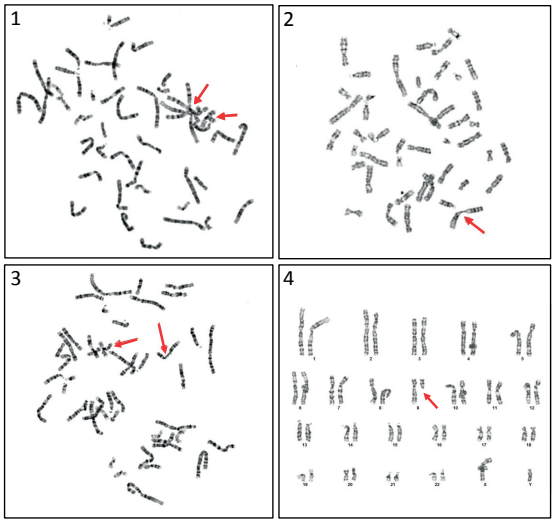
To investigate *DUX4* expression in ICF1 we trans-differentiated three primary fibroblast cell lines of ICF1 patients (Rf699.2 (II-1), Rf614.1 (I-1) and Rf1178.2 (II-1); Fig. 3). In ICF1 patient Rf699.2 (II-1) with a permissive D4Z4 array of 32 units we detected DUX4 in the MyoD transduced fibroblasts (Fig. S3B). *DUX4* could not be detected in ICF1 patient Rf614.1 (I-1), because she carries two 4qB alleles, which are unable to produce a stable DUX4 transcript (Fig. S3B). Our *DUX4* primers recognize the most common *DUX4*-4A-S variant, but not the *DUX4*-4A-L variant, which is produced from 4qA-L repeats. Since ICF1 patient Rf1178.2 (II-1) carries a 4qA-L repeat, we were unable to directly detect DUX4 in this patient (Fig. S3B). However, the expression of DUX4 target genes was detected in Rf1178.2 (II-1), suggesting that there is DUX4 produced in these fibroblasts (Fig. 4A and Fig. S3A). These results show that ICF1 patients can express small amounts of DUX4 in MyoD-transduced fibroblasts indicating that when both *DNMT3B* alleles are mutated, the epigenetic derepression is sufficient to facilitate *DUX4* expression from D4Z4 repeats (Fig. S3B). Additionally, we had myotubes available of one ICF1 patient from a different family (Rf285.1 (I-1), Fig. 3) with a D4Z4 repeat of 11 units on a FSHD permissive chromosome 4, where we detected small amounts of DUX4 by immunofluorescent staining (Fig. 4B). This ICF1 patient (Rf285.1 (I-1)) might still be too young (15 years) to develop FSHD. Possibly, the short life expectancy of ICF1 patients in general may obscure the diagnosis of muscle weakness.

Conversely, although ICF1 mutation carriers are reported to be unaffected, we explored the possibility that dominant *DNMT3B* mutations identified in our FSHD2 families may have epigenetic consequences similar to what is found in ICF1, or clinical features reminiscent of ICF syndrome. Metaphase analysis of PHA stimulated peripheral mononuclear blood cell cultures of FSHD *DNMT3B* mutation carrier Rf210.319 (III-6, Fig. 1B), but not Rf732.3 (II-1, Fig. 1C), indicated a low frequency of formation of multi-branched chromosomes (Fig. 5A and 5B). Chromosome decondensations, breaks and deletions, can be found at low frequency also in ICF1 mutation carriers and controls[32]. But the formation of multi-branched chromosomes may be specific to the presence of *DNMT3B* mutations, even in heterozygous carriers. Rf210.319 (III-6) also showed evidence for mild NBL2 hypomethylation in a Southern blot assay, as the NBL2 repeat is sensitive to digestion by the methylation-sensitive endonuclease Eco52I, albeit to a lesser degree than observed in ICF1 patients (Fig. 5C). Similarly, one heterozygous ICF1 mutation carrier with strongly reduced Delta1 values for D4Z4 (Rf699.1 (I-1), Fig. 3), also showed mild NBL2 hypomethylation (Fig. 5C). Since not all carriers of the same variant showed NBL2 hypomethylation, this suggests that heterozygous *DNMT3B* variants can cause mild and variable NBL2 hypomethylation. Clinically, however, *DNMT3B* mutation carrier Rf210.319 (III-6) and his siblings Rf210.316 (III-4) and Rf210.317 (III-5), do not show signs or features of ICF syndrome and have normal serum immunoglobulin levels and normal numbers of B-cells and T-cell subsets (Fig. S4).

A

				DNMT3B mutation			
Family	Nr	Coriell ID	Gender	NM_006892.3	Metaphases	Anomalies	Details anomalies
Rf210	316	-	F	WT	100 (2013)	0	
					100 (2015)	0	
Rf210	317	-	M	WT	100	0	
Rf210	319	-	M	c.[1579T>C];[=]	100 (2013)	3	One cell with a multiradial of chromosome 1 (p,p,q,q,q) and a triradial of chromosome 16 (p,q,q) (see B-1) Two cells with decondensation (stretching) in the pericentromeric region of chromosome 1 (see B-2)
					100 (2015)	4	One cell with a triradial of chromosome 16 (p,q,q) One cell with a triradial of chromosome 16 (p,q,q) and decondensation in the pericentromeric region of chromosome 1 (see B-3) One cell with a deletion of 9q (see B-4) One cell with a deletion of 16q
Rf732	3	-	M	c.[2072C>T];[=]	100	0	
Rf1178	1	GM08729	M	WT	95	0	See ref. 32 and 35
Rf1178	2	GM08714	M	c.[1807G>A];[2421-11G>A]	28	2	One cell with a deletion of 1q and one cell with an extra 1q. See also ref. 32 and 43
Rf1178	3	GM08728	F	c.[1807G>A];[=]	61	0	See ref. 32 and 35

B



C

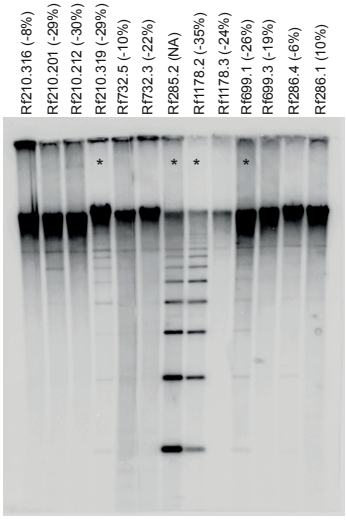


FIGURE 5. Metaphase analysis and NBL2 Southern blot analysis of Rf210, Rf732 and ICF1 families **(a)** Metaphases were analyzed from 3 heterozygous *DNMT3B* mutation carriers (Rf210.319; Rf732.3 and Rf1178.3), one ICF1 patient (Rf1178.2) and three individuals without *DNMT3B* variant (Rf210.316, Rf210.317 and Rf1178.1). Identifiers from LUMC and Coriell are indicated, the mutation in *DNMT3B* (NM_006892.3) and the number of analyzed metaphases. Chromosomal anomalies are listed in the last column. **(b)** Four panels with examples of chromosomal anomalies identified in individual Rf210.319. Chromosomal anomalies are indicated with red arrows. **(c)** *NBL2* Southern blot analysis in Rf210, Rf732 and ICF1 families after digestion of 2 µg genomic DNA with the methylation-sensitive endonuclease Eco52I using previously described protocols[45]. Numbers correspond with pedigrees in Fig. 1 and 3. Delta1 score is indicated between brackets. *NBL2* is only hypomethylated in the four individuals indicated with an asterisk.

These observations raise the question why *DNMT3B* mutations can cause such discordant phenotypes. Mutations that affect the ADD domain of DNMT3B have never been reported in ICF syndrome, but mutations disrupting the ADD domain of DNMT3A (OMIM 602769) have been associated with Tatton-Brown-Rahman syndrome (OMIM 615879), an overgrowth syndrome with intellectual disability[38]. Similarly, mutations that disturb the ADD domain of ATRX (OMIM 300032) have been reported in alpha thalassemia-mental retardation, X-linked (ATR-X; OMIM 301040) syndrome[22]. The ADD domains of ATRX, DNMT3A and DNMT3B bind to the N-terminus of the histone 3 (H3) tail lacking the active lysine 4 (H3K4) methylation mark, where they integrate histone modification status with DNA methylation[39]. Binding of the ADD domain of DNMT3A to the H3 tail stimulates the catalytic activity of this enzyme[40-42]. Likewise, it is possible that the mutation that affects the ADD domain of DNMT3B in family Rf210 also disrupts the DNA methylation activity of DNMT3B. Most of the ICF1 mutations, however, are located in exons that encode for the catalytic domain of DNMT3B, like the mutation in family Rf732. It is not well known why mutations in *DNMT3B* cause a primary immunodeficiency, but the absence of an immunological phenotype in our FSHD2 families may be explained by the presence of one wild type *DNMT3B* allele, as heterozygous ICF1 mutation carriers also do not present immunological abnormalities.

Our study implicates that mutations in *DNMT3B* act as a modifier in FSHD. We propose that, like for *SMCHD1*, the effect of *DNMT3B* mutations on *DUX4* expression and disease presentation depends on the presence of a *DUX4*-PAS and on D4Z4 repeat array size. This, combined with the relative young age at which ICF1 patients typically succumb from their immunodeficiency, may explain the absence of FSHD in ICF1 families. These observations also suggest that FSHD1 and FSHD2 represent polar extremes of a continuous disease mechanism determined by the interaction of D4Z4 repeat size, the presence of a *DUX4*-PAS, and variations in genes that modify the D4Z4 epigenetic state, and provides a firm basis for understanding reduced disease penetrance in the FSHD population.

Supplemental data

Supplemental data include four figures and five tables.

Acknowledgements

We thank all families for participating in our studies. We thank Marcellus Ubbink, Ph.D. (Leiden Institute of Chemistry, Leiden University) for assistance in modeling the mutations and Nisha Verwey, B.Sc. (Human Genetics, LUMC), for assistance with the Cellomics platform. Our studies are supported by grants from the US National Institutes of Health (NIH NINDS P01NS069539), The Prinses Beatrix Spierfonds (W.OR12-20 and

W.OP14-01), the European Union Framework Programme 7 agreement 2012-305121 (NEUROMICS), the FSH Society, Spieren voor Spieren, the FSHD Global Research Foundation, the FSHD Stichting, and Friends of FSH Research.

Web Resources

1000 genomes, http://www.1000genomes.org/
Alamut visual, http://www.interactive-biosoftware.com/alamut-visual/
dbSNP, http://www.ncbi.nlm.nih.gov/SNP/
Exome Aggregation Consortium (ExAC), http://exac.broadinstitute.org/
Leiden Open Variation Database (LOVD), http://www.lovd.nl/3.0/home
Mutalyzer, https://mutalyzer.nl/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
OMIM, http://www.omim.org/
NIGMS Human Genetic Cell Repository, https://catalog.coriell.org/1/NIGMS
RefSeq, http://www.ncbi.nlm.nih.gov/refseq/
Richard Fields Center for FSHD Research, https://www.urmc.rochester.edu/fields-center.aspx

Accession numbers

The mutations have been submitted to the LOVD database, individual IDs 00059205, 00059206, 00059223, 00059224 and 00059225.

REFERENCES

1. Padberg, G.W., Lunt, P.W., Koch, M., and Fardeau, M. (1991). Diagnostic criteria for facioscapulohumeral muscular dystrophy. Neuromuscul. Disord. 1, 231–234.
2. Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2014). Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. Skelet. muscle 4, 12.
3. van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., and Frants, R.R. (1993). 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.
4. Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat. Genet. 2, 26–30.
5. Lemmers, R.J., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.M., de Knijff, P., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2007). Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am. J. Hum. Genet. 81, 884–894.
6. Scionti, I., Fabbri, G., Fiorillo, C., Ricci, G., Greco, F., D'Amico, R., Termanini, A., Vercelli, L., Tomelleri, G., Cao, M., et al. (2012). Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. J. Med. Genet. 49, 171–178.
7. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., et al. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet. 6, e1001181.
8. Tassin, A., Laoudj-Chenivesse, D., Vanderplanck, C., Barro, M., Charron, S., Ansseau, E., Chen, Y.W., Mercier, J., Coppee, F., and Belayew, A. (2013). DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? J. Cell. Mol. Med. 17, 76–89.
9. Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329, 1650-1653.
10. van Overveld, P.G., Lemmers, R.J., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat. Genet. 35, 315-317.
11. Huichalaf, C., Micheloni, S., Ferri, G., Caccia, R., and Gabellini, D. (2014). DNA methylation analysis of the macrosatellite repeat associated with FSHD muscular dystrophy at single nucleotide level. PloS one 9, e115278.

12. Jones, T.I., King, O.D., Himeda, C.L., Homma, S., Chen, J.C., Beermann, M.L., Yan, C., Emerson, C.P., Jr., Miller, J.B., Wagner, K.R., et al. (2015). Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with disease in facioscapulohumeral muscular dystrophy. Clin. Epigenetics 7, 37.

13. Hartweck, L.M., Anderson, L.J., Lemmers, R.J., Dandapat, A., Toso, E.A., Dalton, J.C., Tawil, R., Day, J.W., van der Maarel, S.M., and Kyba, M. (2013). A focal domain of extreme demethylation within D4Z4 in FSHD2. Neurology 80, 392-399.

14. de Greef, J.C., Lemmers, R.J., Camano, P., Day, J.W., Sacconi, S., Dunand, M., van Engelen, B.G., Kiuru-Enari, S., Padberg, G.W., Rosa, A.L., et al. (2010). Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology 75, 1548-1554.

15. Balog, J., Thijssen, P.E., de Greef, J.C., Shah, B., van Engelen, B.G., Yokomori, K., Tapscott, S.J., Tawil, R., and van der Maarel, S.M. (2012). Correlation analysis of clinical parameters with epigenetic modifications in the DUX4 promoter in FSHD. Epigenetics 7, 579–584.

16. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2014). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Hum. Mol. Genet. 24, 659-669.

17. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat. Genet. 44, 1370-1374.

18. Blewitt, M.E., Gendrel, A.V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., et al. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat. Genet. 40, 663-669.

19. Gendrel, A.V., Tang, Y.A., Suzuki, M., Godwin, J., Nesterova, T.B., Greally, J.M., Heard, E., and Brockdorff, N. (2013). Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Mol. Cell. Biol. 33, 3150-3165.

20. Sacconi, S., Lemmers, R.J., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. Am. J. Hum. Genet. 93, 744-751.

21. Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., and Muller, C.R. (2014). Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. Eur. J. Hum. Genet. 23, 808-816.

22. Argentaro, A., Yang, J.C., Chapman, L., Kowalczyk, M.S., Gibbons, R.J., Higgs, D.R., Neuhaus, D., and Rhodes, D. (2007). Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. Proc. Natl. Acad. Sci. USA 104, 11939-11944.

23. Hashimoto, H., Vertino, P.M., and Cheng, X. (2010). Molecular coupling of DNA methylation and histone methylation. Epigenomics 2, 657-669.

24. Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I., and Whitelaw, E. (2002). Metastable epialleles in mammals. Trends Genet. 18, 348-351.

25. Blewitt, M.E., Vickaryous, N.K., Hemley, S.J., Ashe, A., Bruxner, T.J., Preis, J.I., Arkell, R., and Whitelaw, E. (2005). An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. Proc. Natl. Acad. Sci. USA 102, 7629–7634.

26. Youngson, N.A., Epp, T., Roberts, A.R., Daxinger, L., Ashe, A., Huang, E., Lester, K.L., Harten, S.K., Kay, G.F., Cox, T., et al. (2013). No evidence for cumulative effects in a Dnmt3b hypomorph across multiple generations. Mamm. Genome 24, 206-217.

27. Ricci, G., Scionti, I., Sera, F., Govi, M., D'Amico, R., Frambolli, I., Mele, F., Filosto, M., Vercelli, L., Ruggiero, L., et al. (2013). Large scale genotype-phenotype analyses indicate that novel prognostic tools are required for families with facioscapulohumeral muscular dystrophy. Brain 136, 3408-3417.

28. Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M., and Gartler, S.M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc. Natl. Acad. Sci. USA 96, 14412-14417.

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

30. Hagleitner, M.M., Lankester, A., Maraschio, P., Hulten, M., Fryns, J.P., Schuetz, C., Gimelli, G., Davies, E.G., Gennery, A., Belohradsky, B.H., et al. (2008). Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). J. Med. Genet. 45, 93-99.

31. Weemaes, C.M., van Tol, M.J., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C., Thijssen, P.E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E., et al. (2013). Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. Eur. J. Hum. Genet. 21, 1219-1225.

32. Tuck-Muller, C.M., Narayan, A., Tsien, F., Smeets, D.F., Sawyer, J., Fiala, E.S., Sohn, O.S., and Ehrlich, M. (2000). DNA hypomethylation and unusual chromosome instability in cell lines from ICF syndrome patients. Cytogenet. Cell Genet. 89, 121-128.

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

34. Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A., and Viegas-Pequignot, E. (1993). An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. Hum. Mol. Genet. 2, 731-735.

35. Larsen, J., Pettersson, O.J., Jakobsen, M., Thomsen, R., Pedersen, C.B., Hertz, J.M., Gregersen, N., Corydon, T.J., and Jensen, T.G. (2011). Myoblasts generated by lentiviral mediated MyoD transduction of myotonic dystrophy type 1 (DM1) fibroblasts can be used for assays of therapeutic molecules. BMC Res. Notes 4: 490.

36. Racca, A.W., Beck, A.E., McMillin, M.J., Korte, F.S., Bamshad, M.J., and Regnier, M. (2015). The embryonic myosin R672C mutation that underlies Freeman-Sheldon syndrome impairs cross-bridge detachment and cycling in adult skeletal muscle. Hum. Mol. Genet. 24, 3348-3358.

37. Yao, Z., Snider, L., Balog, J., Lemmers, R.J., Van Der Maarel, S.M., Tawil, R., and Tapscott, S.J. (2014). DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. Hum. Mol. Genet. 23, 5342-5352.

38. Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., et al. (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nature genet. 46, 385-388.

39. Noh, K.M., Allis, C.D., and Li, H. (2015). Reading between the Lines: "ADD"-ing Histone and DNA Methylation Marks toward a New Epigenetic "Sum". ACS Chem. Biol. DOI: 10.1021/acschembio.5b00830

40. Li, B.Z., Huang, Z., Cui, Q.Y., Song, X.H., Du, L., Jeltsch, A., Chen, P., Li, G., Li, E., and Xu, G.L. (2011). Histone tails regulate DNA methylation by allosterically activating de novo methyltransferase. Cell Res. 21, 1172-1181.

41. Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., et al. (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res. 38, 4246-4253.

42. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517, 640-644.

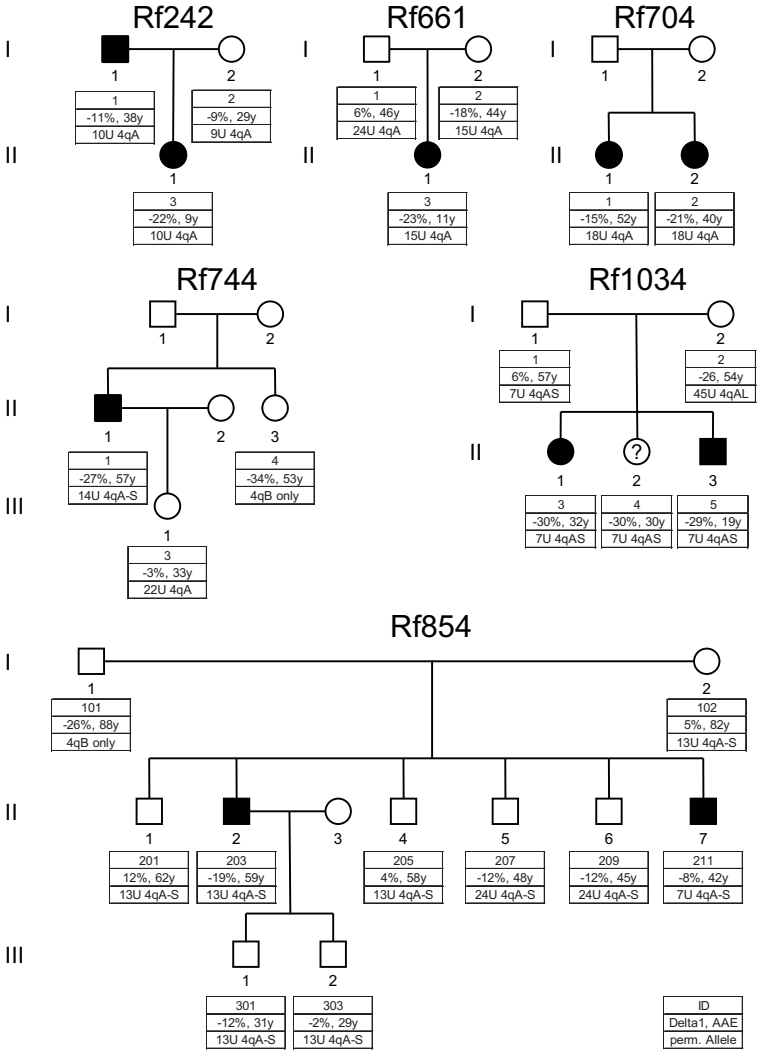
43. Carpenter, N.J., Filipovich, A., Blaese, R.M., Carey, T.L., and Berkel, A.I. (1988). Variable immunodeficiency with abnormal condensation of the heterochromatin of chromosomes 1, 9, and 16. J. Pediatr. 112, 757-760.

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

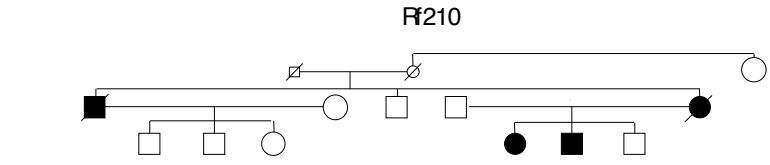
45. de Greef, J.C., Wohlgemuth, M., Chan, O.A., Hansson, K.B., Smeets, D., Frants, R.R., Weemaes, C.M., Padberg, G.W., and van der Maarel, S.M. (2007). Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. Neurology 69, 1018-1026.

SUPPLEMENTARY INFORMATION

Supplementary figures

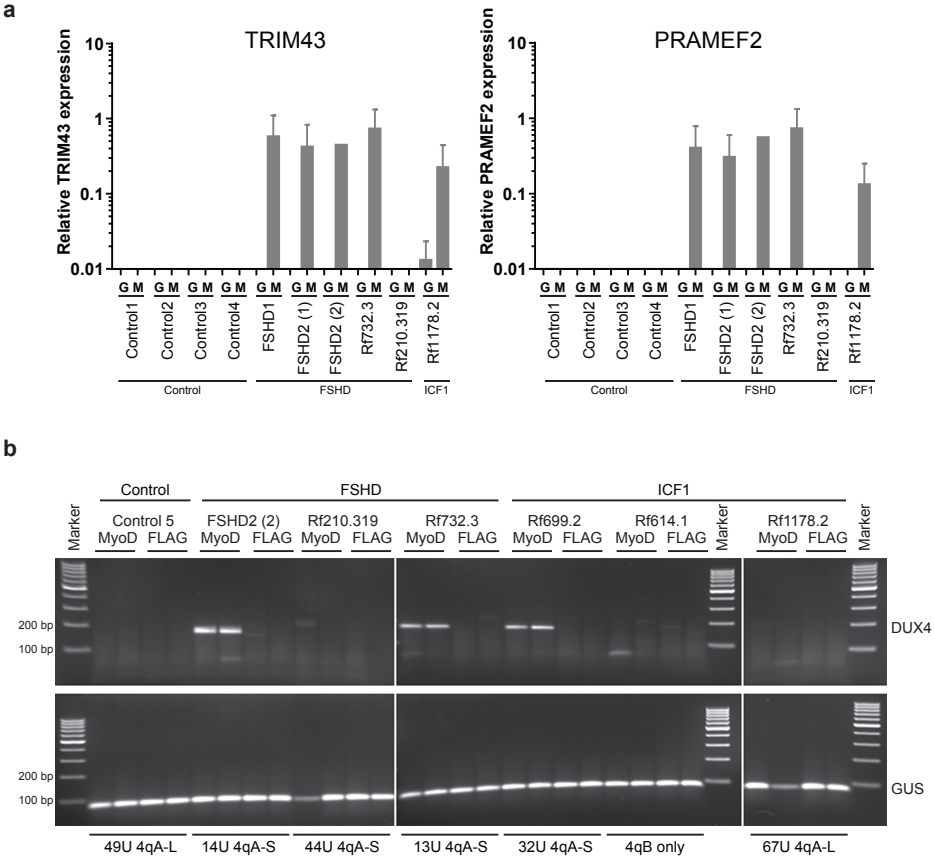


SUPPLEMENTARY FIGURE 1. FSHD2 families without *SMCHD1* mutation. Families with evidence for hereditary D4Z4 hypomethylation that were tested negative for exonic *SMCHD1* and *DNMT3B* mutations. In all families, the Delta1 score was moderately to strongly reduced with possibility of dominant or recessive inheritance of D4Z4 hypomethylation. Family Rf854 was presented previously as FSHD1 (Rf854.211) and *SMCHD1* mutation negative FSHD2 family.[1] Key: ID = identifier, Delta1 score, AAE = age at examination, number of repeat units (U) on smallest permissive allele.



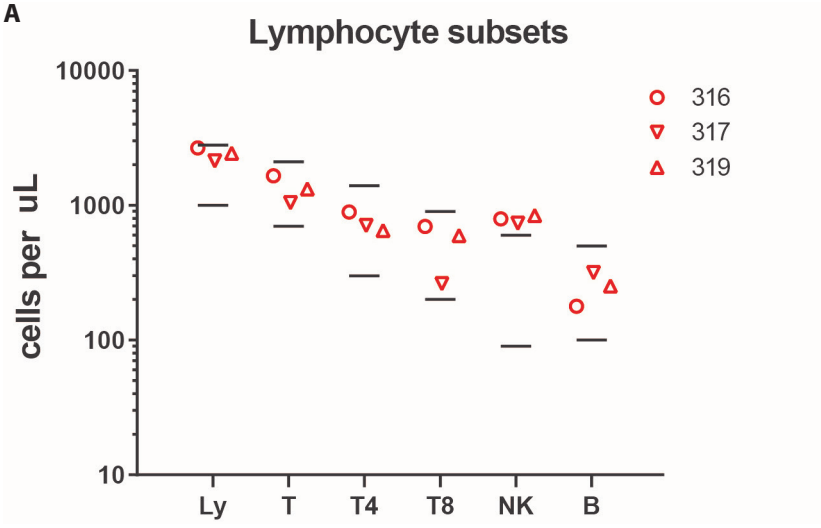
nr	201		301		303		306		202		209		211		316		317		319		212		104			
allele	a	b	a	a'	a	a'	b	a'	a'	a'	a	c	d	e	d	a	d	a	d	a	b	a	d'	c	position	dbSNP
	G	G	G	G	G	G	G	G	NA	NA	G	C	C	C	C	G	C	G	C	G	G	G	C	C	exon 1	rs2430853
	A	A	A	A	A	A	NA	NA	A	A	A	G	G	G	G	A	G	A	G	A	A	A	NA	NA	intron 1	rs555117
	A	A	A	A	A	A	NA	NA	A	A	A	A	G	G	G	A	G	A	NA	NA	A	A	NA	NA	intron 1	rs611187
	T	A	T	T	T	T	A	T	T	T	T	T	A	A	A	T	A	T	A	T	A	T	A	T	exon 14	rs8090988
	G	G	G	G	G	G	G	G	G	G	G	G	A	A	A	G	A	G	A	G	G	G	A	G	exon 14	rs635132
	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	A	A	A	A	A	G	exon 16	rs2276092
	T	T	T	T	T	T	T	T	T	T	T	A	A	A	A	T	A	T	A	T	T	T	A	A	exon 21	rs633422
	A	A	A	A	A	A	A	A	A	A	A	G	G	G	G	A	G	A	G	A	A	A	G	G	exon 28	rs521992
	A	A	A	A	A	A	A	A	A	A	A	G	A	G	A	A	A	A	A	A	A	A	A	G	exon 28	rs12327477
	A	A	A	A	A	A	A	A	A	A	A	A	G	A	G	A	G	A	G	A	A	A	NA	NA	intron 31	rs2304860
	A	A	A	A	A	A	A	A	A	A	A	G	A	G	A	A	A	A	A	A	A	A	NA	NA	intron 31	rs2304861
	T	T	T	T	T	T	T	T	T	T	T	C	C	C	C	T	C	T	C	T	T	T	C	C	exon 37	rs483547
	A	A	A	A	A	A	A	A	A	A	A	G	A	G	A	A	A	A	A	A	A	A	NA	NA	intron 43	rs3213926
Delta1	-29%		-8%		-6%		5%		-8%		8%		-2%		-8%		-8%		-29%		-30%		-2%			

SUPPLEMENTARY FIGURE 2. Exclusion of *SMCHD1* in Rf210. Haplotype analysis in family Rf210 based on common SNPs defined different alleles. Segregation analysis showed that the common *SMCHD1* allele (allele a) found in individuals with D4Z4 hypomethylation (201, 212 and 319; Delta1 values -29% and -30%) was also found in individuals with normal methylation (301, 303, 209, 316 and 317). The position of the SNPs (NA means not analyzed) and dbSNP identifier are shown on the right. Delta1 methylation values are shown below.



SUPPLEMENTARY FIGURE 3. DUX4 target gene expression and DUX4 expression in FSHD and ICF1. **(a)** Expression of DUX4 target genes *TRIM43* and *PRAMEF2* by Q-PCR in GFP- (G) or MyoD- (M) transduced fibroblasts from controls, FSHD1, FSHD2, Rf210.319, Rf732.3 and ICF1 (Rf1178.2). All transductions were performed twice for each cell line, except for control 4 (1x transduced with GFP, 2x transduced with MyoD) and FSHD2 (2) (transduced 1x with GFP and 1x with MyoD). Mean expression values with standard deviations are shown relative to the reference genes *GUSB* and *RPL27*. **(b)** Gel of Q-PCR for DUX4 and GUS in Flag and MyoD transduced fibroblasts from control, FSHD2, Rf210.319, Rf732.3 and ICF1 (Rf699.2, Rf614.1, Rf1178.2). Technical duplicates are shown. The smallest D4Z4 repeat array on a FSHD permissive allele (4qA) in each individual is indicated below the gel. A PCR product for DUX4 is only detected in MyoD transduced fibroblasts from FSHD2, Rf732.3 and Rf699.2. The DUX4 RT-PCR is performed with primers for the most common DUX4-4A-S variant, but the primers do not recognize DUX4-4A-L. The fibroblast from Rf1178.2 carries a 4qA-L allele.

A



B

ID	IgG (g/L)	IgA (g/L)	IgM (g/L)
316	12.2	2.81	0.99
317	9.38	1.92	0.74
319	10.5	2.29	0.56
reference values	7-16	0.7-4	0.42-2.3

SUPPLEMENTARY FIGURE 4. Immunological analysis. **(a)** The numbers of lymphocytes (Ly), T-cells (T), CD4+ and CD8+ T-cell subsets (T4 and T8), NK cells (NK) and B-cells **(b)** in blood for siblings from family Rf210. The range of the normal values is depicted and represents the 5th and 95th percentiles. **(b)** Serum levels of IgG, IgA and IgM for siblings from family Rf210 and the reference values.

Supplementary tables

Supplementary table 1. Detailed genotype and phenotype FSHD families.

		Chromosome position	Transcript position				4q allele 1		4q allele 2	
Nr	gender	Chr20(GRCh37)	NM_006892.3	AAE	ACSS	delta1	units	A/B	units	A/B
Rf210.101	M	NA	NA	NA	NA	NA	20	4A161L	29	4B168
Rf210.201	M	g.[31386354T>C];[=]	c.[1579T>C];[=]	63	111	-29	9	4A161S	20	4A161L
Rf210.301	M	WT	WT	38	0	-8	20	4A161L	71	4B168
Rf210.303	M	WT	WT	35	0	-6	9	4A161S	33	4B163
Rf210.306	F	WT	WT	33	0	5	20	4A161L	33	4B163
Rf210.202	F	WT	WT	60	0	-8	33	4B163	71	4B168
Rf210.209	M	WT	WT	55	0	8	17	4B163	29	4B168
Rf210.211	M	WT	WT	56	0	-2	23	4B168	44	4A161S
Rf210.316	F	WT	WT	30	67	-8	9	4A161S	44	4A161S
Rf210.317	M	WT	WT	40	50	-8	9	4A161S	23	4B168
Rf210.319	M	g.[31386354T>C];[=]	c.[1579T>C];[=]	37	0	-29	29	4B168	44	4A161S
Rf210.212	F	g.[31386354T>C];[=]	c.[1579T>C];[=]	52	135	-30	9	4A161S	29	4B168
Rf210.102	F	NA	NA	NA	NA	NA	9	4A161S	17	4B163
Rf210.104	F	WT	WT	70	0	-2	9	4A161S	15	4B163
Rf732.1	M	g.[31389159C>T];[=]	c.[2072C>T];[=]	74	0	-22	13	4A161S	24	4B168
Rf732.3	M	g.[31389159C>T];[=]	c.[2072C>T];[=]	45	89	-22	13	4A161S	15	4B163
Rf732.5	M	WT	WT	38	0	-10	13	4A161S	25	4B168
Rf732.2	F	WT	WT	68	0	1	15	4B163	25	4B168

Table summarizing the clinical, D4Z4 methylation and genetic data from FSHD families Rf210 and Rf732. Column 1 shows the individual number, column 2 the gender (F= female, M = male), column 3 and 4 show the mutation in DNMT3B at the chromosome and transcript position, respectively, column 5 shows the age at examination (AAE), column 6 shows the age corrected clinical severity score (ACSS)[2,3], column 7 shows the Delta1 value for D4Z4 methylation at the *Fse*I site, column 8-11 show the D4Z4 array sizes, and haplotype (including S or L for 4qA-S or 4qA-L) of the 4q alleles. NA means not analyzed.

SUPPLEMENTARY TABLE 2A. Overview of individuals screened for exonic *DNMT3B* mutations. Table summarizing the clinical, D4Z4 methylation and genetic data from 20 FSHD cases screened for exonic mutations in *DNMT3B* at the LUMC, Leiden, the Netherlands.

				4q allele 1		4q allele 2		
nr	**gender**	**ACSS**	**Delta1**	**units**	**A/B**	**units**	**A/B**	**Pedigree**
Rf201.309	M	143	-21	7	4A161S	27	4B163	
Rf242.3	M	NA	-22	10	4A161	105H2	4A157	see figure S2
Rf537.1	M	NA	-18	7	4A161S	37	4B163	
Rf584.2	F	19	-18	9	4A161S	32	4B168	
Rf661.2	F	NA	-18	15	4A161	20	4A161	see figure S2
Rf704.2	F	150	-21	18	4A161	48	4A161	see figure S2
Rf744.1	M	88	-27	14	4A161S	29	4B163	see figure S2
Rf838.1	F	100	-32	11	4A161L	58	4A161L	
Rf854.203	F	34	-19	13	4A161S	15	4B163	see figure S2
Rf901.1	F	49	-17	18	4B168	37	4A161S	
Rf946.2	M	63	-22	8	4A161S	27H1	4A166	
Rf982.1	M	127	-18	7	4A161S	20	4B163	
Rf1010.1	F	NA	-21	12	4A161S	51	4B163	
Rf1034.5	M	158	-29	7	4A161S	45	4A161L	see figure S2
Rf1049.1	F	54	-22	6	4A161S	37	4A161S	
Rf1093.1	F	NA	-17	11	4A161	13	4A161	
Rf1154.2	F	NA	-32	37	4A161	43	4B163	
Rf1239.1	M	140	-19	8	4A161	36	4A161	
Rf1449a.1	F	196	-27	7	4A161	36	4B168	
Rf1464.1	M	94	-19	14	4A161S	12	4B163	

Column 2 shows gender (F= female, M = male), column 3 shows the age corrected clinical severity score (ACSS), column 4 shows the Delta1 score for D4Z4 methylation at the FseI site, column 5-8 show the sizes, SSLP size and haplotype (including S or L for 4qA-S or 4qA-L when available) of the 4q alleles. NA means not analyzed.

SUPPLEMENTARY TABLE 2B. Overview of individuals screened for exonic *DNMT3B* mutations. Table summarizing the clinical, D4Z4 methylation and genetic data from 6 cases screened for exonic mutations in *DNMT3B* at the NCNP, Tokyo, Japan.

nr	gender	age	CSS	ACSS	DR1 methylation	4qA allele
1	M	4	1	NA	23	>10
2	F	47	6	128	21	11A
3	F	44	7	159	24	8A
4	F	53	10	189	21	7A
5	F	62	6	97	17	8A
6	M	52	7	135	13	13A

Column 4 shows the DR1 methylation percentage, column 5 shows the size of the shortest 4qA allele, size information of the other 4q allele and SSLP sizes are not available.

SUPPLEMENTARY TABLE 3. Overview of ICF1 patients included in this study.

Patient identifier	DNMT3B mutations				Hansen et al. 1999 (ref 4)	Hagleitner et al. 2008 (ref 5)	Weemaes et al, 2013 (ref 6)
	Transcript position NM_006892.3		Protein position NP_008823.1				
	Allele 1	Allele 2	Allele 1	Allele 2			
Rf285.1	c.2421-11G>A	c.2421-11G>A	p.E806_R807insSTP	p.E806_R807insSTP	-	Patient 33	Patient 33
Rf285.2	c.2421-11G>A	c.2421-11G>A	p.E806_R807insSTP	p.E806_R807insSTP	Family 2	Patient 29	Patient 29
Rf286.2	c.2177T>G	c.2177T>G	p.V726G	p.V726G	Family 1	Patient 16	Patient 16
Rf614.1	c.2292G>T	c.2342_2343del	p.R764S	p.I781KfsX23	-	Patient 45	Patient 45
Rf699.2	c.1918G>C	c.1918G>C	p.G640R	p.G640R	-	-	Patient 50
Rf1178.2	c.1807G>A	c.2421-11G>A	p.A603T	p.E806_R807insSTP	Family 3	Patient 7	Patient 7

Column 1 shows the patient identifier in this study. Columns 2-5 show the positions of the DNMT3B mutations on the transcript and protein level. Columns 6-8 show the identifiers from the patients in previous studies.

SUPPLEMENTARY TABLE 4. *DNMT3B* primers used for screen for exonic *DNMT3B* mutations in this study.

DNMT3Bex2F	GGCAAGAGCATCACCCTAAG
DNMT3Bex2R	TTGTGGTGGAGGTTGTCAGAGA
DNMT3Bex3F	GACGGACTGAGAGCAAATCC
DNMT3Bex3R	CGTGATGAAAGCCAAAGACA
DNMT3Bex4F	GTGTGTTGTGATGAGTGACCCG
DNMT3Bex4R	GCTCCCCTAAGGAGCTATGC
DNMT3Bex5F	CAGGCCTCCAGTCACCTAAG
DNMT3Bex5R	AGCCACAACCAGTAGTGCAG
DNMT3Bex6F	TTCTTTTTGCCTAGGAGCCA
DNMT3Bex6R	GGTAACTGGTTTTTCCCCGT
DNMT3Bex7F	GCCTCTCCTCACTGGGATTT
DNMT3Bex7R	TTTGTCTTCAAAGGGAGGCA
DNMT3Bex8F	CACCTGGGACACACCTGTAG
DNMT3Bex8R	TCTCTTGCTTCATCCCTGC
DNMT3Bex9F	GGAATGTAGGCCCTGGCT
DNMT3Bex9R	GTGGCTGACTCTCCCAAGAA
DNMT3Bex10F2	AGGCTGAGGTGGGAGAATTG
DNMT3Bex10R2	GCAAAGAAATCAGAAGAAAGTGC
DNMT3Bex11-12F	CTGGTACCCAGGCATAGCAT
DNMT3Bex11-12R	AGGACAAGGCAGGCCTAGAG
DNMT3Bex13-14F	ACTGAGAGACCCCAGGCTTT
DNMT3Bex13-14R	GACTGCAGGAACGTAGGAGC
DNMT3Bex15F	TCCCTGTGGAAGTGGTAAGG
DNMT3Bex15R	TTCCAGAGCTTTCCAACACCC
DNMT3Bex16F	CAAGGTTTGAAGCCCTCTGA
DNMT3Bex16R	TAATCCCCAGGGACCTTTCT
DNMT3Bex17F	GCTGCTGTGTGCTCAGCATCATT
DNMT3Bex17R	GGAGGACTGGGGAAAAAGAC
DNMT3Bex18F	TGACCTCAGGTAATCCACCC
DNMT3Bex18R	CCAGTAACTTGGCCAGAAGC
DNMT3Bex19F	CCTGCTGGTCTCAGGGAATA
DNMT3Bex19R	GACCAAGAACGGGAAAGTCA
DNMT3Bex20F	GCCTCATCCATAGTCAGGGA
DNMT3Bex20R	CAGAGCCAGGTCTTTCT
DNMT3Bex21F	TGCCAGGATCATTTTCATCA
DNMT3Bex21R	TCACCAAGTGCATTTTTCCA
DNMT3Bex22F2	CAGCCCTGCCACTCTTCT
DNMT3Bex22R	TCTGCCCATTTGTGTTTTGA
DNMT3Bex23F	ACTGATGGGACTGAGGGATG
DNMT3Bex23R	ATGCCTTCAGGAATCACACC

SUPPLEMENTARY TABLE 5. Q-PCR primers used in this study.

Target	Forward	Reverse
PRAMEF2	GCAAGTTAAGCCTGGAGACG	CCCTAGCAGCAAAGATGGAG
LEUTX	AAGGAGGAGACTCCCTCAGC	AAAGAGAGTGGAGGCCCAAG
TRIM43	ACCCATCACTGGACTGGTGT	CACATCCTCAAAGAGCCTGA
RPL27	CCCACATCAAGGAACTGGAG	TGTTGGCATCCAAGGTCATA
DUX4	TCCAGGAGATGTAACTCTAATCCA	CCCAGGTACCAGCAGACC
GUSB	CCGAGTGAAGATCCCCTTTTTA	CTCATTTGGAATTTTGCCGATT
MYOG	GCCAGACTATCCCCTTCCTC	GGGGATGCCCTCTCCTCTAA
MYH3	GATTGCAGGATCTGGTGGAT	CCTGCTGGAGGTGAAGTCTC

REFERENCES

1. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat. Genet. 44, 1370-1374.
2. Ricci, E., Galluzzi, G., Deidda, G., Cacurri, S., Colantoni, L., Merico, B., Piazzo, N., Servidei, S., Vigneti, E., Pasceri, V. (1999). Progress in the molecular diagnosis of facioscapulohumeral dystrophy and correlation between the number of KpnI repeat at the 4q35 locus and clinical phenotype. Ann. Neurol. 45, 751-757.
3. van Overveld, P.G., Enthoven, L., Ricci, E., Rossi, M., Felicetti, L., Jeanpierre, M., Winokur, S.T., Frants, R.R., Padberg, G.W., and van der Maarel, S.M. (2005). Variable hypomethylation of D4Z4 in facioscapulohumeral muscular dystrophy. Ann. Neurol. 58, 569-576.
4. Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M., and Gartler, S.M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc. Natl. Acad. Sci. USA 96, 14412-14417.
5. Hagleitner, M.M., Lankester, A., Maraschio, P., Hulten, M., Fryns, J.P., Schuetz, C., Gimelli, G., Davies, E.G., Gennery, A., Belohradsky, B.H., et al. (2008). Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). J. Med. Genet. 45, 93-99.
6. Weemaes, C.M., van Tol, M.J., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C., Thijssen, P.E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E., et al. (2013). Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. Eur. J. Hum. Genet. 21, 1219-1225.

Marlinde L van den Boogaard[1,+], Peter E Thijssen[1,2,+], Caner Aytekin[3], Francesco Licciardi[4], Ayça A Kiykim[5], Lucia Spossito[6], Virgil ASH Dalm[7,8], Gertjan J Driessen[9,10], Rogier Kersseboom[11,12], Femke de Vries[11], Monique M van Ostaijen-ten Dam[13], Aydan Ikinciogullari[14], Figen Dogu[14], Matias Oleastro[6], Edgardo Bailardo[15], Lucia Daxinger[1], Ercan Nain[5], Safa Baris[5], Maarten JD van Tol[13], Corry Weemaes[16,] Silvère van der Maarel[1]

1 Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
2 Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
3 Department of Pediatric Immunology, Dr Sami Ulus Maternity and Children's Research and Educational Hospital, Ankara, Turkey
4 Department of Paediatrics II, Regina Margherita Hospital Città della Salute e della Scienza di Torino, Torino, Italy
5 Pediatric Allergy and Immunology, Marmara University Pendik Training and Research Hospital, Istanbul, Turkey
6 Department of Immunology and Rheumatology, Hospital "J.P Garrahan", Buenos Aires, Argentina
7 Department of Immunology, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands
8 Department of Internal Medicine, Division of Clinical Immunology, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands
9 Department of Paediatric Infectious Diseases, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands
10 Department of Pediatrics, Juliana Children's Hospital/Haga Teaching Hospital, The Hague, the Netherlands
11 Department of Clinical Genetics, Erasmus MC, University Medical Centre, Rotterdam, The Netherlands
12 Medical service, Stichting Zuidwester, Middelharnis, The Netherlands
13 Department of Pediatrics, Laboratory Immunology, Leiden University Medical Center, Leiden, The Netherlands
14 Department of Pediatric Immunology and Allergy, Ankara University School of Medicine, Ankara, Turkey
15 Department of Genetics, Hospital "J.P. Garrahan", Buenos Aires, Argentina
16 Department of Pediatric Infectious Diseases and Immunology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
+ These authors contributed equally

6

EXPANDING THE MUTATION SPECTRUM IN ICF SYNDROME: EVIDENCE FOR A GENDER BIAS IN ICF2

Van den Boogaard et al. 2017, Clinical Genetics Oct;92(4):380-387

ABSTRACT

Background: Immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome is a rare, genetically heterogeneous, autosomal recessive disorder. Patients suffer from recurrent infections caused by reduced levels or absence of serum immunoglobulins. Genetically, 4 subtypes of ICF syndrome have been identified to date: ICF1 (*DNMT3B* mutations), ICF2 (*ZBTB24* mutations), ICF3 (*CDCA7* mutations), and ICF4 (*HELLS* mutations).

Aim: To study the mutation spectrum in ICF syndrome.

Materials and methods: Genetic studies were performed in peripheral blood lymphocyte DNA from suspected ICF patients and family members.

Results: We describe 7 ICF1 patients and 6 novel missense mutations in *DNMT3B*, affecting highly conserved residues in the catalytic domain. We also describe 5 new ICF2 patients, one of them carrying a homozygous deletion of the complete *ZBTB24* locus. In a meta-analysis of all published ICF cases, we observed a gender bias in ICF2 with 79% male patients.

Discussion: The biallelic deletion of *ZBTB24* provides strong support for the hypothesis that most ICF2 patients suffer from a *ZBTB24* loss of function mechanism and confirms that complete absence of *ZBTB24* is compatible with human life. This is in contrast to the observed early embryonic lethality in mice lacking functional Zbtb24. The observed gender bias seems to be restricted to ICF2 as it is not observed in the ICF1 cohort.

Conclusion: Our study expands the mutation spectrum in ICF syndrome and supports that *DNMT3B* and *ZBTB24* are the most common disease genes.

INTRODUCTION

Immunodeficiency, centromeric instability, and facial anomalies (ICF; OMIM #242860/#614069) syndrome is a rare, autosomal recessive disorder. Patients suffering from this characteristic triad have a severely decreased life expectancy if left untreated. In ICF syndrome, the primary immunodeficiency is characterized by an absence or strong reduction of serum immunoglobulins in the presence of circulating B-cells, leading to recurrent and eventually fatal infections of mainly the gastro-intestinal and respiratory tracts[1; 2]. Although humoral immunodeficiency is pronounced in patients with ICF syndrome, opportunistic infections observed in about 20% of the patients also suggest involvement of a T-cell dysfunction[2]. Currently, patients are most often treated by intravenous immunoglobulin therapy and few cases underwent successful hematopoietic stem cell transplantation[1-4]. ICF syndrome patients present with a distinct but variable set of facial anomalies, including, but not limited to, hypertelorism, flat nasal bridge, and epicanthus[1; 2].

The cytogenetic hallmark of ICF syndrome is centromeric instability. Upon phytohemagglutinin stimulation of lymphocytes, aberrant chromosomal configurations can be observed in metaphase spreads. These aberrant configurations mainly involve the centromeric regions of chromosomes 1, 9, and 16 and correlate with CpG hypomethylation of repeats located in these (peri)centromeric regions[5; 6]. The combination of a compromised humoral immune response, the abnormal facial appearance, and cytogenetic abnormalities is diagnostic for all ICF syndrome patients.

ICF syndrome is a genetically heterogeneous disease. In recent years the number of disease genes has expanded from one to four, with still few cases left unexplained. About half of the cases carry biallelic mutations in the DNA methyltransferase 3B (*DNMT3B*; ICF1) gene, the first identified disease gene[7; 8]. ICF1 patients most often carry missense mutations in the catalytic methyltransferase domain of *DNMT3B*, leading to partial loss of function. Nonsense mutations in *DNMT3B* are only found in combination with missense mutations, suggesting that complete loss of *DNMT3B* activity is not compatible with life, which is consistent with *Dnmt3b* knock-out studies in mice[2; 9].

We and others have described mutations in the zinc finger and BTB (Broad-Complex, Tramtrack, and Bric a Brac) domain containing 24 (*ZBTB24*; ICF2) gene in some patients negative for mutations in *DNMT3B*[2; 4; 10-14]. Most often, homozygous nonsense mutations were identified, which are predicted to lead to an absence of full length protein in patients. In addition, a few ICF2 cases with homozygous or compound heterozygous missense mutations in *ZBTB24* have been described. More recently, missense mutations in the cell division cycle associated 7 (*CDCA7*; ICF3) gene and splice site, missense, and

nonsense mutations in the helicase, lymphoid specific (*HELLS*; ICF4) gene were found to underlie ICF syndrome in almost all patients negative for mutations in *DNMT3B* and *ZBTB24*[15].

Here we report the identification of seven ICF1 cases with previously unreported mutations in *DNMT3B*. Additionally, we present five ICF2 cases, carrying both known and novel mutations, including one case with a homozygous deletion of the complete *ZBTB24* gene. With these new cases included, the total number of genetically confirmed ICF cases reported to date is 77 (including patient 2 from Franceschini *et al*.[16] with *ZBTB24* mutations (unpublished F.Licciardi)), of which 56% have mutations in *DNMT3B*, 31% have mutations in *ZBTB24*, and the remaining 13% have mutations in *CDCA7* or *HELLS*. This study, combined with earlier mutation reports[1; 2; 4; 11; 12; 15; 17-21], also revealed a gender bias specific to the ICF2 population with 79% of all ICF2 patients being male. In summary, we expand the mutation spectrum for the two most common genetic causes of ICF, show that complete absence of the *ZBTB24* locus is compatible with human life, and identify a gender bias in ICF2.

MATERIALS AND METHODS

Patients

Nine ICF families were studied after informed consent and the study protocol was approved by the relevant institutional review boards. All patients fit the immunological, cytogenetic, and clinical criteria for ICF syndrome. Detailed clinical and immunological information is provided in the results section and in the Tables S1 and S2.

Sanger sequencing of ICF disease genes

For all index cases, *DNMT3B* and/or *ZBTB24* variant analysis was performed by Sanger sequencing after polymerase chain reaction (PCR) amplification. For *DNMT3B* the exons encoding the catalytic domain (exons 16-23) were analysed using intronic primers at a position of at least 50 nucleotides from the splice donor or acceptor site. For *ZBTB24* all coding exons (exons 2-7) were analysed using intronic primers at a position of at least 50 nucleotides from the splice donor or acceptor site. Primers and amplification conditions were published previously[10; 22]. For exon 19 of *DNMT3B* a new primer pair was used (DNMT3Bex19fwd2: 5'-ACTGGTAGGCATCACCCTGA-3' and DNMT3Bex19rev2: 5'-CACCCACCAATCATCACTGC-3'). The *DNMT3B* genomic sequence was obtained from Ensemble (build 37) [GRCh37:20:31350191:31397162] (Genomic Refseq: NG_007290.1, Transcript Refseq: NM_006892.3). The *ZBTB24* genomic sequence was obtained from Ensemble (build 37) [GRCh37:6:109781719:109806439] (Genomic Refseq: NG_029388.1,

Transcript Refseq: NM_014797.2). The functional consequences of variants were predicted using SIFT[23], Mutation Taster[24], PolyPhen-2[25], and PROVEAN[26]. Identified variants have been submitted to the Leiden Open Variant Database (http://databases.lovd.nl/shared/genes/DNMT3B; http://databases.lovd.nl/shared/genes/ZBTB24).

SNP array analysis

For family Rf1875 a single nucleotide polymorphism (SNP) Array analysis was performed with the Human Cyto 12-SNP genotyping array (Illumina, San Diego, California, USA), according to the manufacturer's instructions, to identify the breakpoints of the chromosome 6 deletion.

Statistical analysis

For the review of patient numbers and gender distributions we only included genetically confirmed cases with sufficient clinical follow up (including patient 2 from Franceschini *et al*.[16] with *ZBTB24* mutations (unpublished F.Licciardi)). A possible gender bias in the different subtypes of ICF was analysed using a non-parametric Chi-square test (SPPS version 23, IBM) under the null hypothesis of equal distribution between genders, which is expected for autosomal recessive disorders. Two confirmed ICF1 patients were added to the analysis as missing values because of discrepancies in gender between different publications they were reported in (PB and PM from Jiang et al.)[1; 21; 27].

RESULTS

Clinical description of ICF syndrome patients

All twelve ICF patients that we describe in this report display facial dysmorphism characteristics and delayed development as previously described (Table S1). Recurrent infections were frequent: six out of seven ICF1 patients and all five ICF2 patients had suffered from bronchopneumonia, and one ICF1 patient and two ICF2 patients had suffered from meningitis. Until now, opportunistic infections, such as Candida and *Pneumocystis jirovecii*, have not been reported in the ICF1 patients in this study and were only observed in one ICF2 case (Table S1). Almost all ICF2 patients reported previously display a variable degree of intellectual disability[2; 12]. Among the ICF2 patients reported in this study there is one with normal intelligence, whereas the others showed intellectual disability (Table S1). At first analysis and before initiation of intravenous immunoglobulin substitution (IVIG), all evaluable ICF1 and ICF2 patients displayed (strongly) decreased serum IgG and IgM levels, except for one ICF1 patient having a normal IgM concentration (Table S2). IgA showed a normal level in one ICF2

case and was (strongly) decreased in the remaining seven evaluable ICF patients. After diagnosis, all patients were treated with IVIG. Peripheral blood B cells were present in normal numbers in six out of seven ICF1 patients and one out of five ICF2 patients and (strongly) decreased in the remaining cases.

ICF1 patient mutations

In seven ICF patients from four different families, mutations in *DNMT3B* were identified (Fig. 1A). In total, six missense mutations in *DNMT3B* were found, which have not previously been reported in ICF syndrome. Five of them have not previously been described in dbSNP, the National Heart, Lung, and Blood Institute Exome Sequencing Project (ESP) Exome Variant Server or the Exome Aggregation Consortium (ExAC) Browser. The mutation *DNMT3B* c.2221G>A (p.G741R) has been described heterozygously in three out of 121,408 exomes in the ExAC database and dbSNP (rs769823434), a frequency concordant with the recessive nature of ICF syndrome.

All six missense mutations affect highly conserved residues in the catalytic domain of DNMT3B (Fig 1B). Each missense mutation is predicted as damaging or deleterious by at least two out of four prediction tools (Table S3). In two ICF1 families the patients (Rf1648-2.1, Rf1771-2.1, 4.2, and 4.3) carry a homozygous missense mutation in *DNMT3B* (Fig. 1C, 1D, S1A and S1B). In the two other ICF1 families the patients (Rf1840-2.1 and 2.2, Rf1693-2.4) carry compound heterozygous missense mutations (Fig 1E, 1F, S1C and S1D) in *DNMT3B*. All homozygous or compound heterozygous mutations were identified in both or only one parent, respectively, as expected for the recessive inheritance pattern of ICF syndrome (Table S1).

ICF2 patient mutations

In four ICF2 patients from four different families we identified nonsense mutations in *ZBTB24*. In a fifth family a microdeletion affecting several genes, including the complete *ZBTB24* locus, was identified (Fig. 2A). The affected individuals from the non-consanguineous family Rf1225 and the consanguineous families Rf1461 and Rf1462 carry a one base pair deletion on both alleles, which is predicted to create a frameshift and a premature stopcodon in ZBTB24 (Fig. 2B, 2C, 2D, S2A, S2B and S2C). Their parents are all heterozygous carriers of this mutation. This mutation has been previously described in one other ICF2 patient[10]. The ICF2 patient Rf1759B-2.1 (case 1 in Franceschini *et al*. 1995[16]) carries a homozygous duplication in *ZBTB24* (Fig. 2E and S2D), which creates a premature stopcodon. Both non-consanguineous parents are heterozygous carriers of this mutation. This mutation is described heterozygously in one out of 121,392 exomes in the ExAC database and dbSNP (rs770082593), again in concordance with the rare and recessive nature of ICF syndrome. Rf1875 is a consanguineous family (Fig. 2F, 2G),

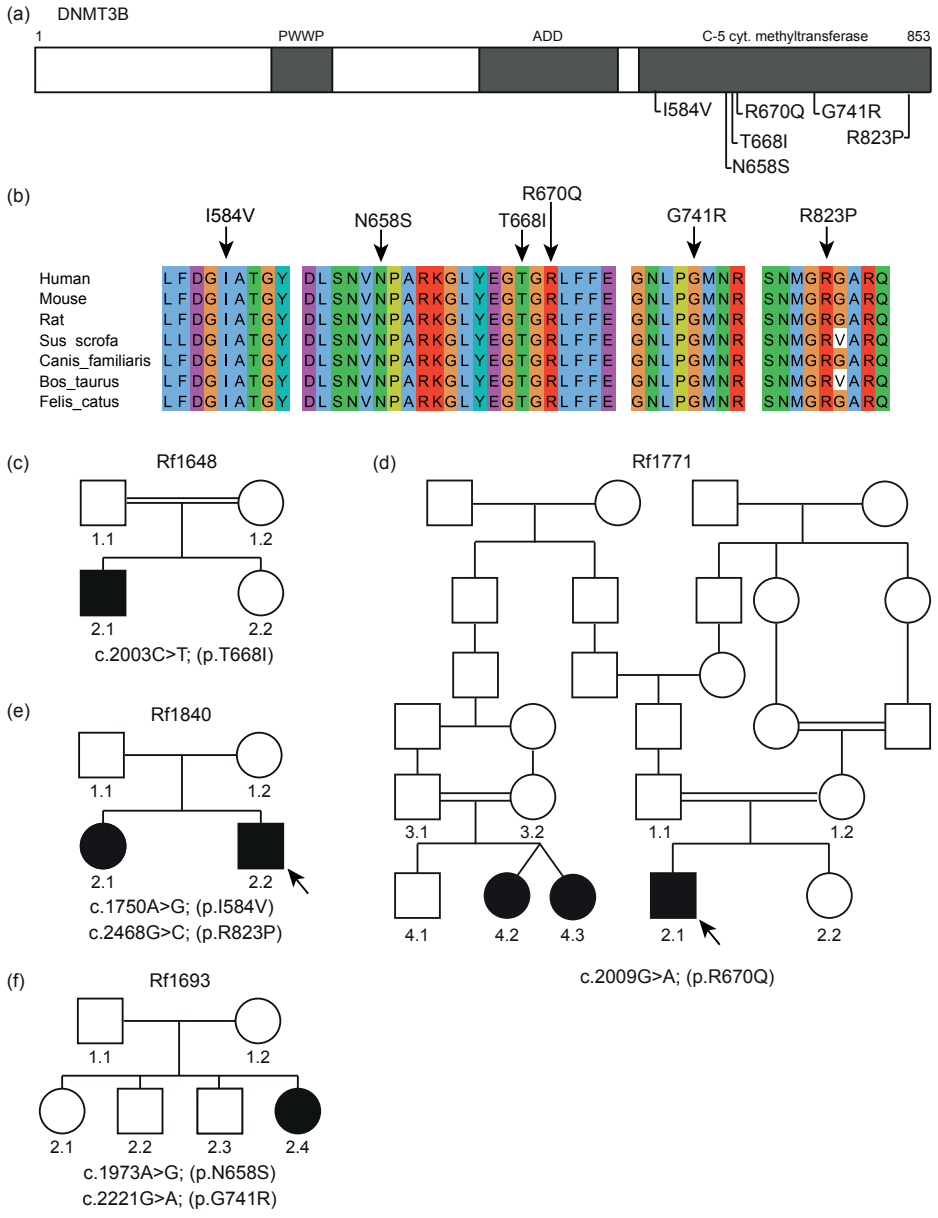


FIGURE 1. DNMT3B mutations in ICF1 families. **(a)** Schematic representation of the DNMT3B protein (GenBank: NP_008823.1). The identified amino acid changes of the ICF1 families are indicated. **(b)** Multiple-sequence alignment (MSA) of DNMT3B across distinct species for DNMT3B mutations in the ICF1 families. MSA was performed with ClustalOmega and alignment was viewed in Jalview and colored as in ClustalX. **(c-f)** Overview of the ICF1 pedigrees and their *DNMT3B* mutations. For families with multiple ICF patients, probands are indicated with an arrow. ICF, immunodeficiency, centromeric instability, and facial anomalies.

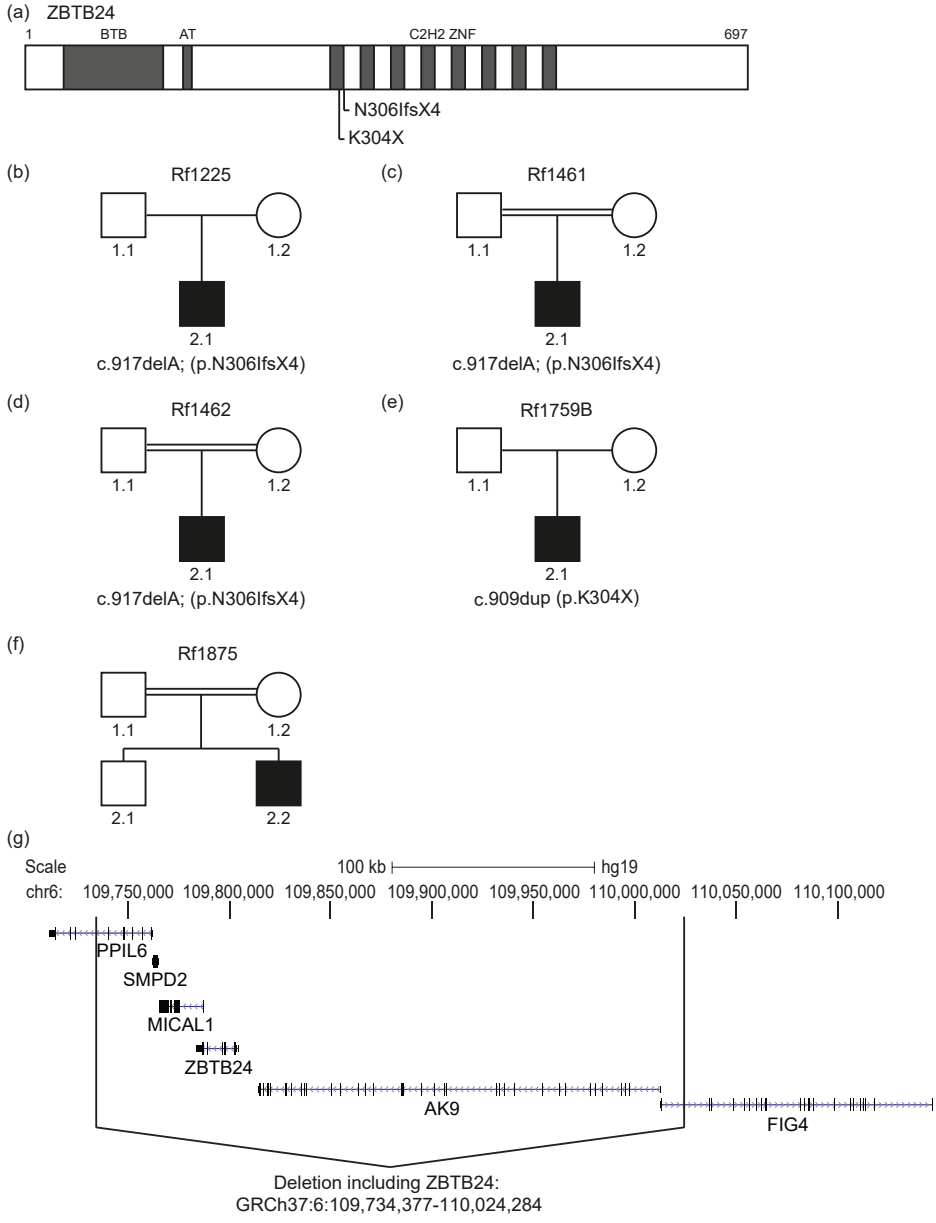


FIGURE 2. ZBTB24 mutations in ICF2 families. **(a)** Schematic representation of the ZBTB24 protein (GenBank: NP_055612.2). The identified amino acid changes of the ICF2 families are indicated. **(b-e)** Overview of the ICF2 pedigrees and their *ZBTB24* mutations. **(f)** Overview of pedigree Rf1875 and **(g)** schematic overview of the deletion on chromosome 6 which was identified in this family. ICF, immunodeficiency, centromeric instability, and facial anomalies.

in which patient Rf1875-2.2 carries a homozygous deletion on chromosome 6 that includes *ZBTB24* and several other genes. Both parents carry this deletion heterozygously, while the genotype of the brother is unknown.

Clinical heterogeneity

The *ZBTB24* mutation c.917delA was found in a total of four ICF2 patients from four unrelated families from Turkey. This suggests that this is a founder mutation. Three of these families are described in this paper, while the fourth family was previously described (P1 in de Greef *et al*. 2011[10] and P49 in Weemaes *et al*. 2013[2]). To assess whether there is clinical heterogeneity among patients carrying an identical mutation, we compared the clinical characteristics of the four male ICF2 patients carrying the c.917delA mutation. We found that the patients from these four families have similar facial anomalies and they all show signs of intellectual disability. In addition, they all have infections, such as bronchopneumonia, however only two (including patient P1 with a reported Candida infection[10]) have opportunistic infections. IgM serum levels are strongly decreased in all of them, whereas IgA levels are strongly decreased in one of them (out of three evaluable cases), and IgG levels are strongly decreased in three of them (except in P1[10]). Finally, B cell numbers are strongly decreased in one, decreased in two, and normal in one (patient P1) of them. They were all treated with IVIG. We also identified two ICF1 patients from one family (Rf1840) which are compound heterozygotes for two missense mutations in *DNMT3B* (c.1750A>G and c.2468G>C). They share similar facial anomalies and bronchopneumonia, however only one of them shows signs of intellectual disability (Rf1840.2.2). They both have strongly decreased IgA levels and normal B-cell numbers. IgM and IgG are strongly decreased in the boy (Rf1840.2.2) and normal and decreased, respectively, in the girl (Rf1840.2.1).

Finally, we identified three ICF1 patients from one family (Rf1771) with the same *DNMT3B* mutation (c.2009G>A). However, two of these patients (twin Rf1771-4.2 and 4.3) were born in 2016, which currently limits the comparison of their phenotypes with that of the third patient (Rf1771-2.1, born in 2013).

Gender bias in ICF2

With the identification of mutations in twelve ICF syndrome cases, a total of 77 genetically confirmed ICF patients has been reported[1; 2; 4; 11; 12; 15; 17-21]. 56% of the ICF syndrome patients carry mutations in *DNMT3B*, 31% carry mutations in *ZBTB24*, 6.5% carry mutations in *CDCA7*, and 6.5% carry mutations in *HELLS* (Fig. 3A). We observed a gender bias in the ICF2 population, with 19 male patients and only 5 female patients out

of 24 patients reported (Fig. 3B). This gender bias is statistically significant (χ^2 (1) = 8.167, p = 0.004) and seems to be specific for ICF2 as it is not observed for the ICF1 cohort or the overall ICF syndrome patient population (Fig. 3B).

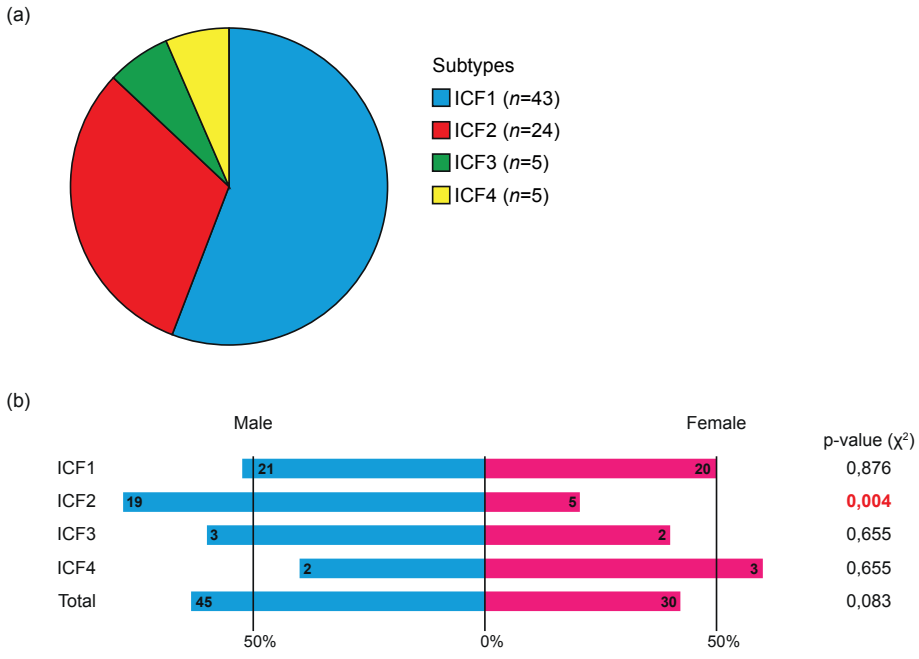


FIGURE 3. Subtype and gender distribution in ICF, immunodeficiency, centromere instability, and facial anomalies (ICF) syndrome. **(a)** Subtype distribution of the ICF patients described in the literature and in this study. **(b)** Gender distribution of the ICF patients described in the literature and in this study. P-values are calculated with the Chi-square test. Out of the 43 reported ICF1 cases, 41 were included in this analysis, 2 were added as missing values. P-values in red reflect a statistically significant difference.

DISCUSSION

In seven ICF1 patients we identified a total of six *DNMT3B* missense mutations, which have not previously been described in ICF1. However, in Rf1840 the arginine residue at postion 823 is replaced by a proline (R823P) and this residue was previously reported to be mutated into a glycine (R823G) in ICF1[19]. All mutations we identified affect highly conserved residues in the catalytic domain of DNMT3B, which is in line with the majority of the previously described ICF1 mutations[2]. Multiple missense mutations

in the catalytic domain have been shown to reduce the methyltransferase activity of DNMT3B[28; 29]. Therefore, we expect that the mutations identified in this study also reduce the catalytic activity of DNMT3B. Three of the newly identified ICF1 patients are from one family and carry the same homozygous missense mutation. Comparison of their clinical phenotypes may facilitate a better understanding of clinical variability at later ages, which is observed in the ICF1 affected siblings from the family Rf1840.

We also identified *ZBTB24* mutations in five previously unreported ICF2 patients. In four ICF2 patients homozygous nonsense mutations in *ZBTB24* were found. Nonsense mutations are the most common type of mutations in ICF2, indicating that most ICF2 patients are expected to completely lack production of ZBTB24 protein. In concordance, we also identified an ICF2 patient (Rf1875) with a homozygous deletion on chromosome 6 that includes *ZBTB24* and several other genes. This indicates that complete absence of ZBTB24 is compatible with human life and supports the conclusion that most ICF2 patients suffer from a ZBTB24 loss of function. In contrast to humans, a Zbtb24 BTB domain deletion in mice, which leads to non-functional Zbtb24 protein, was found to be embryonic lethal[30].

The deletion in ICF2 patient Rf1875-2.2 is not confined to *ZBTB24* but includes *SMPD2*, *MICAL1*, *AK9*, and parts of *PPIL6* and *FIG4*. Most of these genes have not been implicated in disease. PPIL6 is a member of the Cyclophilin Family of Peptidyl-Prolyl Isomerases with unknown activity[31]. *SMPD2* is predicted to encode a sphingomyelin phosphodiesterase, which is able to hydrolyze sphingomyelin *in vitro*, but its *in vivo* function is unclear[32]. MICAL1 is involved in cytoskeleton dynamics via its mono-oxygenase domain that can oxidize target proteins, such as actin, or produce reactive oxygen species[33]. Reduced levels of MICAL1 were detected in the temporal neocortex of patients with temporal lobe epilepsy (TLE) as well as in the hippocampus and temporal lobe of a rat model for TLE after seizures[34]. AK9 is one of the nine adenylate kinases (AK), which regulate adenine nucleotide ratios. Some AK isoforms have been related to human disease, but for *AK9* no relation to disease has been identified[35]. In contrast, mutations in *FIG4* are associated with Charcot-Marie-Tooth disease type 4J (CMT4J)[36], Yunis-Varón syndrome (YVS)[37], amyotrophic lateral sclerosis (ALS)[38], and polymicrogyria[39]. *FIG4* encodes a 5'-phosphoinositide phosphatase involved in endosomal trafficking and autophagy[39]. In patient Rf1875-2.2 the first exon of *FIG4*, which contains the start codon, is deleted, possibly resulting in the absence of full length FIG4 protein in Rf1875-2.2. This patient suffers from epilepsy, which is not a common feature of the ICF syndrome, however was reported in a family where *FIG4* mutations were linked to polymicrogyria. Although none of the other deleted genes have been linked to human disease, we cannot rule out contribution of any of those to the clinical presentation in this patient. Therefore, careful follow up with extra attention on co-morbidities normally not seen in ICF is warranted for this patient.

With the addition of twelve genetically confirmed patients, we re-analyzed the frequencies of the different ICF subtypes. For ICF1 and ICF2 we find similar frequencies, as previously reported[2]. ICF3 and ICF4 are much less frequent than ICF1 and ICF2. Unexpectedly, for ICF2 we found a significant gender bias, with 79% males and only 21% females. This suggests that lacking functional ZBTB24 is more deleterious for females. The mechanistic basis for this difference is currently unclear and warrants further studies focusing on the function of ZBTB24 during (female) development and perhaps X-chromosome inactivation, since ZBTB24 seems to colocalize with heterochromatin[12]. Currently, the limited number of female patients makes it difficult to study phenotype/genotype relations or to compare their phenotypes with those of male ICF2 patients.

In summary, in this report we describe the mutations and clinical characteristics of seven ICF1 cases and five ICF2 cases. We expand the ICF1 mutation spectrum with six new *DNMT3B* mutations and the ICF2 mutation spectrum is extended with a nonsense mutation and a homozygous deletion of chromosome 6, which includes the entire *ZBTB24* locus. For ICF2 we identify a gender bias towards male cases, which, in combination with the embryonic lethality in mice lacking functional Zbtb24, is intriguing and will require follow up studies into the function of ZBTB24.

Acknowledgements

We thank all families for participating in these studies. This study was made possible with financial support from the LUMC van de Kamp fonds, and the Dutch Cancer Society (KWF 2012-5460).

Conflict of interest

The authors declare that there is no conflict of interest associated with this manuscript.

REFERENCES

1. Hagleitner, M.M., Lankester, A., Maraschio, P., Hulten, M., Fryns, J.P., Schuetz, C., Gimelli, G., Davies, E.G., Gennery, A., Belohradsky, B.H., et al. (2008). Clinical spectrum of immunodeficiency, centromeric instability and facial dysmorphism (ICF syndrome). Journal of medical genetics 45, 93-99.
2. Weemaes, C.M., van Tol, M.J., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C., Thijssen, P.E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E., et al. (2013). Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. European journal of human genetics : EJHG 21, 1219-1225.
3. Gennery, A.R., Slatter, M.A., Bredius, R.G., Hagleitner, M.M., Weemaes, C., Cant, A.J., and Lankester, A.C. (2007). Hematopoietic stem cell transplantation corrects the immunologic abnormalities associated with immunodeficiency-centromeric instability-facial dysmorphism syndrome. Pediatrics 120, e1341-1344.
4. Harnisch, E., Buddingh, E.P., Thijssen, P.E., Brooks, A.S., Driessen, G.J., Kersseboom, R., and Lankester, A.C. (2016). Hematopoietic Stem Cell Transplantation in a Patient With ICF2 Syndrome Presenting With EBV-Induced Hemophagocytic Lymphohystiocytosis. Transplantation 100, e35-36.
5. Smeets, D.F., Moog, U., Weemaes, C.M., Vaes-Peeters, G., Merkx, G.F., Niehof, J.P., and Hamers, G. (1994). ICF syndrome: a new case and review of the literature. Hum Genet 94, 240-246.
6. Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A., and Viegas-Pequignot, E. (1993). An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. Human molecular genetics 2, 731-735.
7. Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M., and Gartler, S.M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. ProcNatlAcadSciUSA 96, 14412-14417.
8. Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 402, 187-191.
9. Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247-257.
10. de Greef, J.C., Wang, J., Balog, J., den Dunnen, J.T., Frants, R.R., Straasheijm, K.R., Aytekin, C., van der Burg, M., Duprez, L., Ferster, A., et al. (2011). Mutations in ZBTB24 are associated with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. American journal of human genetics 88, 796-804.

11. von Bernuth, H., Ravindran, E., Du, H., Frohler, S., Strehl, K., Kramer, N., Issa-Jahns, L., Amulic, B., Ninnemann, O., Xiao, M.S., et al. (2014). Combined immunodeficiency develops with age in Immunodeficiency-centromeric instability-facial anomalies syndrome 2 (ICF2). Orphanet J Rare Dis 9, 116.

12. Nitta, H., Unoki, M., Ichiyanagi, K., Kosho, T., Shigemura, T., Takahashi, H., Velasco, G., Francastel, C., Picard, C., Kubota, T., et al. (2013). Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. J Hum Genet 58, 455-460.

13. Cerbone, M., Wang, J., Van der Maarel, S.M., D'Amico, A., D'Agostino, A., Romano, A., and Brunetti-Pierri, N. (2012). Immunodeficiency, centromeric instability, facial anomalies (ICF) syndrome, due to ZBTB24 mutations, presenting with large cerebral cyst. Am J Med Genet A 158A, 2043-2046.

14. Chouery, E., Abou-Ghoch, J., Corbani, S., El Ali, N., Korban, R., Salem, N., Castro, C., Klayme, S., Azoury-Abou Rjeily, M., Khoury-Matar, R., et al. (2012). A novel deletion in ZBTB24 in a Lebanese family with immunodeficiency, centromeric instability, and facial anomalies syndrome type 2. Clin Genet 82, 489-493.

15. Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., et al. (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nat Commun 6, 7870.

16. Franceschini, P., Martino, S., Ciocchini, M., Ciuti, E., Vardeu, M.P., Guala, A., Signorile, F., Camerano, P., Franceschini, D., and Tovo, P.A. (1995). Variability of clinical and immunological phenotype in immunodeficiency-centromeric instability-facial anomalies syndrome. Report of two new patients and review of the literature. Eur J Pediatr 154, 840-846.

17. Kutlug, S., Ogur, G., Yilmaz, A., Thijssen, P.E., Abur, U., and Yildiran, A. (2016). Vesicourethral reflux-induced renal failure in a patient with ICF syndrome due to a novel DNMT3B mutation. Am J Med Genet A 170, 3253-3257.

18. Velasco, G., Walton, E.L., Sterlin, D., Hedouin, S., Nitta, H., Ito, Y., Fouyssac, F., Megarbane, A., Sasaki, H., Picard, C., et al. (2014). 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.

19. Ehrlich, M., Buchanan, K.L., Tsien, F., Jiang, G., Sun, B., Uicker, W., Weemaes, C.M., Smeets, D., Sperling, K., Belohradsky, B.H., et al. (2001). DNA methyltransferase 3B mutations linked to the ICF syndrome cause dysregulation of lymphogenesis genes. Human molecular genetics 10, 2917-2931.

20. Rechavi, E., Lev, A., Eyal, E., Barel, O., Kol, N., Barhom, S.F., Pode-Shakked, B., Anikster, Y., Somech, R., and Simon, A.J. (2016). A Novel Mutation in a Critical Region for the Methyl Donor Binding in DNMT3B Causes Immunodeficiency, Centromeric Instability, and Facial Anomalies Syndrome (ICF). J Clin Immunol 36, 801-809.

21. Jiang, Y.L., Rigolet, M., Bourc'his, D., Nigon, F., Bokesoy, I., Fryns, J.P., Hulten, M., Jonveaux, P., Maraschio, P., Megarbane, A., et al. (2005). DNMT3B mutations and DNA methylation defect define two types of ICF syndrome. Hum Mutat 25, 56-63.

22. van den Boogaard, M.L., Lemmers, R.J., Balog, J., Wohlgemuth, M., Auranen, M., Mitsuhashi, S., van der Vliet, P.J., Straasheijm, K.R., van den Akker, R.F., Kriek, M., et al. (2016). Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. American journal of human genetics 98, 1020-1029.

23. Ng, P.C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. Genome Res 11, 863-874.

24. Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11, 361-362.

25. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248-249.

26. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. PloS one 7, e46688.

27. Tauris, P., and Hansen, P.W. (1983). Functional studies on lymphocytes from two siblings with congenital hypogammaglobulinaemia. Scand J Haematol 30, 117-124.

28. Gowher, H., and Jeltsch, A. (2002). Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. J Biol Chem 277, 20409-20414.

29. Moarefi, A.H., and Chedin, F. (2011). ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. J Mol Biol 409, 758-772.

30. Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K., Wang, J., den Hamer, B., Aytekin, C., van der Maarel, S.M., and Daxinger, L. (2016). Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. Human molecular genetics.

31. Davis, T.L., Walker, J.R., Campagna-Slater, V., Finerty, P.J., Paramanathan, R., Bernstein, G., MacKenzie, F., Tempel, W., Ouyang, H., Lee, W.H., et al. (2010). Structural and biochemical characterization of the human cyclophilin family of peptidyl-prolyl isomerases. PLoS Biol 8, e1000439.

32. Horres, C.R., and Hannun, Y.A. (2012). The roles of neutral sphingomyelinases in neurological pathologies. Neurochem Res 37, 1137-1149.

33. Giridharan, S.S., and Caplan, S. (2014). MICAL-family proteins: Complex regulators of the actin cytoskeleton. Antioxid Redox Signal 20, 2059-2073.

34. Luo, J., Xu, Y., Zhu, Q., Zhao, F., Zhang, Y., Peng, X., Wang, W., and Wang, X. (2011). Expression pattern of Mical-1 in the temporal neocortex of patients with intractable temporal epilepsy and pilocarpine-induced rat model. Synapse 65, 1213-1221.

35. Panayiotou, C., Solaroli, N., and Karlsson, A. (2014). The many isoforms of human adenylate kinases. Int J Biochem Cell Biol 49, 75-83.

36. Chow, C.Y., Zhang, Y., Dowling, J.J., Jin, N., Adamska, M., Shiga, K., Szigeti, K., Shy, M.E., Li, J., Zhang, X., et al. (2007). Mutation of FIG4 causes neurodegeneration in the pale tremor mouse and patients with CMT4J. Nature 448, 68-72.

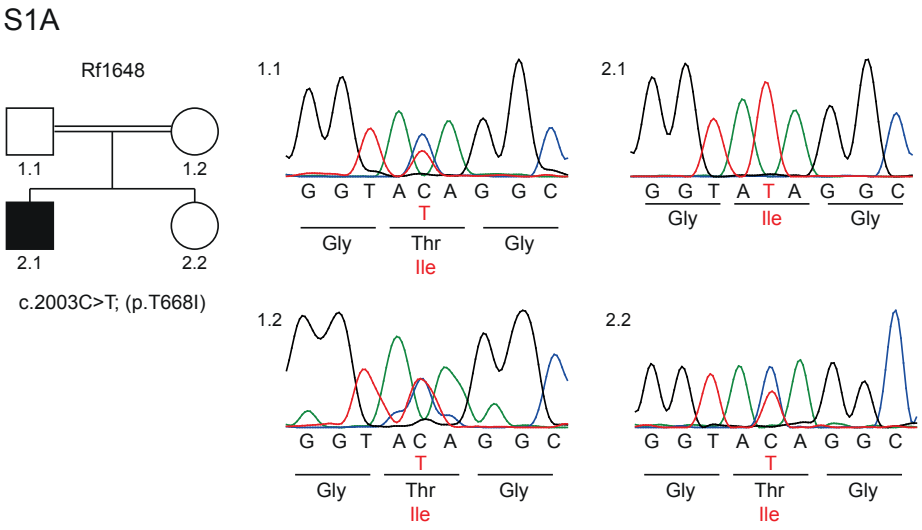
37. Campeau, P.M., Lenk, G.M., Lu, J.T., Bae, Y., Burrage, L., Turnpenny, P., Roman Corona-Rivera, J., Morandi, L., Mora, M., Reutter, H., et al. (2013). Yunis-Varon syndrome is caused by mutations in FIG4, encoding a phosphoinositide phosphatase. American journal of human genetics 92, 781-791.

38. Chow, C.Y., Landers, J.E., Bergren, S.K., Sapp, P.C., Grant, A.E., Jones, J.M., Everett, L., Lenk, G.M., McKenna-Yasek, D.M., Weisman, L.S., et al. (2009). Deleterious variants of FIG4, a phosphoinositide phosphatase, in patients with ALS. American journal of human genetics 84, 85-88.

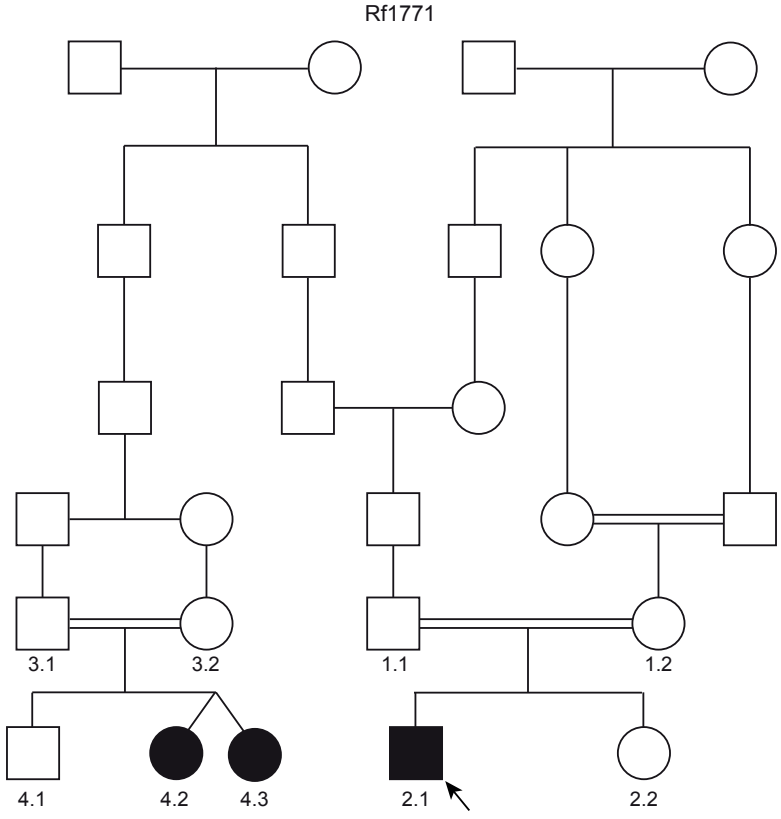
39. Baulac, S., Lenk, G.M., Dufresnois, B., Ouled Amar Bencheikh, B., Couarch, P., Renard, J., Larson, P.A., Ferguson, C.J., Noe, E., Poirier, K., et al. (2014). Role of the phosphoinositide phosphatase FIG4 gene in familial epilepsy with polymicrogyria. Neurology 82, 1068-1075.

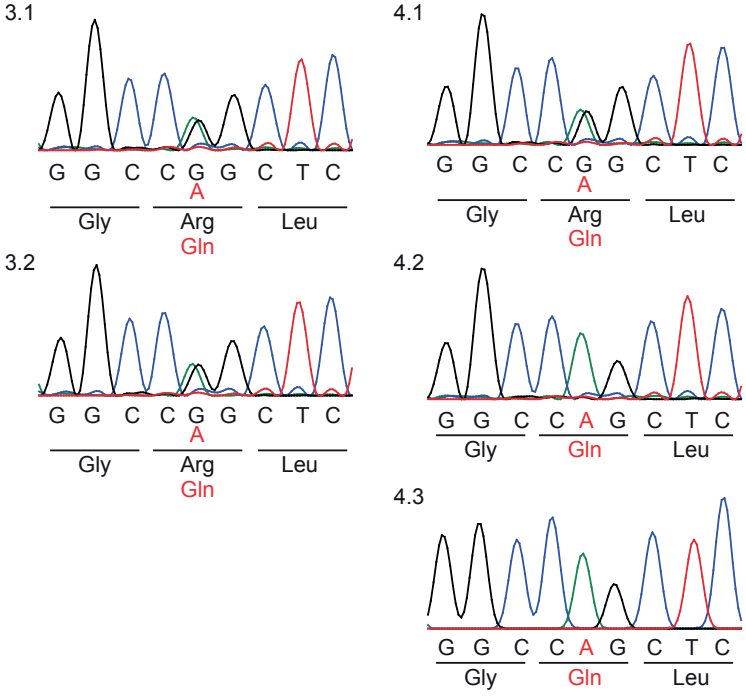
SUPPLEMENTARY INFORMATION

Supplementary figures

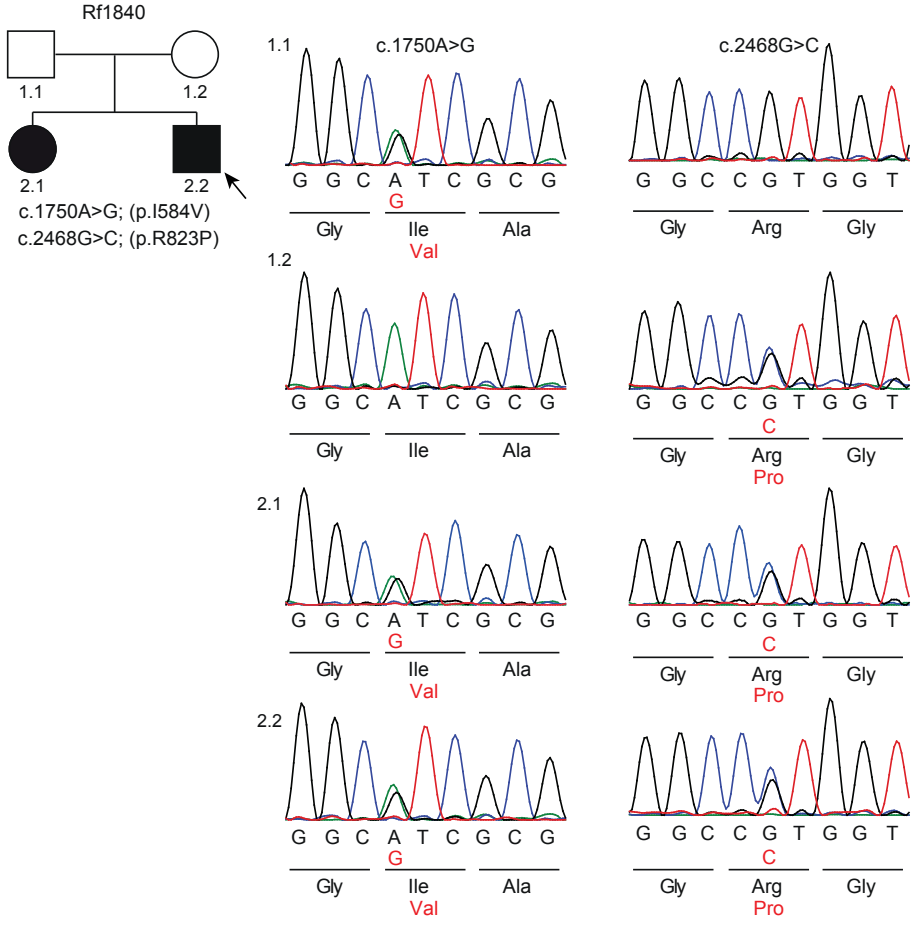


S1B

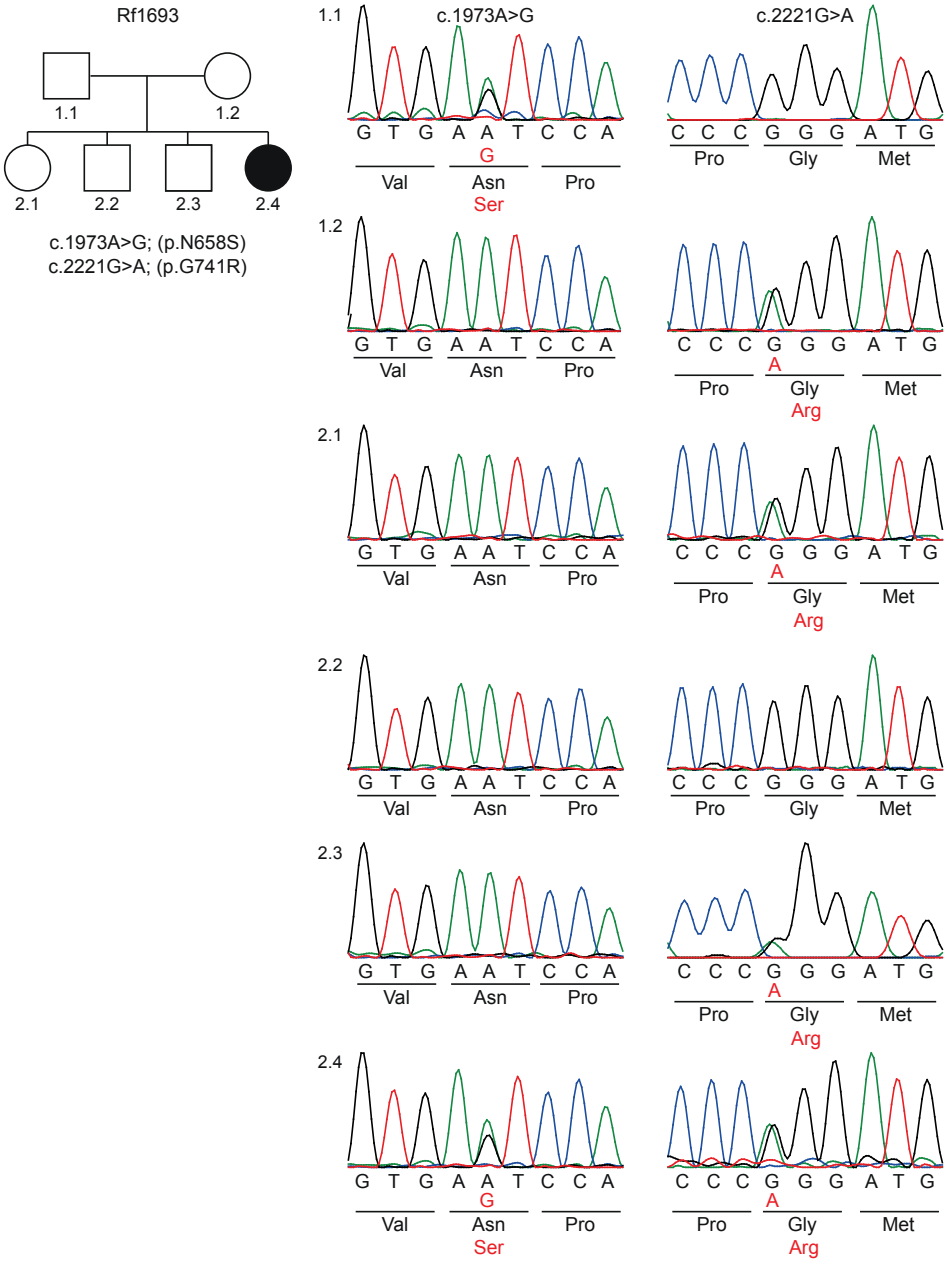

Rf1771
3.1
3.2
1.1
1.2
4.1
4.2
4.3
2.1
2.2
c.2009G>A; (p.R670Q)
1.1
G G C C G G C T C
A
Gly
Arg
Gln
Leu
2.1
G G C C A G C T C
Gly
Gln
Leu
1.2
G G C C G G C T C
A
Gly
Arg
Gln
Leu
2.2
G G C C G G C T C
A
Gly
Arg
Gln
Leu


3.1
G G C C G G C T C
A
Gly
Arg
Gln
Leu
3.2
G G C C G G C T C
A
Gly
Arg
Gln
Leu
4.1
G G C C G G C T C
A
Gly
Arg
Gln
Leu
4.2
G G C C A G C T C
Gly
Gln
Leu
4.3
G G C C A G C T C
Gly
Gln
Leu

S1C

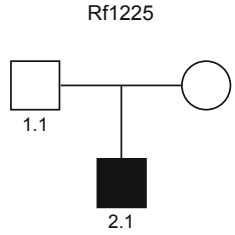

Rf1840
1.1
1.2
2.1
2.2
c.1750A>G; (p.I584V)
c.2468G>C; (p.R823P)
1.1
c.1750A>G
c.2468G>C
G G C A T C G C G
G
Gly Ile Ala
Val
G G C C G T G G T
Gly Arg Gly
1.2
G G C A T C G C G
Gly Ile Ala
G G C C G T G G T
C
Gly Arg Gly
Pro
2.1
G G C A T C G C G
G
Gly Ile Ala
Val
G G C C G T G G T
C
Gly Arg Gly
Pro
2.2
G G C A T C G C G
G
Gly Ile Ala
Val
G G C C G T G G T
C
Gly Arg Gly
Pro

S1D

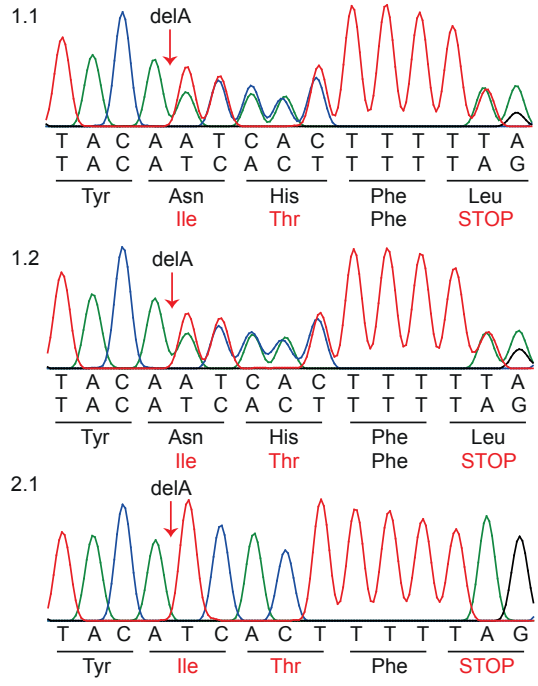


SUPPLEMENTARY FIGURE S1. *DNMT3B* mutations in ICF1 families. **(a-d)** Overview of the ICF1 pedigrees, their DNMT3B mutations and the Sanger sequence track of these mutations. For families with multiple ICF patients, probands are indicated with an arrow.

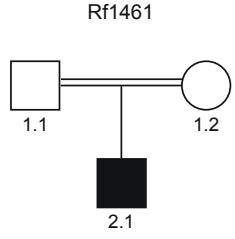
S2A



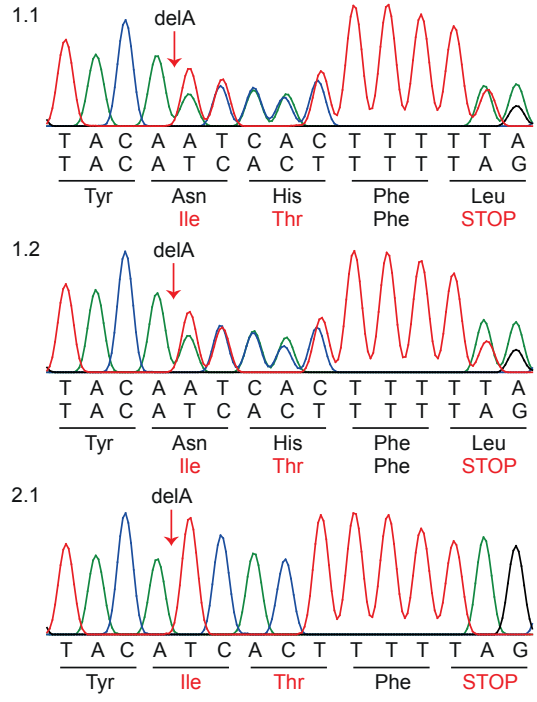
c.917delA; (p.N306IfsX4)

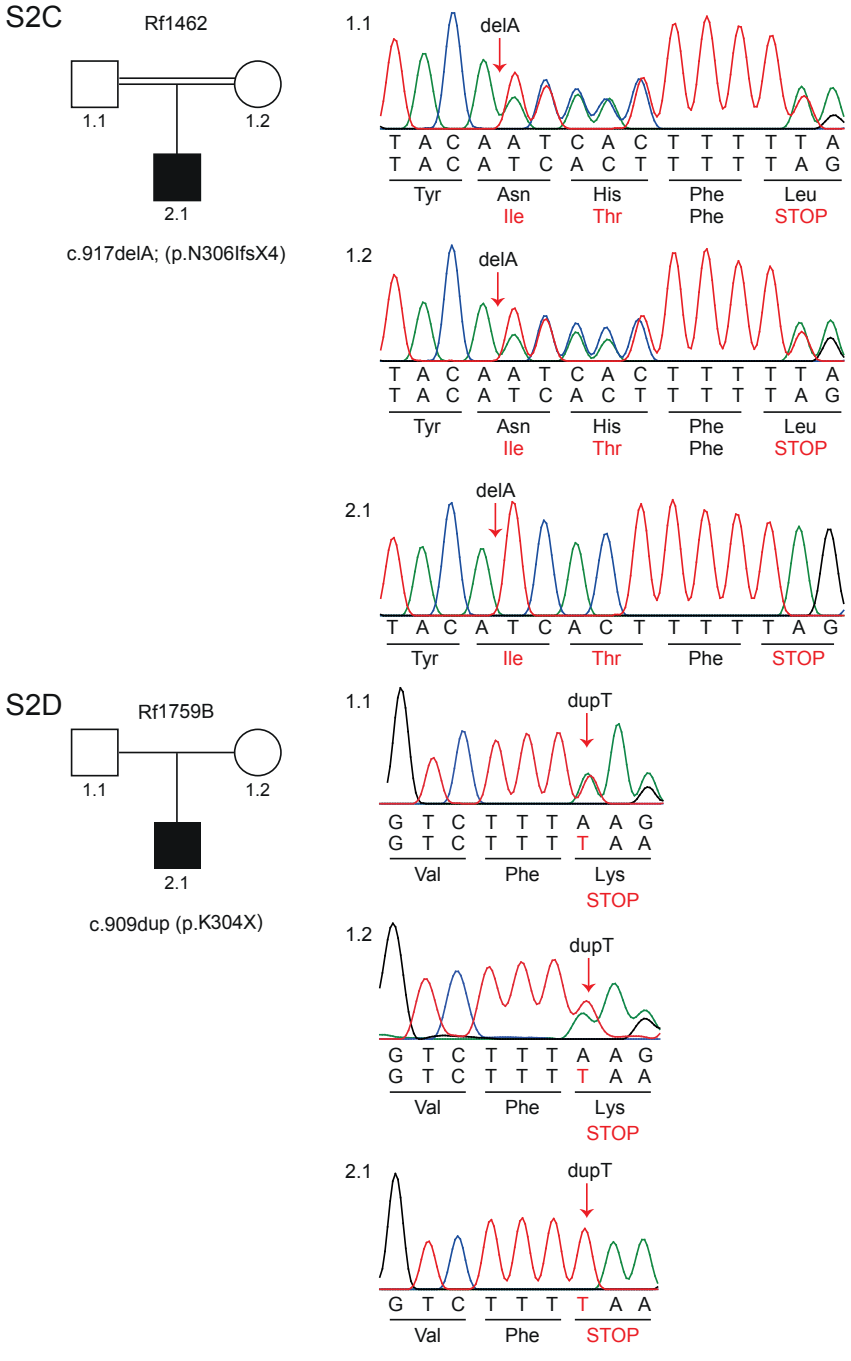


S2B



c.917delA; (p.N306IfsX4)





SUPPLEMENTARY FIGURE S2. *ZBTB24* mutations in ICF2 families. **(a-d)** Overview of the ICF2 pedigrees, their *ZBTB24* mutations and the Sanger sequence track of these mutations.

Supplementary tables

SUPPLEMENTARY TABLE 1. Overview of the clinical data of the ICF patients.

	ICF1								
Patient	**Rf1648-2.1**	**Rf1771-2.1**	**Rf1771-4.2**	**Rf1771-4.3**	**Rf1840-2.1**	**Rf1840-2.2**	**Rf1693-2.4**	**Total this study**	**Weemaes et al. EJHG 2013**
Year of birth	2004	2013	2016	2016	2007	2009	2009		
Gene defect	DNMT3B	DNMT3B	DNMT3B	DNMT3B	DNMT3B	DNMT3B	DNMT3B	DNMT3B	DNMT3B
Mutation Paternal Allele	c.2003C>T	c.2009G>A	c.2009G>A	c.2009G>A	c.1705A>G	c.1705A>G	c.1973A>G		
Mutation Maternal Allele	c.2003C>T	c.2009G>A	c.2009G>A	c.2009G>A	c.2468G>C	c.2468G>C	c.2221G>A		
Patients included								7	23
Facial anomalies								7/7	21/22
hypertelorism	+	+			+	+	-	4/5	14/18
flat nasal bridge	-	+	+	+	+	+	+	6/7	13/16
epicanthus	+	-	+	+	+	+	+	6/7	14/17
up-turned nose	+	-	+	+	+	+	-	5/7	6/9
macroglossia	-	-	-	-		-	-	0/6	5/11
telecanthus	-	-	-	-		+	+	2/6	3/11
micrognathia	+	-	+	+		-	+	4/6	5/12
low-set ears	+	+	-	-	+	+	-	4/7	6/14
round face	+	+	+	+	+	+	+	7/7	8/10
Growth and development									
gestional age < 37 weeks	-	-	-	-	-	-	-	0/7	3/20
birth weight < P10	+		-	-	-	+	-	2/6	9/20
failure to thrive	-	-	+	-	-	+	-	2/7	8
delay in motor development	-	-			-	+		1/4	9/16
delay in speech development	+				-	+		2/3	14/16
Malformations									
congenital			+					1	7
cerebral									2
Intelligence									
normal		+			+			2/5	11/20
retardation	+					+	+	3/5	9/20
Neurology									
seizures								0	3
Gastrointestinal problems									
diarrhea	+							1	7/14
Infections									
otitis	+							1	8/13
bronchopneumonia	+	+	+	+	+	+	-	6/7	16/16
sepsis									5
meningitis							+*	1	
Candida									4
Pneumocystis jirovecii									2
cytomegalovirus	-	+	-	-	-	-	-	1	2
Malignancy									2

The gene defects and clinical data of the ICF patients described in this study are listed per patient. Columns 9 and 16 list the total number of patients displaying the respective trait/total number of patients of which data on the respective trait is reported. As a comparison the same data are listed from an earlier study in column 10 and 17 [1]. *: Pseudomonas, **: Streptococcus pneumoniae

	ICF2						
	Rf1225-2.1	**Rf1461-2.1**	**Rf1462-2.1**	**Rf1759B-2.1**	**Rf1875-2.2**	**Total this study**	**Weemaes et al. EJHG 2013**
	2011	2012	2001	1985	1995		
	ZBTB24	ZBTB24	ZBTB24	ZBTB24	ZBTB24	ZBTB24	ZBTB24
	c.917delA	c.917delA	c.917delA	c.909dup	Deletion		
	c.917delA	c.917delA	c.917delA	c.909dup	Deletion		
						5	13
						5/5	13/13
	+	+	+	+	+	5/5	7/13
	+	+	+	+	+	5/5	8/9
	+	+	+	+	-	4/5	7/8
	+	+	+	-		3/4	4/7
	+	-	+	-		2/4	1/5
	+	+	+	+	-	4/5	2/4
	+	+	-	-	+	3/5	3/8
	+	+	+		+	4/4	5/7
	+	+	+	-		3/4	6/8
	+	-	-	+	-	2/5	1/7
	+	+	+	+	-	4/5	4/6
	-	-	-	-	-	0/5	3
	-	+	+	-	+	3/5	7/8
	+	+	+	+	+	5/5	11/13
							2
							2
				+		1/5	0/13
	+	+	+		+	4/5	13/13
	-	-	-	+	+	2	1
	-	-	-	-	-	0	2/6
				+		1	2/6
	+	+	+	+	+	5/5	5/7
							1
			+**		+**	2	
	+					1	2
	+					1	2
	+	-	-	-	-	1	
							1

SUPPLEMENTARY TABLE 2. Serum levels of the different immunoglobulin isotypes and numbers of peripheral blood B cells at first analysis.

	ICF1									ICF2						
	Rf1648-2.1	Rf1771-2.1	Rf1771-4.2	Rf1771-4.3	Rf1840-2.1	Rf1840-2.2	Rf1693-2.4	This study	Weemaes et al. EJHG 2013	Rf1225-2.1	Rf1461-2.1	Rf1462-2.1	Rf1759B-2.1	Rf1875-2.2	This study	Weemaes et al. EJHG 2013
Patients included	x	x	x	x	x	x	x	7	23	x	x	x	x	x	5	13
IgG																
Patients analyzed	x	x	x	x	x	x	x	7	13	x	x	x	x	x	5	12
Normal								0	1						0	2
Decreased		x				x		2	1				x		1	4
Strongly decreased	x				x		x	3	11	x	x	x		x	4	6
Inconclusive*			x	x				2	0						0	0
IgA																
Patients analyzed	x	x	x	x	x	x	x	7	19	x	x	x	x	x	5	13
Normal								0	0				x		1	5
Decreased		x						1	0			x			1	0
Strongly decreased	x				x	x		3	13	x				x	2	4
Inconclusive*			x	x			x	3	6		x				1	4
IgM																
Patients analyzed	x	x	x	x	x	x	x	7	19	x	x	x	x	x	5	13
Normal						x		1	3						0	1
Decreased								0	0				x	x	2	3
Strongly decreased	x	x	x	x	x		x	6	16	x	x	x			3	9
B-cells																
Patients analyzed	x	x	x	x	x	x	x	7	9	x	x	x	x	x	5	8
Normal	x		x	x	x	x	x	6	4				x		1	7
Decreased		x						1	2	x	x				2	0
Strongly decreased								0	3			x		x	2	1

The serum immunoglobulin levels of the ICF patients described in this study are categorized as normal, decreased and strongly decreased for each isotype using reference values for age-matched healthy controls according to Kanariou *et al.,* 1995[2]. Peripheral blood B-cell numbers are categorized as normal, decreased and strongly decreased based on reference values for age-matched healthy controls according to Comans-Bitter *et al.*, 1997 [3]. As a comparison, the same data are listed from an earlier study [1]. For some patients from this earlier study IgG data is not included because of IVIG treatment already initiated at time of first analysis. In other cases data are lacking because they were not included at first analysis.

*: Inconclusive: measured IgA concentrations as well as lower threshold for normal range in healthy age-matched controls are at or below the sensitivity level of the quantitative assay applied; two ICF1 cases were diagnosed at 4 months of age and the measured decreased IgG levels might be from maternal origin.

SUPPLEMENTARY TABLE 3. Variant predictions for identified *DNMT3B* missense variants

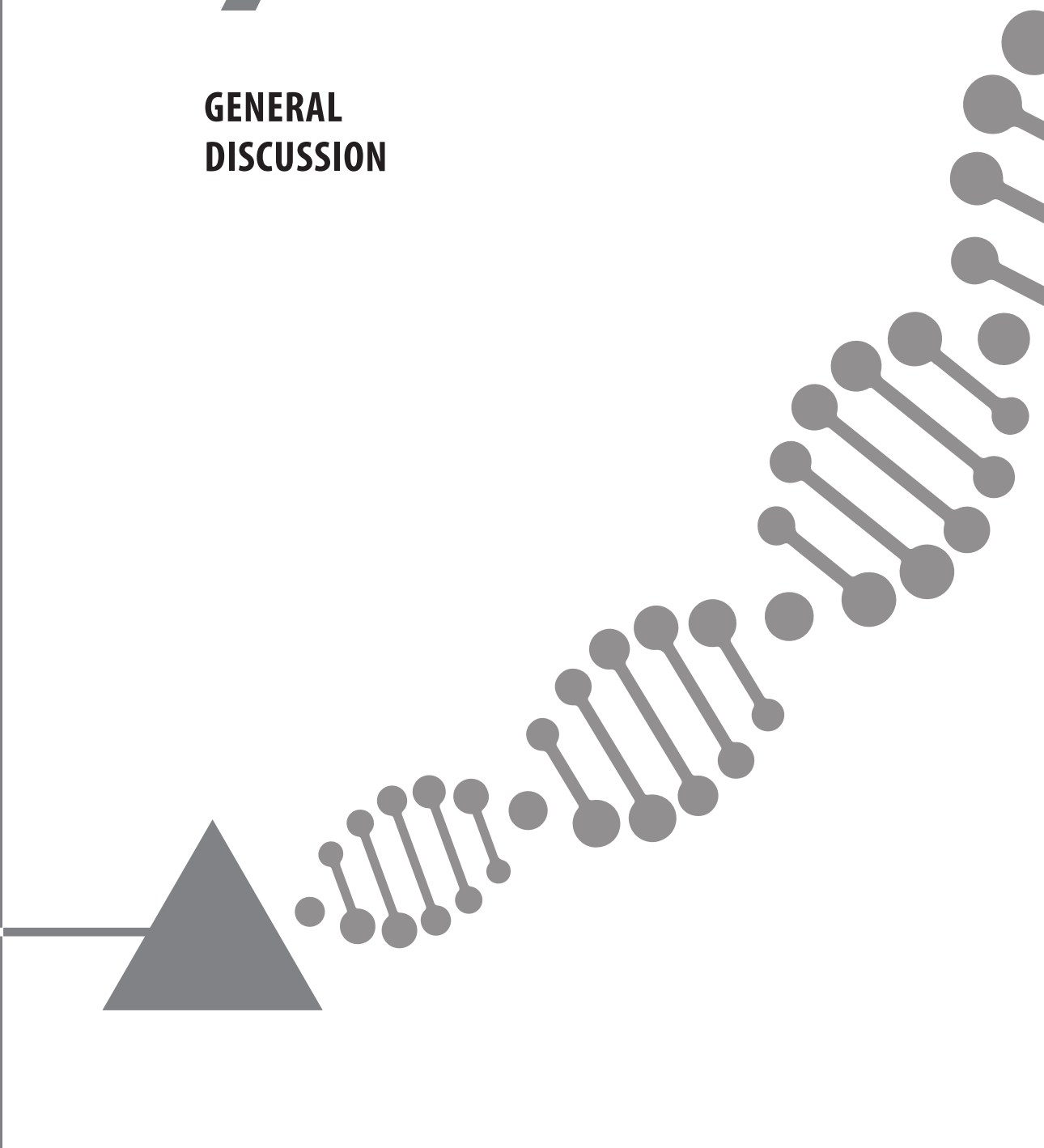
	c.2003C>T, (p.T668I)	**c.1750A>G, (p.I584V)**	**c.2468G>C, (p.R823P)**
SIFT	Deleterious (score: 0)	Deleterious (score: 0)	Deleterious (score: 0.01)
Mutation Taster	Disease causing (p-value: 1)	Disease causing (p-value: 0.995)	Disease causing (p-value: 1)
PolyPhen-2 HumVar	Probably damaging (score: 0.984)	Possibly damaging (score: 0.610)	Benign (score: 0.180)
PROVEAN	Deleterious (score: -5.692)	Neutral (score: -0.854)	Deleterious (score: -6.419)
	c.1973A>G, (p.N658S)	**c.2221G>A, (G741R)**	**c.2009G>A, p.(R670Q)**
SIFT	Tolerated (score: 0.16)	Deleterious (score: 0.02)	Deleterious (score: 0)
Mutation Taster	Disease causing (p-value: 0.924)	Disease causing (p-value: 1)	Disease causing (p-value: 1)
PolyPhen-2 HumVar	Benign (score: 0.266)	Probably damaging (score: 0.993)	Possibly damaging (score: 0.630)
PROVEAN	Deleterious (score: -4.638)	Deleterious (score: -7.378)	Deleterious (score: -3.817)

REFERENCES

1. Weemaes, C.M., van Tol, M.J., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C., Thijssen, P.E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E., et al. (2013). Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. European journal of human genetics : EJHG 21, 1219-1225.
2. Kanariou, M., Petridou, E., Liatsis, M., Revinthi, K., Mandalenaki-Lambrou, K., and Trichopoulos, D. (1995). Age patterns of immunoglobulins G, A & M in healthy children and the influence of breast feeding and vaccination status. Pediatr Allergy Immunol 6, 24-29.
3. Comans-Bitter, W.M., de Groot, R., van den Beemd, R., Neijens, H.J., Hop, W.C., Groeneveld, K., Hooijkaas, H., and van Dongen, J.J. (1997). Immunophenotyping of blood lymphocytes in childhood. Reference values for lymphocyte subpopulations. The Journal of pediatrics 130, 388-393.

7

GENERAL DISCUSSION



FSHD and ICF are two chromatin diseases characterized by hypomethylation of repetitive DNA. In FSHD there is hypomethylation of the macrosatellite repeat D4Z4 and the associated DUX4 gene, which is caused by a repeat contraction and/or variants in chromatin modifiers essential for a repressive D4Z4 chromatin structure in somatic cells. In ICF there is hypomethylation of centromeric repeats, which is caused by recessive variants in one of four ICF genes, of which two are established chromatin modifiers. In this thesis, the mutation spectrum of FSHD and ICF has been expanded, and this includes the identification of a new modifier gene for FSHD as well as variant types that had not previously been described for FSHD and ICF. These findings are important to genetically explain disease in patients, and they contribute to the understanding of disease mechanisms.

Synergistic effect of double *SMCHD1* variants

In chapter 2 we have described three FSHD2 families in which we identified two potentially damaging variants in *SMCHD1* in the proband of each family. For each family, we investigated whether the variants were present *in cis* or *in trans*, used variant effect prediction tools to predict the effect of the variants and determined whether the *SMCHD1* variants contribute to D4Z4 hypomethylation. There were no cell lines available from these families, making it impossible to study the effects of the variants on protein level.

Southern blot-based methylation analysis using probe p13E-11 and methylation sensitive restriction enzyme FseI was used to measure the D4Z4 methylation in the DUX4 promoter of the most proximal unit of the D4Z4 repeat arrays of chromosomes 4 and 10 simultaneously, as previously described[1]. With this method, the percentage of D4Z4 methylation was calculated, which we refer to as the observed FseI methylation, and a threshold for FSHD2 was determined as ≤25%[1]. Since the D4Z4 methylation level is dependent on repeat size, we also calculated the Delta1 value, which represents the D4Z4 methylation corrected for the size of the D4Z4 repeat array (observed FseI methylation minus predicted methylation based on the four D4Z4 repeat array sizes). To determine the Delta1 value, complete information on the D4Z4 repeat array sizes on chromosomes 4 and 10 was necessary, which required high molecular weight DNA. In individuals that carry an *SMCHD1* variant affecting function, the Delta1 value ranges between –42% and –22% (5th and 95th percentile, respectively)[2].

In the first family containing two *SMCHD1* variants, a splice site variant and a nonsense variant, both in exon 21 were identified *in cis* in the proband. By RT-PCR analysis we were able to demonstrate that the splice site variant disrupts the splice donor site and results in skipping of exon 21. This removes the nonsense variant from the transcript. However, skipping of exon 21 disrupts the open reading frame resulting

in a predicted stop codon in exon 22. Thus, although both variants are individually expected to affect function, in this patient only the splice site variant seems to affect the transcript and to be responsible for causing FSHD2.

In the second family containing two *SMCHD1* variants, a 4 nucleotides deletion in the splice donor site of exon 25 and a missense variant in exon 28, were identified. These variants were found to be located *in trans*. The missense variant in exon 28 was suggested to affect function by two prediction tools, however segregation analysis revealed that this variant is not causing D4Z4 hypomethylation independently (based on Delta1 value), and thus is likely a neutral variant. In this case it was helpful to have DNA available from family members to measure D4Z4 methylation and to determine the functional consequence of the missense variant in exon 28. Deletions in the splice donor site of exon 25 were already identified in 8 FSHD2 families, suggesting that this is a mutation hotspot[1-3]. In this proband we expect that only the variant in exon 25 contributes to D4Z4 hypomethylation and FSHD2.

In the third family, the proband carries two *SMCHD1* missense variants in exon 24 and 45 *in trans*. The father and mother carry the missense variant in exon 24 and 45, respectively. Since both parents also show low D4Z4 methylation values (based on Delta1 value), these variants independently cause D4Z4 hypomethylation. A more deleterious effect of both variants on D4Z4 methylation was detected in the proband and his sister who also carries both variants. The proband is severely affected, which is probably explained by the synergistic effect of the two variants. Due to other phenotypes, it was impossible to establish whether the sister of the proband was also affected with FSHD. Both parents of the proband were unaffected. For the father this is explained by the absence of a permissive 4qA allele. The mother carries a permissive 4qA allele of 27 units, similar to the proband, which might be too long to develop FSHD2 with only a single *SMCHD1* variant. A study of a large cohort of FSHD2 families indeed provided evidence that D4Z4 repeat array sizes up to 20 units are most sensitive to SMCHD1 variants[2]. This family also demonstrates that having two *SMCHD1* copies, each containing a missense variant affecting function, is compatible with life in males and females. Possibly, these missense variants create hypomorphic alleles in which SMCHD1 protein function is only partially diminished. In mouse models, homozygous knockout of Smchd1 is embryonic lethal in females, because of a defect in X-chromosome inactivation[4]. However, no models with heterozygous or homozygous missense variants have been reported.

Hemizygosity for *SMCHD1* in FSHD2 and 18p deletion syndrome

In chapter 3 we describe microdeletions that delete the entire *SMCHD1* locus and neighboring genes in two distantly related FSHD2 families. These families were first described as negative for *SMCHD1* variants because of technical limitations of the

variant analysis. This study shows that in FSHD2 patients *SMCHD1* hemizygosity should be considered in the absence of heterozygous *SMCHD1* SNPs. Four genes neighboring *SMCHD1* were also deleted in these families, however hemizygosity of these genes has not been linked to disease thus far. This study also has important implications for patients with 18p deletion syndrome, who most often also have only one copy of *SMCHD1*. As expected, D4Z4 hypomethylation (based on observed FseI methylation) could be detected in most of these patients. However, only 18p deletion syndrome patients with an intermediate-sized permissive D4Z4 allele (≤16 units) are expected to be at risk of developing FSHD. Based on the prevalence of these intermediate-sized permissive D4Z4 alleles in the population, 1:8 18p deletion syndrome patients might be at risk of developing FSHD. From the current study, two 18p deletion syndrome patients have a D4Z4 repeat array in this size range, and are thus at risk to develop FSHD. However, information on their clinical status was unavailable. So far no muscular dystrophy phenotypes have been reported in 18p deletion syndrome patients. Nevertheless, clinicians should be aware of possible co-occurrence of 18p deletion syndrome with FSHD. To study possible D4Z4 derepression in 18p deletion patients with an intermediate-sized D4Z4 repeat array, it would be interesting to obtain muscle or skin biopsies from these individuals to study *DUX4* expression. Since *DUX4* is not expressed in fibroblasts derived from skin biopsies, a MyoD transduction would need to be performed to transdifferentiate them into myogenic cells in order to explore *DUX4* expression.

(Deep) intronic variants in *SMCHD1* in FSHD2

In chapter 4 we identified an intronic variant and a deep intronic variant in *SMCHD1* in two FSHD families. As a consequence of these variants, the intron is partially retained into the transcript. In the first family, a splice region variant was identified at c.1843-15A>G (15 base pairs upstream of exon 14), which creates a 3' splice site and results in the inclusion of 14 nucleotides of intron 13 into the transcript. The inclusion of these 14 nucleotides in the transcript is predicted to disrupt the open reading frame by creating a premature stop codon. The intronic variant was detected in the proband and his unaffected sister, which both show D4Z4 hypomethylation (based on Delta1 values). The daughter of the proband was negative for the variant and does not show D4Z4 hypomethylation. The proband carries a 14 unit permissive D4Z4 repeat array, which is in the common size range for FSHD2 (8-20 units). The sister remained unaffected because she carries two non-permissive alleles.

In the second family, a deep intronic variant was identified at c.4347-236A>G (236 base pairs upstream of exon 35), which results in the exonisation of 53 nucleotides from intron 34 into the transcript. This new exon is predicted to disrupt the open reading

frame by creating a premature stop codon in exon 35. The mother, the proband, and his siblings are all carriers of this deep intronic variant which segregates with D4Z4 hypomethylation (based on Delta1 value). In addition, the father, the proband and his siblings are carriers of a 7 unit FSHD1 allele. In this family, the deep intronic variant in *SMCHD1* is modifying disease severity since all three siblings are (severely) affected, while the father remained asymptomatic. The modifying role of *SMCHD1* for FSHD disease severity has already been described in multiple FSHD1 families with borderline repeat array sizes where carriers of both an *SMCHD1* variant and a FSHD1-sized allele are more severely affected than family members with only one of these conditions[3; 5]. Indeed, in this family the father carries a 7 units D4Z4 repeat array and remained unaffected as is seen more often for these upper sized FSHD1 alleles[6; 7]. The mother carries a permissive allele of 44 units, which is likely too long to develop FSHD2. This may explain why she remained unaffected.

Genome wide consequences of *SMCHD1* variants

SMCHD1 is a chromatin modifier with an important role in silencing of D4Z4 and in X chromosome inactivation[1; 4; 5]. Additionally, SMCHD1 might have a role in epigenetic regulation of other loci, which also might be affected in FSHD2. Genome wide consequences of loss of *Smchd1* have mainly been studied in mice, where Smchd1 has been shown to play an important role in methylation of CpG islands and gene silencing on the inactive X-chromosome as well as at some autosomal loci, including imprinted genes and the clustered protocadherin (*Pcdh*) genes[4;8-11]. Also in human cell lines, human embryonic kidney (HEK293) cells and SH-SY5Y cells (a human neuroblastoma derived cell line), knockdown of *SMCHD1* results in misregulation of X-chromosomal genes and autosomal genes, such as the *PCDHβ* gene cluster and imprinted genes[12]. However, the genome wide consequences of heterozygous *SMCHD1* variants have not been studied extensively. A recent study compared the methylome of female carriers of an *SMCHD1* variant with female controls, and revealed that the majority of autosomal differentially methylated regions (DMRs) were located within or near gene clusters, with the *PCDH* gene cluster being most severely affected[13]. Using ChIP-qPCR, binding of SMCHD1 at *PCDHβ* cluster control region (CCR) hypersensitivity (HS) sites was identified in control primary myoblasts, which was reduced in FSHD2 primary myoblasts. In addition, expression of some of the *PCDHβ* isoforms was upregulated in FSHD2 myoblasts[13]. Also, in mouse neural stem cells binding of Smchd1 to the *Pcdh* cluster was identified, at the *Pcdha* cluster regulatory HS site HS5-1, coinciding with upregulation of *Pcdh* genes in the alpha, beta and gamma clusters in mouse *Smchd1*-null neural stem cells and whole brain of male adult *SMCHD1*-null mice[14]. This suggests that there is tissue specific or species specific binding and regulation of the *Pcdh* genes by SMCHD1/Smchd1.

Besides the *PCDH* gene cluster, some other gene clusters, such as the *HOXB* and *HOXD* clusters and the tRNA and 5S rRNA clusters, were also hypomethylated in *SMCHD1* variant carriers. In FSHD2 myoblasts expression of some tRNA isoforms and *HOXB7* was upregulated. This study emphasizes that, besides the effect on D4Z4, heterozygous *SMCHD1* variants also affect methylation at a limited number of autosomal gene clusters and influence expression levels of some genes in these regions. However, the changes in gene expression are limited.

Limited effects of heterozygous *SMCHD1* variants on transcriptional profiles in FSHD patients were already suggested by an earlier study of Yao et al., which revealed very few differentially expressed genes between FSHD1 and FSHD2 biopsies. However, the small sample size and large transcriptional effect due to *DUX4* upregulation in FSHD patients might obscure subtle transcriptional changes because of SMCHD1 dysfunction[15].

The relevance of genome wide effects of *SMCHD1* variants for the FSHD phenotype remains to be investigated. The phenotypic effects of *SMCHD1* variants could be studied best in carriers of *SMCHD1* variants without an FSHD permissive allele. No overt phenotypes have been described for carriers of an *SMCHD1* variant in combination with two 4qB chromosomes and comorbidities in FSHD2 have not been reported. The limited effects of heterozygous *SMCHD1* variants on genome wide methylation and gene expression suggest that upregulation of SMCHD1 might also have limited side-effects. This would be beneficial when exploring SMCHD1 upregulation as a therapeutic strategy for FSHD.

DNMT3B variants in FSHD families

In chapter 5 we describe heterozygous missense variants in *DNMT3B* identified in two FSHD families. The first family is an FSHD1 family with a D4Z4 repeat array of 9 units, and D4Z4 hypomethylation (based on Delta1 value) in some family members. *SMCHD1* was excluded as the causative gene for D4Z4 hypomethylation by haplotype analysis of common SNPs in *SMCHD1*. Whole exome sequencing (WES) identified a missense variant in the ADD domain of DNMT3B segregating in three family members with D4Z4 hypomethylation (based on Delta1 value). This variant was absent in all other family members with normal D4Z4 methylation levels. In this family, the *DNMT3B* variant explains the high clinical severity in two family members with both the FSHD1 allele and D4Z4 hypomethylation. Family members who carry only the FSHD1 allele of 9 units were either mildly affected, or even unaffected, as is seen more often for borderline FSHD1 alleles[6; 7]. One of the family members carries the *DNMT3B* missense variant and shows hypomethylation (based on Delta1 value), but is unaffected, which can be explained by his relatively long permissive D4Z4 repeat array of 44 units. In concordance, no *DUX4* expression was detected in MyoD transduced fibroblasts from this individual with the

array of 44 units. Also for carriers of *SMCHD1* variants the likelihood to develop FSHD is higher for individuals with smaller, normally sized D4Z4 repeat arrays (8-20 units) than for individuals with larger arrays[2]. In the second family, a missense variant in the catalytic domain of DNMT3B was identified in the proband and his unaffected father that shares a similar degree of D4Z4 hypomethylation (based on Delta1 value). Hypomethylation was absent in the brother of the proband with normal D4Z4 methylation. The father and the proband both carry a permissive D4Z4 repeat array of 13 units, which is in the common size range for FSHD2 patients. Incomplete penetrance is seen more often in FSHD families and this may explain the absence of disease in the father. In both carriers of this variant the Delta1 value is -22%, which is not as low as seen in most carriers of an *SMCHD1* variant (average Delta1 value -31%)[2]. Possibly, this milder degree of D4Z4 chromatin relaxation might contribute to the incomplete penetrance in this family. We were able to detect *DUX4* expression in MyoD transduced fibroblasts from the proband of this family. Together, this suggests that heterozygous *DNMT3B* variants can only derepress *DUX4* in combination with smaller D4Z4 repeat arrays and have incomplete penetrance.

Homozygous or compound heterozygous variants in *DNMT3B* cause ICF1 syndrome. Due to the identification of heterozygous *DNMT3B* variants in FSHD2, we compared the clinical and epigenetic characteristics of carriers of heterozygous and homozygous/compound heterozygous variants in DNMT3B from ICF1 and FSHD2 families. Severe D4Z4 hypomethylation was already described in ICF1 patients, and was also identified in the ICF1 patients we studied. Some of their parents, being carriers of a *DNMT3B* variant, also showed D4Z4 hypomethylation (based on Delta1 value), in a similar range as in the FSHD2 families with a *DNMT3B* variant. More severe hypomethylation (based on Delta1 value) is seen in the ICF1 patients, which suggests an additive effect of both *DNMT3B* variants, similar to what we have seen for carriers of two *SMCHD1* variants. The D4Z4 hypomethylation detected in ICF1 patients and some of their parents, suggests that they might be at risk for expressing *DUX4* and developing FSHD if they carry a permissive D4Z4 array. Indeed, we were able to detect DUX4 protein in myotubes of one ICF1 patient with a permissive D4Z4 array of 11 units. Additionally, we were able to detect *DUX4* expression in MyoD transduced fibroblasts from two other ICF1 patients, with permissive D4Z4 repeat arrays of 32 and 67 units, respectively. These *in vitro* experiments suggest that with homozygous or compound heterozygous *DNMT3B* variants there is sufficient epigenetic derepression to facilitate *DUX4* expression, also from longer D4Z4 repeat arrays. However, muscular dystrophy has not been reported in ICF1 patients. The short life-expectancy of ICF1 patients might explain why they do not develop FSHD. There were no fibroblasts or myoblasts available from the parents of ICF1 patients that carry heterozygous *DNMT3B* variants associated with D4Z4

hypomethylation. However, none of the ICF1 parents we studied carried a permissive D4Z4 array in the FSHD2 range, so we expect that they are unlikely to develop FSHD2. Nevertheless, clinicians should be aware of the possible co-occurrence of FSHD in ICF1 families.

We also examined some characteristics of ICF syndrome (chromosomal anomalies, NBL2 hypomethylation and immune deficiency) in carriers of heterozygous *DNMT3B* variants from the FSHD families. For one individual in the first family, which carries a missense variant in the ADD domain of DNMT3B, mild NBL2 hypomethylation and a low frequency of chromosomal anomalies was detected. However, no immunological features of ICF syndrome were identified in this individual. Similarly, no immunological abnormalities have been reported in the parents of ICF patients. The FSHD2 patient in the second family with a variant in the catalytic domain of DNMT3B did not show NBL2 hypomethylation or chromosomal anomalies. The variant in this patient might cause a milder methylation defect, which also corresponds with the less severe effect of this variant on D4Z4 methylation.

Unexplained FSHD cases

With the identification of intronic variants in *SMCHD1* in two families, *SMCHD1* hemizygosity in two families, and *DNMT3B* missense variants in two other families, we were able to explain D4Z4 hypomethylation in six families for which the disease etiology had remained enigmatic. These families emphasize the importance of further variant screening when D4Z4 hypomethylation is present in the absence of exonic *SMCHD1* variants. However, some FSHD families with D4Z4 hypomethylation (based on Delta1 value) cannot be explained by exonic variants in *SMCHD1* or *DNMT3B*. We expect that D4Z4 hypomethylation in these families could be explained by other variants in *SMCHD1* or *DNMT3B*, other gene defects, or a combination of variants. Possibly, similar to the family we described in chapter 4, deep intronic variants in *SMCHD1* or *DNMT3B* might be involved in unexplained families with D4Z4 hypomethylation. Whole genome sequencing (WGS) or RNA sequencing could be performed to identify deep intronic variants. In addition, variants in introns, promotor, or regulatory regions of *SMCHD1* might alter SMCHD1 transcript and protein levels and could result in D4Z4 hypomethylation and disease presentation. These types of variants can be identified by WGS or Sanger sequencing of these regions. However, reduced SMCHD1 protein levels might be muscle specific, making it more difficult to assess effects of these variants when no myogenic cells are available. Furthermore, the information on regulatory regions of *SMCHD1* is limited and therefore the functional consequences of identified variants will be difficult to predict. Indeed, a recent study performed an *in silico* prediction of regulatory regions of *SMCHD1* and screened these regions for variants in FSHD patients. In two FSHD

families a variant in a putative regulatory region was identified, however segregation with D4Z4 hypomethylation was inconclusive[16]. To study the functional effects of variants in regulatory regions reporter assays should be developed, in combination with segregation analysis of D4Z4 hypomethylation.

In addition to variants in *SMCHD1* and *DNMT3B*, there might be other gene defects, possibly in genes encoding proteins of the D4Z4 silencing complex, involved in FSHD patients with unexplained D4Z4 hypomethylation (based on Delta1 value). However, with WES we were so far unable to explain D4Z4 hypomethylation in these individuals. For these cases identification of causative variants will be more successful when DNA from multiple family members with and without hypomethylation is available. Alternatively, (combinations of) rare variants in chromatin modifiers, might have an unknown effect on protein function and therefore contribute to defects in D4Z4 methylation.

We also identified some patients clinically diagnosed with FSHD who don't have a contracted allele nor D4Z4 hypomethylation (Lemmers et al. unpublished results). Some of these patients might be affected with another muscular dystrophy with a (partially) overlapping phenotype, as differential diagnosis is sometimes challenging[17]. For those of whom a firm clinical FSHD diagnosis has been established, it would be interesting to focus not only on one aspect of D4Z4 chromatin relaxation, i.e. hypomethylation, but rather to assess *DUX4* expression. Since muscle biopsies are quite invasive, it would be more convenient to obtain skin biopsies from these patients and to perform a MyoD transduction on skin fibroblasts, followed by a transdifferentiation to study *DUX4* expression. Recently, a protocol describing MyoD transformation and differentiation of urine derived cells was published[18]. This makes urine cells very useful as a non-invasive source of myogenic cells. However, it remains to be tested if *DUX4* expression can be detected in urine derived myotubes from FSHD patients, and is absent in urine cells from controls. Another recent publication identified *DUX4* expression in Epstein–Barr virus (EBV) transformed B-LCLs (lymphoblastoid cell lines) from FSHD patients[19]. Upregulation of DUX4 target genes was also identified in a LCLs from an ICF1 patient[20]. Since LCLs can be derived from blood, this suggest that this might also be a less invasive source of cells, compared to skin or muscle biopsies, to study *DUX4* expression.

Gender bias in ICF2

In chapter 6 we extend the mutation spectrum for the two most common forms of ICF syndrome, ICF1 (recessive *DNMT3B* variants) and ICF2 (recessive *ZBTB24* variants). In seven ICF patients recessive missense variants in *DNMT3B* were identified (ICF1). Four ICF1 patients from consanguineous families had homozygous *DNMT3B* variants, and the remaining three were compound heterozygotes for *DNMT3B* variants. All missense variants we identified are located in the catalytic domain of DNMT3B, and are expected to

decrease the catalytic activity of DNMT3B. In four ICF patients we identified homozygous nonsense variants in *ZBTB24* (ICF2). Homozygous nonsense variants are often detected in ICF2 patients, and are suggested to result in the absence of ZBTB24 protein. In one ICF2 patient we identified a homozygous deletion on chromosome 6, which includes *ZBTB24* and several neighboring genes. This patient confirms that complete absence of ZBTB24 is compatible with life. This is in contrast to homozygous Broad-Complex, Tramtrack, and Bric a Brac (BTB) domain deletion in mice, a condition that is embryonic lethal. This suggests that ZBTB24 has (partial) different functions in human and mice. With the addition of these 12 patients, a total of 77 genetically confirmed ICF patients have been reported. We calculated the frequencies of the different ICF subtypes: 56% of the ICF patients carry *DNMT3B* variants (ICF1), 31% carry *ZBTB24* variants (ICF2), 7% carry *CDCA7* variants (ICF3) and 7% carry *HELLS* variants (ICF4). A statistically significant gender bias was detected for ICF2 patients, with 79% male and 21% female patients. This gender bias seems to be restricted to ICF2, since no gender bias was identified in the ICF1 cohort or the total ICF patient population. So far five female ICF2 patients have been reported, three of them carry homozygous or compound heterozygous nonsense variants. This suggests that although ZBTB24 variants seem to be more deleterious in females, female embryos can survive in the absence of full length ZBTB24 protein. To explain the mechanism behind this gender bias in ICF2, studies addressing the function of ZBTB24 are required. So far, studies of ZBTB24 function revealed that it acts as a transcriptional activator and/or repressor[21; 22]. Since GFP-ZBTB24 colocalizes with heterochromatin in NIH3T3 cells[23], it might have a function in X-inactivation, although ZBTB24 colocalization with the Barr body has not been reported.

It is important to perform genotype-phenotype studies in ICF patients to better understand the clinical spectrum of this disease. However, since the disease is very rare and patients are seen by different clinicians, it is currently impossible to draw conclusions about genotype-phenotype correlations. One of the ICF2 variants (*ZBTB24* c.917delA) is found within multiple families, and the effect on immunoglobulin subtype abundance and B-cell numbers vary among these patients. Similarly, from two siblings with ICF1, only one shows intellectual disability. This indicates that already with the same variant, the phenotype is variable.

Remarkably, homozygous missense variants in *ZBTB24* were recently also identified in a male patient with congenital anomalies of the kidneys and urinary tract (CAKUT) syndrome, which was characterized by bilateral vesicoureteral reflux (VUR). The patient also had intellectual disability, growth retardation and deafness, but no overt immunodeficiency. Intellectual disability and growth retardation are also seen in ICF patients, and specifically in ICF2 intellectual disability is very common[24]. Furthermore, congenital abnormalities in ICF patients have been reported more frequently[24], and

in one ICF1 patient vesicoureteral reflux has been described[25]. However, it is very remarkable that this patient doesn't have an immune defect, since no asymptomatic carriers of homozygous *ZBTB24* variants that affect function have been reported. The missense variant in this patient (c.1457G>A p.Gly486Asp) affects a highly conserved amino acid in the seventh Zinc finger of ZBTB24[26]. Missense variants are rare in ICF2, and no missense variants in this part of the protein have been identified in ICF2 patients. Possibly, the absence of an overt immunodeficiency might be related to difficulties in establishing disease causality, to the location of the variant in the protein, or the patient might be protected by an unknown mechanism. It would be interesting to investigate whether this patient has centromeric instability and hypomethylation of pericentromeric repeats. Furthermore, this case exemplifies the need for functional tests for ZBTB24. Since ZBTB24 has been reported to regulate CDCA7 transcriptional activity[21], a Cdca7 reporter assay in mouse embryonic stem (ES) cells might help to define the functional consequences of *Zbtb24* variants (Wu et al. in preparation).

DNMT3B variants cause discordant phenotypes: ICF and FSHD

The identification of *DNMT3B* variants in FSHD families, indicates that recessive or dominant variants in the chromatin modifier DNMT3B can cause very discordant phenotypes. In one of the FSHD families a variant in the ADD domain of DNMT3B was identified. Variants in the ADD domain of DNMT3B have never been reported in ICF1 syndrome. However, variants in the highly homologous ADD domain of DNMT3A and ATRX have been reported in Tatton-Brown-Rahman overgrowth syndrome (TBRS [OMIM: 615879])[27] and alpha thalassemia-mental retardation syndrome, X-linked (ATR-X [OMIM: 301040])[28], respectively.

For DNMT3A a model was proposed where DNMT3A exists in two forms, an active and an auto-inhibitory form. In the auto-inhibitory form, the ADD domain interacts with the catalytic domain, preventing it from binding to DNA. When DNMT3A is recruited to chromatin, unmethylated histone H3 lysine 4 (H3K4me0) binds to the ADD domain and stimulates a conformational change into the active form of DNMT3A, which is not induced by H3K4me3. This mechanism connects histone modification status with DNA methylation[29]. Moreover DNMT3B binds H3K4me0, and this interaction is disrupted by methylation of H3K4[30]. Since DNMT3A and DNMT3B are highly homologous, a similar auto-inhibitory mechanism of the ADD domain could be envisioned for DNMT3B. This suggests that the variant in the ADD of DNMT3B could disturb the auto-inhibitory form of DNMT3B, but it might also disturb the total structure of the protein.

Although variants in the ADD domain have never been reported in ICF1 patients, the missense variant in the ADD domain described in our FSHD family, might cause ICF1 when present in a homozygous or compound heterozygous state. This is supported by the observation of mild ICF features (NBL2 hypomethylation, chromosomal anomalies) detected in one of the carriers.

The *DNMT3B* variant in the second FSHD family is located in the catalytic domain of DNMT3B, similar to most variants in ICF1 patients. This type of variants is expected to reduce the methyltransferase function of DNMT3B. The variant identified in this family has never been reported in ICF patients, but might also cause ICF when present in a homozygous or compound heterozygous state.

To study the effect of heterozygous *DNMT3B* variants on D4Z4 methylation, we measured D4Z4 methylation in some parents and siblings of ICF1 patients. It is interesting that some but not all heterozygous missense variants in the catalytic domain of DNMT3B cause D4Z4 hypomethylation (based on Delta1 value) in the parents or siblings of ICF1 patients. The *DNMT3B* c.2177T>G (p.Val726Gly) variant did not cause D4Z4 hypomethylation in three heterozygous carriers from one ICF1 family. On the other hand the *DNMT3B* c.1807G>A (p.Ala603Thr) variant and *DNMT3B* c.1918G>C (p.Gly640Arg) variant present in two other ICF1 families are associated with D4Z4 hypomethylation in one and two heterozygous carrier(s), respectively. Possibly the methylation status at D4Z4 could be related to differences in residual activity, protein interactions, or localization of mutant DNMT3B protein. For some ICF1 variants (including p.Val726Gly and p.Ala603Thr) the functional consequences were studied by *in vitro* assays using human DNMT3B or the corresponding variants in mouse Dnmt3b. Although normal levels of D4Z4 methylation are present in heterozygous carriers of DNMT3B p.Val726Gly, recombinantly produced Dnmt3b containing the corresponding variant in mouse (p.Val732Gly) showed no or very little residual activity in *in vitro* functional assays[31-33]. Similarly, DNMT3B p.Ala603Thr or the corresponding variant in mouse (p.Ala609Thr) also shows little residual activity when recombinantly produced or introduced in Dnmt3a$^{-/-}$Dnmt3b$^{-/-}$ embryonic stem (ES) cells[32; 34], although in one study the *in vitro* activity of Dnmt3b p.Ala609Thr was slightly higher than for Dnmt3b p.Val732Gly[32]. Thus, the difference in D4Z4 hypomethylation does not seem to be explained by the presence/absence of residual activity in *in vitro* assays. Alternatively, the difference might be related to the disturbance of protein-protein interactions of DNMT3B. Co-immunoprecipitation indicated that Dnmt3b p.Ala609Thr disturbs the interaction with Dnmt3a as well as homodimerization of Dnmt3b[34]. The interaction between Dnmt3a and Dnmt3b was not disrupted by Dnmt3b p.Val732Gly[34]. However, in a glutathione S-transferase (GST) pull down assay this variant reduced the strength of the interaction with Dnmt3L[31]. In addition, both Dnmt3b p.Ala609Thr and Dnmt3b

p.Val732Gly showed altered subcellular localization in NIH3T3 embryonic fibroblasts[34]. Although the disturbance of protein-protein interactions might be relevant, thus far these functional studies were limited to few ICF1 variants, and therefore they cannot provide an explanation why some heterozygous *DNMT3B* variants do not cause D4Z4 hypomethylation. To better understand these differences, more information on D4Z4 methylation status (Delta1 values) from ICF1 family members, in combination with extended functional studies of mutant proteins is necessary. To determine the Delta1 value, full D4Z4 repeat array sizing information is necessary, for which high molecular weight DNA in agarose plugs is required. Unfortunately, in most cases only low quality DNA is available from ICF1 patients and their family members, which is insufficient to determine their Delta1 value. Therefore, for many heterozygous *DNMT3B* variants the effect on D4Z4 methylation based on the Delta1 value is unknown. It would be interesting to study the effect of more heterozygous *DNMT3B* variants on D4Z4 methylation in more carriers by bisulfite sequencing of D4Z4, for which less DNA of a lower quality is required[35; 36].

An alternative way to study the effects of *DNMT3B* variants on D4Z4 chromatin structure and *DUX4* expression, is to individually introduce these variants by CRISPR/Cas9 genome editing in a controlled genetic background with an intermediate-sized D4Z4 repeat array, for example in genetically defined human induced pluripotent stem (iPS) cells or in embryonic stem (ES) cells from the transgenic D4Z4-12.5 mice carrying a human D4Z4 repeat array of 12.5 units (12.5 D4Z4 mES cells). In human iPS cells *DUX4* is expressed, which is silenced upon differentiation to embryoid bodies in control iPS cells, but not in FSHD iPS cells[37]. Similarly, *DUX4* is expressed in 12.5 D4Z4 mES cells, and silenced during differentiation. iPS cells or 12.5 D4Z4 mES cells with heterozygous or homozygous DNMT3B variants could be differentiated to assess the effect of the variants on D4Z4 methylation and *DUX4* expression. Furthermore, to study if the FSHD variants in *DNMT3B* might also cause ICF1 when present homozygously, these variants could be introduced homozygously in iPS cells to compare genome wide methylation patterns with ICF1 iPS cells.

For the other ICF subtypes (ICF2, ICF3 and ICF4), no D4Z4 hypomethylation has been reported, although the availability of high molecular weight DNA often limits the calculation of Delta1 values. From the observed FseI methylation we can conclude that the D4Z4 methylation is not as low as for ICF1 patients. Currently, it is also unclear if variants in *ZBTB24*, *CDCA7*, and/or *HELLS* have other effects on D4Z4 chromatin structure. With MyoD transductions and ChIP experiments the effects of these variants on D4Z4 repression and chromatin structure could be studied. However, the limited

availability of patient material (with a permissive D4Z4 allele) hampers these studies. Therefore, perhaps a better way to study these effects is to make use of CRISPR/CAS9 edited human iPS cells or 12.5 mES cells, as suggested above for *DNMT3B* variants.

SMCHD1 variants cause discordant phenotypes: FSHD and arhinia/BAMS

Very recently, two papers described variants in *SMCHD1* in a total of 40 probands with arhinia, a rare congenital malformation characterized by the complete or partial absence of an external nose[38; 39]. Other craniofacial defects can co-occur with arhinia, and a severe combination of arhinia with ocular and reproductive defects is referred to as Bosma arhinia microphthalmia syndrome (BAMS, OMIM 603457). *SMCHD1* variants were identified in individuals with BAMS as well as isolated arhinia.

In contrast to FSHD2, where *SMCHD1* variants can be found over the entire locus and include both haploinsufficiency and loss of function variants[2], the *SMCHD1* variants found in arhinia/BAMS patients are all missense variants in exons 3 to 13[38; 39]. These exons encode the GHKL type ATPase domain and an adjacent region of SMCHD1. Missense variants in these exons have also been identified in FSHD2 patients, and one of the *SMCHD1* variants identified in an arhinia patient, was previously identified in an FSHD2 patient[38].

D4Z4 methylation was tested in both studies with bisulfite sequencing of D4Z4 at chromosomes 4 and 10 simultaneously. Interestingly, Shaw et al. detected D4Z4 hypomethylation in 74% of the arhinia/BAMS patients. Gordon et al. identified only a trend towards D4Z4 hypomethylation in arhinia/BAMS patients compared to their unaffected relatives and controls. Since the D4Z4 repeat array size influences D4Z4 methylation, some of the arhinia/BAMS patients might not show hypomethylation with this assay because of long D4Z4 repeat arrays. Nevertheless, this shows that FSHD2 patients and (some) arhinia/BAMS patients share D4Z4 hypomethylation.

Gordon et al. suggest that the *SMCHD1* variants found in arhinia/BAMS patients might result in a gain of function of the SMCHD1 protein, as opposed to a loss of function mechanism in FSHD. This was based on the observation that three different arhinia/BAMS variants introduced in the recombinant N-terminal domain of mouse Smchd1 gave increased ATPase activity compared to wildtype SMCHD1. One other arhinia/BAMS variant did not have an effect in this assay. In contradistinction, two FSHD2 variants in *SMCHD1* resulted in slightly or strongly decreased ATPase activity. Injections in Xenopus embryos with mutant and wildtype human *SMCHD1* revealed craniofacial defects only in tadpoles with arhinia/BAMS variants or very high levels of wildtype SMCHD1, but not with an FSHD2 variant. Together these experiments suggest that the arhinia/BAMS variants in *SMCHD1* might have a gain of function effect, in contrast to FSHD2 variants, which result in a loss of function effect. However, this effect has only been tested for a

small number of *SMCHD1* variants. On the contrary, Shaw et al. performed experiments with a zebrafish model which gave opposite results for arhinia/BAMS variants. Knockdown of the orthologue of *SMCHD1* in zebrafish results in craniofacial, ocular defects and a reduction of gonadotropin-releasing hormone (GnRH)-positive terminal nerve length. However, overexpression of *SMCHD1* wildtype or mutants did not result in craniofacial defects, arguing against a gain of function mechanism. In the zebrafish with a knockdown for endogenous Smchd1 the GnRH-positive terminal nerve length could be rescued with wildtype *SMCHD1*, but not with *SMCHD1* with an arhinia/BAMS or FSHD2 variant. This suggests that a loss of function effect is shared by FSHD2 and arhinia/BAMS variants in *SMCHD1*, which fits with the fact that one *SMCHD1* variant was found in both an FSHD2 and an arhinia family and with the shared D4Z4 hypomethylation. On the other hand, the specificity for missense variants in exons 3-13 in arhinia/BAMS patients, indicates that mutant SMCHD1 protein is produced in all arhinia/BAMS patients and that this domain has a specific role in the disease mechanism of arhinia/BAMS. Although the gain of function and loss of function mechanisms suggested by Gordon et al, and Shaw et al. respectively, seem to be contradictory, they might both result in the absence of SMCHD1 on chromatin, by either more active release or by less binding[40].

Besides an *SMCHD1* variant and D4Z4 hypomethylation, FSHD2 requires a permissive 4qA allele with a D4Z4 repeat array <20 units, therefore only some of the arhinia/BAMS patients might be at risk for FSHD2. Phenotypic evaluation of two arhinia/BAMS patients, which likely met all these criteria for FSHD2, was performed and at least one of them showed FSHD2 symptoms. This confirms that at least some *SMCHD1* variants can cause both FSHD2 and arhinia/BAMS, even in the same individual. However, the question remains why FSHD2 patients, especially with a missense variant in exons 3 to 13, do not show signs of arhinia/BAMS. One explanation could be that *SMCHD1* variants are insufficient by themselves to cause an arhinia/BAMS phenotype, like *SMCHD1* variants by themselves are insufficient to cause FSHD2. Therefore, it seems reasonable that there is a second locus involved in arhinia/BAMS, similar to FSHD. Although, one of the groups reports that they were unable to identify significant differences in genome-wide de-repression of repeats between arhinia/BAMS subjects and familial controls, there might be a specific repeat that is involved in arhinia/BAMS. There is large clinical variability among carriers of an *SMCHD1* variant within arhinia/BAMS families, with some carriers remaining completely unaffected. If a repeat would be involved, it is possible that size and/or haplotype differences might help to explain clinical variability, similar to what is observed in FSHD.

Shaw et al. also performed studies to identify differential gene expression between arhinia/BAMS subjects and familial controls. Some genes involved in craniofacial development were downregulated in arhinia/BAMS patients, indicating that SMCHD1

is a regulator of these genes[38]. It would be interesting to study the expression of these genes in FSHD2 patients and to compare genome wide expression profiles of arhinia/BAMS and FSHD2 patients with an *SMCHD1* variant, to assess similarities and differences between the two syndromes.

Identifying functional convergence among ICF genes

The number of ICF genes has recently been increased to four: *DNMT3B*, *ZBTB24*, *CDCA7* and *HELLS*. Almost all patients show the phenotypic characteristics of immunodeficiency, facial anomalies and centromeric instability. The phenotypes vary amongst patients, but this cannot be attributed to a specific gene defect. Besides these characteristics, ICF patients often have a delay in motor and/or speech development. Intellectual disability is also regularly seen in ICF syndrome. Almost all ICF2 patients present with intellectual disability, compared to about half of the ICF1 patients[24]. However, in our study we also identified an ICF2 patient with normal intelligence, indicating that the absence of intellectual disability does not exclude *ZBTB24* variants in ICF patients.

The high similarity in phenotypes between the different ICF patient subgroups indicates that the function of the proteins involved converge somewhere in the disease mechanism. Furthermore, although the molecular phenotype of centromeric instability and hypomethylation of repetitive regions is detected for all ICF disease entities, no clear relation with the immunodeficiency and facial anomalies has been described.

DNMT3B and HELLS interact in the establishment of CpG methylation during early development[41]. Hells recruits Dnmt3b to repeat sequences (IAP elements, minor satellites and Line1 elements) in mouse ES cells[41], and knockdown of Hells leads to reduced methylation at these repetitive regions[41; 42]. Besides its role in methylation of repetitive regions, Hells knockdown also affects methylation at non-repeat sequences sites and alters gene expression of some of the affected genes[43-46]. In addition, the transient depletion of Hells in mouse embryonic fibroblasts (MEFs) resulted in a reduction in DNA methylation at murine minor satellite repeats[47]. A transient depletion of Dnmt3b did not result in a reduction in DNA methylation at murine minor satellite repeats[47]. This indicates that Hells, but not Dnmt3b, is involved in the maintenance of DNA methylation at repetitive regions. In agreement with a role in methylation maintenance, Hells is also involved in the methylation at an imprinted region[44], binds indirectly to Dnmt1[48], and colocalizes with late replication foci[49]. On the other hand, Hells was not required for the maintenance of methylation of an episomal vector[50]. Possibly, Hells has a role in the maintenance of methylation at specific loci.

In MEFs transiently depleted for Zbtb24 and Cdca7, decreased methylation at minor satellite repeats was detected, indicating that these proteins might also have a role in the maintenance of CpG methylation[47]. However, for ZBTB24 and CDCA7 there is currently

no functional explanation for their role in maintaining (and possibly establishing) DNA methylation at repetitive DNA. The functions of ZBTB24 and CDCA7 have not been studied extensively. ZBTB24 belongs to the BTB-ZF family of proteins and might function as a transcription factor[51]. CDCA7 has a role in neoplastic transformation and hematopoietic stem cell emergence, through interactions with Myc and activation by Notch, respectively[52-54]. Recently, a functional connection between ZBTB24 and CDCA7 was discovered, since ZBTB24 controls expression of *CDCA7*[21]. In mouse embryonic stem cells siRNA knockdown of Zbtb24 reduced not only *Zbtb24* expression but also *Cdca7* expression, indicating that ZBTB24 is a transcriptional activator of *CDCA7*. A similar effect was also detected in fibroblast of ICF2 patients, which have reduced expression of both *ZBTB24* and *CDCA7*. ChIP-qPCR also confirmed binding of ZBTB24 to the *CDCA7* promoter, suggesting that ZBTB24 modulates *CDCA7* expression via a direct interaction with the *CDCA7* promoter. siRNA knockdown of the other ICF genes did not reveal any other transcriptional interactions between the ICF genes[21].

Although two links between two ICF genes have been made, the pathological mechanism involved in ICF syndrome has not been unraveled. Most functional studies have thus far been performed in ICF1 patient cell lines, which revealed hypomethylated regions, differentially expressed genes, alternative transcription start sites usage, alternative splicing, deregulated antisense transcription, changes in nuclear architecture and altered replication timing[55-67]. Although these studies did not deduce a complete disease mechanism, it would be valuable to study these alterations in all ICF subtypes to identify common changes, which likely contribute to the development of ICF.

Besides studies in human cell lines, mouse models might also be valuable to study the functions of ICF genes. However, knockout mouse models of *Dnmt3b*, *Hells* and *Zbtb24* were found to be embryonic lethal[21; 68; 69]. For *Hells* and *Zbtb24* this is in contrast to the situation in humans, where homozygous nonsense alleles are viable[24; 47]. For *CDCA7* only homozygous missense variants have been identified so far in ICF3 patients[47], suggesting that it might be more relevant to create missense variants in *Cdca7* to establish a ICF3 mouse model, then to make a knockout. Indeed, while homozygous knock out of Dnmt3b is embryonic lethal, mouse models with homozygous or compound heterozygous ICF missense variants in Dnmt3b are viable and show some characteristics of ICF syndrome such as facial anomalies and hypomethylation of germ line genes and repetitive regions[34; 70; 71]. A B-cell defect is absent in these mice, however, they show increased T-cell apoptosis and reduced numbers of total thymocytes[34]. About 20% of ICF patients show opportunistic infections, which suggest that there might also be T-cell dysfunction in ICF patients. However, the B-cell dysfunction is more pronounced in ICF syndrome[24].

Because of the lethality in mice from the loss of Dnmt3b, Hells and Zbtb24, an alternative approach to study their role in immune cell development would make use of immune compromised recipient mice transplanted with fetal liver cells of embryo's with a knockout of one of these genes. Such an approach was already performed for Hells, where fetal liver cell suspensions from Hells knockout embryos were injected in Rag2 knockout mice, which lack lymphoid cells. The resulting chimeras showed a reduction in peripheral T lymphocytes and a reduction in mature B-cells [69]. To understand the disease mechanism of ICF, it will be essential to study the role of all four ICF genes in immune cell development in more detail.

The role of DUX4 in development and FSHD pathology

DUX4 has been identified as a transcription factor that becomes inappropriately expressed in FSHD myocytes with toxic consequences to the muscle. Although both overexpression and endogenous expression of *DUX4* lead to apoptosis, it is not entirely clear which mechanisms contribute to DUX4 induced cell death and how this results in the development of a muscular dystrophy[72; 73].

DUX4 was first identified as a toxic protein in by Kowaljow et al., who reported that DUX4 activates caspase 3/7 activity, indicative of apoptosis[73]. Since then various pathways have been identified that might contribute to the induction of apoptosis after DUX4 expression. DUX4 induced apoptosis was shown to be dependent on P53 in HEK293T cells transfected with DUX4 as well as in skeletal muscle of mice injected with DUX4 AAV virus[74]. However, in C2C12 mouse myoblasts and a mouse model with a doxycycline-inducible *DUX4* transgene, DUX4 induced pathogenicity was independent of P53[75]. In addition, in human myoblasts and a RD rhabdomyosarcoma cell line DUX4 efficiently induced apoptosis in the absence of functional P53[76]. Furthermore, doxycycline induction of *DUX4* expression increases MYC protein levels by stabilization of MYC mRNA and activates components of the MYC-mediated apoptotic pathway in a RD rhabdomyosarcoma cell line and immortalized human myoblasts. Both cell lines also showed an accumulation of nuclear double stranded RNA (dsRNA) foci. The accumulation of dsRNAs might induce apoptosis via a pro-apoptotic dsRNA-sensing innate immune response. In addition, the accumulation of dsRNAs is associated with the formation of EIF4A3 protein aggregates[76]. Since EIF4A3 plays a role in mRNA surveillance and decay, DUX4 might inhibit nonsense mediated decay (NMD) via EIF4A3 aggregation, amongst other mechanisms such as the degradation of UPF1[77]. The inhibition of NMD might result in an accumulation of alternative transcripts containing premature stop codons and aberrant spliced exons, which is detected in DUX4 expressing myoblasts[72; 77]. In addition, DUX4 also binds and activates transcription of transposable elements such as mammalian apparent LTR-retrotransposons (MaLRs) and

endogenous retrovirus (ERVL and ERVK) elements[78; 79]. At some of these elements DUX4 creates alternative transcription start sites, therefore creating novel transcripts, which are expressed in FSHD muscle cells but not in control muscle cells[79]. DUX4 also induces genes that are involved in muscle atrophy in FSHD myoblasts and elevates levels of reactive oxygen species and DNA damage, which might affect myogenic differentiation of FSHD myoblasts[80; 81]. Furthermore, *DUX4* expression in myotubes also alters nuclear body structure (PML bodies and SC35 speckles) and induces coaggregation of two RNA/DNA handling protein (FUS and TDP43)[82; 83]. All these pathways might play a role in the FSHD disease mechanism, however their exact contributions to FSHD pathology remain to be investigated.

In addition to the toxic effects of DUX4 in FSHD, the evolutionary conservation of DUX4 suggests that there is an important normal function for DUX4. Studies of early human embryos revealed that *DUX4* is specifically expressed at the 4-cell stage, which corresponds to early embryonic genome activation, and activates cleavage specific genes and repetitive elements[84; 85]. Mouse *Dux* is also expressed during the 2-cell stage, which corresponds to the onset of embryonic genome activation in mouse[84]. Mouse Dux also activates cleavage specific genes and repetitive elements, while embryonic genome activation is disrupted in the absence of mouse Dux[84-86]. Therefore, human DUX4 and mouse Dux have a conserved role in activating cleavage specific genes and repetitive elements[84-86], although they show only modest amino acid sequence conservation[87]. These studies indicate a normal function for DUX4 in early embryonic development, besides its role in FSHD pathology. However it remains to be investigated what activates and subsequently silences *DUX4* during early cleavage stage, and how this is relevant to *DUX4* expression in FSHD muscle.

The FSHD disease continuum

FSHD1 and FSHD2 have been described as two separate disease entities, with a shared disease mechanism of D4Z4 chromatin relaxation and *DUX4* derepression. However, older and recent studies emphasize the importance of the size of the D4Z4 repeat array in both FSHD1 and FSHD2, and indicate that there is a repeat array size dependent disease severity in both forms of the disease[2; 88-91]. In addition, multiple families with both an FSHD1-sized allele of 7-10 units and an *SMCHD1* or *DNMT3B* variant have been identified during the last years[2; 3; 5]. Permissive D4Z4 repeat arrays of 7-10 units are also found in 1-3% of the control population[6; 7]. This suggests that, rather than two separate disease entities, FSHD1 and FSHD2 are opposite extremes of a disease continuum. With very short D4Z4 repeat arrays of 1-6 units the contraction by itself is largely sufficient for *DUX4* derepression and disease presentation in somatic cells. With longer repeat arrays the ability to repress *DUX4* depends on a combination of the repeat array size

and the functionality or level of epigenetic modifiers of D4Z4 such as SMCHD1 and DNMT3B. Variability in disease penetrance between families with *SMCHD1* variants is dependent on the type of variant and the repeat length, while variability within families often depends on the repeat array size. Not all variability can be explained by repeat array size and variants in *SMCHD1* or *DNMT3B*. Thus, there are likely other genetic risk factors including *in cis* (e.g. differences in the sequence of the FSHD1 locus) and *in trans* modifiers (other epigenetic modifiers) that influence disease severity, in addition to other modifiers such as those that affect the DUX4 downstream pathways. The involvement of additional genetic risk factors may also explain why variants in *SMCHD1* and *DNMT3B* are associated with the discordant phenotypes: FSHD, ICF1 or BAMS. In isolation, damaging variants in one allele of these chromatin modifiers is insufficient to cause disease, as evidenced by the absence of disease phenotypes in carriers of an *DNMT3B* variant, or carriers of an *SMCHD1* variant in the absence of a permissive DUX4 allele. However, when combined with other genetic risk factors, such as a short D4Z4 repeat array size, the presence of a DUX4 PAS, and/or yet unknown risk factors, these variants are lowering the threshold for disease presentation.

Concluding remarks

In this thesis, the mutation spectrum of FSHD and ICF, two diseases with very distinct phenotypes that share a common feature of hypomethylation of repetitive DNA, has been expanded. The *SMCHD1* mutation spectrum in FSHD2 has been expanded with the discovery of exonic *SMCHD1* variants, intronic *SMCHD1* variants, and whole *SMCHD1* gene deletions. In addition, we identified heterozygous variants in a new FSHD2 gene, *DNMT3B*, in two FSHD2 families. For ICF syndrome we expanded the mutation spectrum in the two most common ICF genes, *DNMT3B* and *ZBTB24*. All of the aforementioned findings enabled us to identify the causative variants in previously unexplainable cases of FSHD and in new ICF patients, expanding the knowledge base for these diseases and setting the stage for further studies.

REFERENCES

1. Lemmers, R.J., Tawil, R., Petek, L.M., Balog, J., Block, G.J., Santen, G.W., Amell, A.M., van der Vliet, P.J., Almomani, R., Straasheijm, K.R., et al. (2012). Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nature genetics 44, 1370-1374.
2. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2014). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Human molecular genetics.
3. Larsen, M., Rost, S., El Hajj, N., Ferbert, A., Deschauer, M., Walter, M.C., Schoser, B., Tacik, P., Kress, W., and Muller, C.R. (2014). Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. European journal of human genetics : EJHG.
4. Blewitt, M.E., Gendrel, A.V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., et al. (2008). SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nature genetics 40, 663-669.
5. Sacconi, S., Lemmers, R.J., Balog, J., van der Vliet, P.J., Lahaut, P., van Nieuwenhuizen, M.P., Straasheijm, K.R., Debipersad, R.D., Vos-Versteeg, M., Salviati, L., et al. (2013). The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. American journal of human genetics 93, 744-751.
6. Lemmers, R.J., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.M., de Knijff, P., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2007). Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. American journal of human genetics 81, 884-894.
7. Scionti, I., Fabbri, G., Fiorillo, C., Ricci, G., Greco, F., D'Amico, R., Termanini, A., Vercelli, L., Tomelleri, G., Cao, M., et al. (2012). Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. Journal of medical genetics 49, 171-178.
8. Gendrel, A.V., Apedaile, A., Coker, H., Termanis, A., Zvetkova, I., Godwin, J., Tang, Y.A., Huntley, D., Montana, G., Taylor, S., et al. (2012). Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Developmental cell 23, 265-279.
9. Gendrel, A.V., Tang, Y.A., Suzuki, M., Godwin, J., Nesterova, T.B., Greally, J.M., Heard, E., and Brockdorff, N. (2013). Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Molecular and cellular biology 33, 3150-3165.

10. Mould, A.W., Pang, Z., Pakusch, M., Tonks, I.D., Stark, M., Carrie, D., Mukhopadhyay, P., Seidel, A., Ellis, J.J., Deakin, J., et al. (2013). Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. Epigenetics & chromatin 6, 19.

11. Liu, R., Chen, K., Jansz, N., Blewitt, M.E., and Ritchie, M.E. (2016). Transcriptional profiling of the epigenetic regulator Smchd1. Genom Data 7, 144-147.

12. Massah, S., Hollebakken, R., Labrecque, M.P., Kolybaba, A.M., Beischlag, T.V., and Prefontaine, G.G. (2014). Epigenetic characterization of the growth hormone gene identifies SmcHD1 as a regulator of autosomal gene clusters. PloS one 9, e97535.

13. Mason, A.G., Slieker, R.C., Balog, J., Lemmers, R., Wong, C.J., Yao, Z., Lim, J.W., Filippova, G.N., Ne, E., Tawil, R., et al. (2017). SMCHD1 regulates a limited set of gene clusters on autosomal chromosomes. Skeletal muscle 7, 12.

14. Chen, K., Hu, J., Moore, D.L., Liu, R., Kessans, S.A., Breslin, K., Lucet, I.S., Keniry, A., Leong, H.S., Parish, C.L., et al. (2015). Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. Proceedings of the National Academy of Sciences of the United States of America 112, E3535-3544.

15. Yao, Z., Snider, L., Balog, J., Lemmers, R.J., Van Der Maarel, S.M., Tawil, R., and Tapscott, S.J. (2014). DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. Human molecular genetics 23, 5342-5352.

16. Mayes, M.B., Morgan, T., Winston, J., Buxton, D.S., Kamat, M.A., Smith, D., Williams, M., Martin, R.L., Kleinjan, D.A., Cooper, D.N., et al. (2015). Remotely acting SMCHD1 gene regulatory elements: in silico prediction and identification of potential regulatory variants in patients with FSHD. Hum Genomics 9, 25.

17. Sacconi, S., Camano, P., de Greef, J.C., Lemmers, R.J., Salviati, L., Boileau, P., Lopez de Munain Arregui, A., van der Maarel, S.M., and Desnuelle, C. (2012). Patients with a phenotype consistent with facioscapulohumeral muscular dystrophy display genetic and epigenetic heterogeneity. Journal of medical genetics 49, 41-46.

18. Kim, E.Y., Page, P., Dellefave-Castillo, L.M., McNally, E.M., and Wyatt, E.J. (2016). Direct reprogramming of urine-derived cells with inducible MyoD for modeling human muscle disease. Skeletal muscle 6, 32.

19. Jones, T.I., Himeda, C.L., Perez, D.P., and Jones, P.L. (2017). Large family cohorts of lymphoblastoid cells provide a new cellular model for investigating facioscapulohumeral muscular dystrophy. Neuromuscular disorders : NMD 27, 221-238.

20. Das, S., and Chadwick, B.P. (2016). Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. PloS one 11, e0160022.

21. Wu, H., Thijssen, P.E., de Klerk, E., Vonk, K.K., Wang, J., den Hamer, B., Aytekin, C., van der Maarel, S.M., and Daxinger, L. (2016). Converging disease genes in ICF syndrome: ZBTB24 controls expression of CDCA7 in mammals. Human molecular genetics.

22. Liang, J., Yan, R., Chen, G., Feng, J., Wu, W.W., Ren, W., Zhu, C., Zhao, Y., Gao, X.M., and Wang, J. (2016). Downregulation of ZBTB24 hampers the G0/1- to S-phase cell-cycle transition via upregulating the expression of IRF-4 in human B cells. Genes Immun.

23. Nitta, H., Unoki, M., Ichiyanagi, K., Kosho, T., Shigemura, T., Takahashi, H., Velasco, G., Francastel, C., Picard, C., Kubota, T., et al. (2013). Three novel ZBTB24 mutations identified in Japanese and Cape Verdean type 2 ICF syndrome patients. J Hum Genet 58, 455-460.

24. Weemaes, C.M., van Tol, M.J., Wang, J., van Ostaijen-ten Dam, M.M., van Eggermond, M.C., Thijssen, P.E., Aytekin, C., Brunetti-Pierri, N., van der Burg, M., Graham Davies, E., et al. (2013). Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. European journal of human genetics : EJHG 21, 1219-1225.

25. Kutlug, S., Ogur, G., Yilmaz, A., Thijssen, P.E., Abur, U., and Yildiran, A. (2016). Vesicourethral reflux-induced renal failure in a patient with ICF syndrome due to a novel DNMT3B mutation. Am J Med Genet A 170, 3253-3257.

26. Vivante, A., Hwang, D.Y., Kohl, S., Chen, J., Shril, S., Schulz, J., van der Ven, A., Daouk, G., Soliman, N.A., Kumar, A.S., et al. (2017). Exome Sequencing Discerns Syndromes in Patients from Consanguineous Families with Congenital Anomalies of the Kidneys and Urinary Tract. J Am Soc Nephrol 28, 69-75.

27. Tatton-Brown, K., Seal, S., Ruark, E., Harmer, J., Ramsay, E., Del Vecchio Duarte, S., Zachariou, A., Hanks, S., O'Brien, E., Aksglaede, L., et al. (2014). Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. Nature genetics 46, 385-388.

28. Argentaro, A., Yang, J.C., Chapman, L., Kowalczyk, M.S., Gibbons, R.J., Higgs, D.R., Neuhaus, D., and Rhodes, D. (2007). Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein ATRX. Proceedings of the National Academy of Sciences of the United States of America 104, 11939-11944.

29. Guo, X., Wang, L., Li, J., Ding, Z., Xiao, J., Yin, X., He, S., Shi, P., Dong, L., Li, G., et al. (2015). Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. Nature 517, 640-644.

30. Noh, K.M., Allis, C.D., and Li, H. (2015). Reading between the Lines: "ADD"-ing Histone and DNA Methylation Marks toward a New Epigenetic "Sum". ACS chemical biology.

31. Van Emburgh, B.O., and Robertson, K.D. (2011). Modulation of Dnmt3b function in vitro by interactions with Dnmt3L, Dnmt3a and Dnmt3b splice variants. Nucleic acids research 39, 4984-5002.

32. Gowher, H., and Jeltsch, A. (2002). Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. J Biol Chem 277, 20409-20414.

33. Xie, Z.H., Huang, Y.N., Chen, Z.X., Riggs, A.D., Ding, J.P., Gowher, H., Jeltsch, A., Sasaki, H., Hata, K., and Xu, G.L. (2006). Mutations in DNA methyltransferase DNMT3B in ICF syndrome affect its regulation by DNMT3L. Human molecular genetics 15, 1375-1385.

34. Ueda, Y., Okano, M., Williams, C., Chen, T., Georgopoulos, K., and Li, E. (2006). Roles for Dnmt3b in mammalian development: a mouse model for the ICF syndrome. Development 133, 1183-1192.

35. Gaillard, M.C., Roche, S., Dion, C., Tasmadjian, A., Bouget, G., Salort-Campana, E., Vovan, C., Chaix, C., Broucqsault, N., Morere, J., et al. (2014). Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. Neurology 83, 733-742.

36. Jones, T.I., Yan, C., Sapp, P.C., McKenna-Yasek, D., Kang, P.B., Quinn, C., Salameh, J.S., King, O.D., and Jones, P.L. (2014). Identifying diagnostic DNA methylation profiles for facioscapulohumeral muscular dystrophy in blood and saliva using bisulfite sequencing. Clinical epigenetics 6, 23.

37. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., et al. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS genetics 6, e1001181.

38. Shaw, N.D., Brand, H., Kupchinsky, Z.A., Bengani, H., Plummer, L., Jones, T.I., Erdin, S., Williamson, K.A., Rainger, J., Stortchevoi, A., et al. (2017). SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. Nature genetics.

39. Gordon, C.T., Xue, S., Yigit, G., Filali, H., Chen, K., Rosin, N., Yoshiura, K.I., Oufadem, M., Beck, T.J., McGowan, R., et al. (2017). De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. Nature genetics.

40. Wilkie, A.O. (2017). Many faces of SMCHD1. Nature genetics 49, 176-178.

41. Ren, J., Briones, V., Barbour, S., Yu, W., Han, Y., Terashima, M., and Muegge, K. (2015). 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 research 43, 1444-1455.

42. Yu, W., McIntosh, C., Lister, R., Zhu, I., Han, Y., Ren, J., Landsman, D., Lee, E., Briones, V., Terashima, M., et al. (2014). Genome-wide DNA methylation patterns in LSH mutant reveals de-repression of repeat elements and redundant epigenetic silencing pathways. Genome Res 24, 1613-1623.

43. Xi, S., Zhu, H., Xu, H., Schmidtmann, A., Geiman, T.M., and Muegge, K. (2007). Lsh controls Hox gene silencing during development. Proceedings of the National Academy of Sciences of the United States of America 104, 14366-14371.

44. Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes Dev 15, 2940-2944.

45. Tao, Y., Xi, S., Shan, J., Maunakea, A., Che, A., Briones, V., Lee, E.Y., Geiman, T., Huang, J., Stephens, R., et al. (2011). Lsh, chromatin remodeling family member, modulates genome-wide cytosine methylation patterns at nonrepeat sequences. Proceedings of the National Academy of Sciences of the United States of America 108, 5626-5631.

46. Myant, K., Termanis, A., Sundaram, A.Y., Boe, T., Li, C., Merusi, C., Burrage, J., de Las Heras, J.I., and Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. Genome Res 21, 83-94.

47. Thijssen, P.E., Ito, Y., Grillo, G., Wang, J., Velasco, G., Nitta, H., Unoki, M., Yoshihara, M., Suyama, M., Sun, Y., et al. (2015). Mutations in CDCA7 and HELLS cause immunodeficiency-centromeric instability-facial anomalies syndrome. Nat Commun 6, 7870.

48. Myant, K., and Stancheva, I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. Molecular and cellular biology 28, 215-226.

49. Yan, Q., Cho, E., Lockett, S., and Muegge, K. (2003). Association of Lsh, a regulator of DNA methylation, with pericentromeric heterochromatin is dependent on intact heterochromatin. Molecular and cellular biology 23, 8416-8428.

50. Zhu, H., Geiman, T.M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E., and Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. The EMBO journal 25, 335-345.

51. Maeda, T. (2016). Regulation of hematopoietic development by ZBTB transcription factors. Int J Hematol 104, 310-323.

52. Prescott, J.E., Osthus, R.C., Lee, L.A., Lewis, B.C., Shim, H., Barrett, J.F., Guo, Q., Hawkins, A.L., Griffin, C.A., and Dang, C.V. (2001). A novel c-Myc-responsive gene, JPO1, participates in neoplastic transformation. J Biol Chem 276, 48276-48284.

53. Gill, R.M., Gabor, T.V., Couzens, A.L., and Scheid, M.P. (2013). The MYC-associated protein CDCA7 is phosphorylated by AKT to regulate MYC-dependent apoptosis and transformation. Molecular and cellular biology 33, 498-513.

54. Guiu, J., Bergen, D.J., De Pater, E., Islam, A.B., Ayllon, V., Gama-Norton, L., Ruiz-Herguido, C., Gonzalez, J., Lopez-Bigas, N., Menendez, P., et al. (2014). Identification of Cdca7 as a novel Notch transcriptional target involved in hematopoietic stem cell emergence. J Exp Med 211, 2411-2423.

55. Simo-Riudalbas, L., Diaz-Lagares, A., Gatto, S., Gagliardi, M., Crujeiras, A.B., Matarazzo, M.R., Esteller, M., and Sandoval, J. (2015). Genome-Wide DNA Methylation Analysis Identifies Novel Hypomethylated Non-Pericentromeric Genes with Potential Clinical Implications in ICF Syndrome. PloS one 10, e0132517.

56. Jin, B., Tao, Q., Peng, J., Soo, H.M., Wu, W., Ying, J., Fields, C.R., Delmas, A.L., Liu, X., Qiu, J., et al. (2008). DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. Human molecular genetics 17, 690-709.

57. Velasco, G., Walton, E.L., Sterlin, D., Hedouin, S., Nitta, H., Ito, Y., Fouyssac, F., Megarbane, A., Sasaki, H., Picard, C., et al. (2014). 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.

58. Huang, K., Wu, Z., Liu, Z., Hu, G., Yu, J., Chang, K.H., Kim, K.P., Le, T., Faull, K.F., Rao, N., et al. (2014). Selective demethylation and altered gene expression are associated with ICF syndrome in human-induced pluripotent stem cells and mesenchymal stem cells. Human molecular genetics 23, 6448-6457.

59. Heyn, H., Vidal, E., Sayols, S., Sanchez-Mut, J.V., Moran, S., Medina, I., Sandoval, J., Simo-Riudalbas, L., Szczesna, K., Huertas, D., et al. (2012). Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. Epigenetics : official journal of the DNA Methylation Society 7, 542-550.

60. Sagie, S., Ellran, E., Katzir, H., Shaked, R., Yehezkel, S., Laevsky, I., Ghanayim, A., Geiger, D., Tzukerman, M., and Selig, S. (2014). Induced pluripotent stem cells as a model for telomeric abnormalities in ICF type I syndrome. Human molecular genetics 23, 3629-3640.

61. Jefferson, A., Colella, S., Moralli, D., Wilson, N., Yusuf, M., Gimelli, G., Ragoussis, J., and Volpi, E.V. (2010). Altered intra-nuclear organisation of heterochromatin and genes in ICF syndrome. PloS one 5, e11364.

62. Matarazzo, M.R., Boyle, S., D'Esposito, M., and Bickmore, W.A. (2007). Chromosome territory reorganization in a human disease with altered DNA methylation. Proceedings of the National Academy of Sciences of the United States of America 104, 16546-16551.

63. Hassan, K.M., Norwood, T., Gimelli, G., Gartler, S.M., and Hansen, R.S. (2001). Satellite 2 methylation patterns in normal and ICF syndrome cells and association of hypomethylation with advanced replication. Hum Genet 109, 452-462.

64. Yehezkel, S., Segev, Y., Viegas-Pequignot, E., Skorecki, K., and Selig, S. (2008). Hypomethylation of subtelomeric regions in ICF syndrome is associated with abnormally short telomeres and enhanced transcription from telomeric regions. Human molecular genetics 17, 2776-2789.

65. Hansen, R.S., Stoger, R., Wijmenga, C., Stanek, A.M., Canfield, T.K., Luo, P., Matarazzo, M.R., D'Esposito, M., Feil, R., Gimelli, G., et al. (2000). Escape from gene silencing in ICF syndrome: evidence for advanced replication time as a major determinant. Human molecular genetics 9, 2575-2587.

66. Lana, E., Megarbane, A., Tourriere, H., Sarda, P., Lefranc, G., Claustres, M., and De Sario, A. (2012). DNA replication is altered in Immunodeficiency Centromeric instability Facial anomalies (ICF) cells carrying DNMT3B mutations. European journal of human genetics : EJHG 20, 1044-1050.

67. Gatto, S., Gagliardi, M., Franzese, M., Leppert, S., Papa, M., Cammisa, M., Grillo, G., Velasco, G., Francastel, C., Toubiana, S., et al. (2017). ICF-specific DNMT3B dysfunction interferes with intragenic regulation of mRNA transcription and alternative splicing. Nucleic acids research 45, 5739-5756.

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

69. Geiman, T.M., and Muegge, K. (2000). Lsh, an SNF2/helicase family member, is required for proliferation of mature T lymphocytes. Proceedings of the National Academy of Sciences of the United States of America 97, 4772-4777.

70. Velasco, G., Hube, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Pequignot, E., and Francastel, C. (2010). Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proceedings of the National Academy of Sciences of the United States of America 107, 9281-9286.

71. Youngson, N.A., Epp, T., Roberts, A.R., Daxinger, L., Ashe, A., Huang, E., Lester, K.L., Harten, S.K., Kay, G.F., Cox, T., et al. (2013). No evidence for cumulative effects in a Dnmt3b hypomorph across multiple generations. Mammalian genome : official journal of the International Mammalian Genome Society 24, 206-217.

72. Rickard, A.M., Petek, L.M., and Miller, D.G. (2015). Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Human molecular genetics 24, 5901-5914.

73. Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O., et al. (2007). The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscular disorders : NMD 17, 611-623.

74. Wallace, L.M., Garwick, S.E., Mei, W., Belayew, A., Coppee, F., Ladner, K.J., Guttridge, D., Yang, J., and Harper, S.Q. (2011). DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. Annals of neurology 69, 540-552.

75. Bosnakovski, D., Gearhart, M.D., Toso, E.A., Recht, O.O., Cucak, A., Jain, A.K., Barton, M.C., and Kyba, M. (2017). p53-independent DUX4 pathology. Dis Model Mech.

76. Shadle, S.C., Zhong, J.W., Campbell, A.E., Conerly, M.L., Jagannathan, S., Wong, C.J., Morello, T.D., van der Maarel, S.M., and Tapscott, S.J. (2017). DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. PLoS genetics 13, e1006658.

77. Feng, Q., Snider, L., Jagannathan, S., Tawil, R., van der Maarel, S.M., Tapscott, S.J., and Bradley, R.K. (2015). A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. Elife 4.

78. Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., et al. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Developmental cell 22, 38-51.

79. Young, J.M., Whiddon, J.L., Yao, Z., Kasinathan, B., Snider, L., Geng, L.N., Balog, J., Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2013). DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. PLoS genetics 9, e1003947.

80. Vanderplanck, C., Ansseau, E., Charron, S., Stricwant, N., Tassin, A., Laoudj-Chenivesse, D., Wilton, S.D., Coppee, F., and Belayew, A. (2011). The FSHD atrophic myotube phenotype is caused by DUX4 expression. PloS one 6, e26820.

81. Dmitriev, P., Bou Saada, Y., Dib, C., Ansseau, E., Barat, A., Hamade, A., Dessen, P., Robert, T., Lazar, V., Louzada, R.A., et al. (2016). DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. Free Radic Biol Med 99, 244-258.

82. Homma, S., Beermann, M.L., Boyce, F.M., and Miller, J.B. (2015). Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. Ann Clin Transl Neurol 2, 151-166.

83. Homma, S., Beermann, M.L., Yu, B., Boyce, F.M., and Miller, J.B. (2016). Nuclear bodies reorganize during myogenesis in vitro and are differentially disrupted by expression of FSHD-associated DUX4. Skeletal muscle 6, 42.

84. Hendrickson, P.G., Dorais, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nature genetics.

85. De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nature genetics.

86. Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. Nature genetics.

87. Eidahl, J.O., Giesige, C.R., Domire, J.S., Wallace, L.M., Fowler, A.M., Guckes, S.M., Garwick-Coppens, S.E., Labhart, P., and Harper, S.Q. (2016). Mouse Dux is myotoxic and shares partial functional homology with its human paralog DUX4. Human molecular genetics 25, 4577-4589.

88. de Greef, J.C., Lemmers, R.J., Camano, P., Day, J.W., Sacconi, S., Dunand, M., van Engelen, B.G., Kiuru-Enari, S., Padberg, G.W., Rosa, A.L., et al. (2010). Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology 75, 1548-1554.

89. Lunt, P.W., Jardine, P.E., Koch, M.C., Maynard, J., Osborn, M., Williams, M., Harper, P.S., and Upadhyaya, M. (1995). Correlation between fragment size at 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). Human molecular genetics 4, 951-958.

90. Lin, F., Wang, Z.Q., Lin, M.T., Murong, S.X., and Wang, N. (2015). New Insights into Genotype-phenotype Correlations in Chinese Facioscapulohumeral Muscular Dystrophy: A Retrospective Analysis of 178 Patients. Chin Med J (Engl) 128, 1707-1713.

91. Park, H.J., Hong, J.M., Lee, J.H., Lee, H.S., Shin, H.Y., Kim, S.M., Ki, C.S., Lee, J.H., and Choi, Y.C. (2015). Low D4Z4 copy number and gender difference in Korean patients with facioscapulohumeral muscular dystrophy type 1. Neuromuscular disorders : NMD 25, 859-864.

A

APPENDIX

ENGLISH SUMMARY

Epigenetic regulation by chromatin modifiers needs to be tightly controlled to ensure correct patterns of gene expression. Misregulation of epigenetic processes can lead to disease, as is the case in FSHD and ICF syndrome. FSHD and ICF share a common hallmark of hypomethylation of repetitive DNA. FSHD is a muscular dystrophy mainly affecting the facial, shoulder girdle, and upper arm muscles. It is characterized by partial chromatin relaxation of the *DUX4* encoding D4Z4 macrosatellite repeat located on chromosome 4, and transcriptional derepression of the DUX4 transcription factor in skeletal muscle. D4Z4 chromatin relaxation is consistently marked by CpG hypomethylation of the D4Z4 repeat, including the *DUX4* promoter. In the most common form, FSHD1, this chromatin relaxation is the result of a repeat contraction of the D4Z4 repeat to 1-10 units (normal range 8-100 units). In the rare form of FSHD, FSHD2, D4Z4 chromatin relaxation occurs on all D4Z4 arrays and is most often caused by variants in the D4Z4 chromatin repressor SMCHD1.

ICF is an immunodeficiency syndrome mainly characterized by hypo- or agammaglobulinemia in the presence of B cells, developmental delay and facial anomalies. Compared to FSHD, ICF patients show more genome-wide hypomethylation at repetitive elements and this results in, amongst others, centromeric instability, which is the cytogenetic hallmark of ICF. Recessive variants in four genes have been identified to cause ICF syndrome: *DNMT3B* (ICF1), *ZBTB24* (ICF2), *CDCA7* (ICF3) and *HELLS* (ICF4). In this thesis the variant spectrum for chromatin modifiers involved in FSHD and ICF has been expanded.

In chapter 2 three FSHD2 families are described in which two potentially damaging variants in *SMCHD1* were identified in the proband. For each family we investigated whether these variants were located *in cis* or *in trans* as well as their contribution to D4Z4 hypomethylation. In the first family, a splice site variant and a nonsense variants *in cis* were identified in exon 21. The splice variant results in skipping of exon 21, which neutralizes the nonsense variant in exon 21. However, skipping of exon 21 is predicted to disrupt the open reading frame by a premature stop codon in exon 22. In the second family, the proband carries a missense variant in exon 28 and a splice site variant in exon 25. Although the missense variant in exon 28 was predicted to affect function, no D4Z4 hypomethylation was detected in the mother of the proband who also carries this variant. This suggests that only the splice site variant in exon 25, which seems to be in an *SMCHD1* mutation hotspot, affects SMCHD1 function and contributes to disease. This family also emphasizes the predictive value of D4Z4 methylation analysis in establishing the functional consequences of *SMCHD1* variants. In the third family, the proband carries two *SMCHD1* missense variants in exon 24 and exon 45 *in trans* and shows severe D4Z4 hypomethylation. D4Z4 hypomethylation was also detected in

the father and the mother of the proband, who carry the missense variant in exon 24 and exon 45, respectively. This revealed that both missense variants contribute to D4Z4 hypomethylation independently. The combination of these missense variants has an additive effect on D4Z4 hypomethylation, since this was more strongly reduced in the proband and his sister who carry both missense variants. Although the mother and the proband both carry a permissive D4Z4 allele of 27 repeat units, only the proband with two *SMCHD1* missense variants is affected. This suggests that these variants also have an additive effect on disease presentation. Interestingly, this family also shows that having two *SMCHD1* missense variants *in trans* is compatible with life in males and females.

In chapter 3 whole *SMCHD1* gene deletions were identified in two FSHD2 families, which were previously tested negative for exonic *SMCHD1* variants. Some neighbouring genes were also deleted in these families, however these genes have not been reported to be sensitive to hemizygosity. This study emphasizes that hemizygosity of *SMCHD1* can also cause FSHD2, and results in a haploinsufficiency mechanism as is also expected for the open reading frame disrupting variants identified in one third of the FSHD2 patients. *SMCHD1* hemizygosity is also present in most patients with 18p-deletion syndrome, however muscular dystrophy has not been reported in these patients. D4Z4 methylation analysis in 18p-deletion patients revealed a similar degree of hypomethylation in most of the 18p-deletion patients as is seen in FSHD2 patients. Based on the prevalence of permissive 4qA alleles with a D4Z4 repeat size in the FSHD2 range (11-16 units) in the control population, it is expected that approximately 1:8 18p-deletion patients with SMCHD1 hemizygosity might be at risk of developing FSHD.

In chapter 4 we described a (deep) intronic variant in two FSHD families. In the first family an intronic variant located 15 base pairs proximal to exon 14 was identified (c.1843-15A>G), which creates a splice acceptor site. The variant results in the inclusion of the distal 14 nucleotides of intron 13 into the transcript and this is predicted to disrupt the open reading frame by a premature stop codon in exon 14. The variant was identified in the proband and his sister, and segregates with D4Z4 hypomethylation. In the second family a deep intronic variant in *SMCHD1* was identified in intron 34 (c.4347-236A>G). This variant creates a splice acceptor site, while a cryptic splice donor site is predicted at position c.4347-183 of *SMCHD1*. Together this results in exonisation of 53 nucleotides from intron 34 into the transcript, which is predicted to disrupt the open reading frame by a premature stop codon in exon 35. In this family this deep intronic variant segregates with D4Z4 hypomethylation and modifies disease severity in family members with a permissive 7 units D4Z4 repeat array.

D4Z4 hypomethylation cannot always be explained by *SMCHD1* variants in all FSHD2 families. In chapter 5 we described the identification of heterozygous missense variants in *DNMT3B* in two of these families. These *DNMT3B* variants segregate with D4Z4

hypomethylation and increased penetrance of FSHD. We identified *DUX4* expression in MyoD transduced fibroblasts from an affected individual from one of these families with a permissive D4Z4 repeat array of 13 units, but not in an unaffected individual from the other family with a permissive D4Z4 repeat array of 44 units. This suggests that heterozygous *DNMT3B* variants can only derepress *DUX4* in combination with smaller D4Z4 repeat arrays.

Since homozygous or compound heterozygous *DNMT3B* variants cause autosomal recessive ICF1 syndrome, we made a clinical and epigenetic comparison between some ICF1 patients and FSHD2 families with a *DNMT3B* variant. D4Z4 hypomethylation is shared between FSHD2 patients and ICF1 patients, and is also found in some parents of ICF1 patients who carry a heterozygous *DNMT3B* variant, like our FSHD2 patients. In addition, DUX4 expression was detected in myoblasts from one ICF1 patient, as well as in MyoD transduced fibroblasts from two other ICF1 patients. Some epigenetic characteristics of ICF were also detected in one *DNMT3B* variant carrier from these FSHD2 families. Although some epigenetic characteristics are shared, muscular dystrophy has not been reported in ICF1 patients or their parents and no immunodeficiency was detected in FSHD2 families, in concordance with parents of ICF1 patients. We propose that the effect of *DNMT3B* variants on *DUX4* expression and disease presentation, like for *SMCHD1*, depends on several aspects associated with the FSHD1 locus including D4Z4 repeat array size, and the presence of a polymorphic *DUX4* polyadenylation signal. This study suggests that multiple factors are involved in the epigenetic state at D4Z4 and the regulation of *DUX4* in muscle cells.

Finally, in chapter 6 we expand the spectrum of ICF1 and ICF2 variants. In seven ICF1 patients from four different families we identified in total six missense variants in the catalytic domain of DNMT3B. These variants were not previously described in ICF1 patients. In addition, we describe five ICF2 cases from five different families. Four of these ICF2 patients carry homozygous nonsense variants in *ZBTB24*, including one which was not previously identified in ICF2 patients. The fifth ICF2 patient carries a homozygous deletion on chromosome 6 which includes *ZBTB24*. This observation shows that complete absence of ZBTB24 protein is compatible with life and supports the hypothesis that most ICF2 patients suffer from a functional loss of ZBTB24. With these 12 ICF patients included, a total of 77 genetically confirmed patients has been reported. 56% of the ICF patients carry *DNMT3B* variants, 31% carry *ZBTB24* variants, 7% carry *CDCA7* variants and 7% carry *HELLS* variants. Only for ICF2, a gender bias was detected, with 79% male patients. This suggests that *ZBTB24* variants might be more deleterious for females.

NEDERLANDSE SAMENVATTING

Een correcte epigenetische regulatie van het genoom is van groot belang om de juiste patronen van genexpressie te garanderen. Hierbij spelen chromatine modifiers, eiwitten die de organisatie van het eiwit-DNA complex "chromatine" reguleren, een grote rol. Verkeerde regulatie van epigenetische processen kan ziektes veroorzaken, zoals het geval is bij Facioscapulohumerale dystrofie (FSHD) en Immunodeficiëntie, Centromeerinstabiliteit, Faciale anomalieën (ICF) syndroom. FSHD en ICF worden beide gekarakteriseerd door lage hoeveelheden DNA methylatie (hypomethylatie) van repeterend DNA.

FSHD is een spierziekte die vooral de spieren in het gezicht, de schoudergordel en de bovenarmen treft. FSHD wordt gekenmerkt door gedeeltelijke relaxatie van de chromatine structuur van de D4Z4 macrosatelliet repeat op chromosoom 4. Dit zorgt ervoor dat het *DUX4* gen, wat zich bevindt in de D4Z4 repeat en codeert voor een transcriptie factor, per abuis tot expressie komt in skeletspieren. Eén van de consistente kenmerken van de relaxatie van de chromatine structuur in FSHD is CpG hypomethylatie van de D4Z4 repeat, inclusief de *DUX4* promotor. In de meest voorkomende vorm, FSHD1, is deze chromatine relaxatie een gevolg van een contractie van de D4Z4 repeat tot 1-10 kopieën (normale lengte: 8-100 kopieën). In de zeldzame vorm van FSHD, FSHD2, ontstaat D4Z4 chromatine relaxatie op alle D4Z4 repeats en dit wordt in de meeste gevallen veroorzaakt door varianten in de D4Z4 chromatine repressor SMCHD1. ICF syndroom is een primaire immuundeficiëntie, gekenmerkt door lage hoeveelheden of afwezigheid van immunoglobulinen in de aanwezigheid van B-cellen in de circulatie, ontwikkelingsachterstand en dysmorfologiën. Vergeleken met FSHD, hebben ICF patiënten hypomethylatie van meer repeterende DNA sequenties verspreid over het genoom en dit zorgt voor onder andere centromeer instabiliteit, het cytogenetische kenmerk van ICF syndroom. Er zijn recessieve varianten in vier genen geïdentificeerd die ICF syndroom kunnen veroorzaken: *DNMT3B* (ICF1), *ZBTB24* (ICF2), *CDCA7* (ICF3) en *HELLS* (ICF4). In dit proefschrift worden het variantspectrum van chromatine modifiers die een rol spelen in FSHD en ICF uitgebreid.

In hoofdstuk 2 worden drie FSHD families beschreven waarin twee mogelijk schadelijke SMCHD1 varianten zijn geïdentificeerd in de proband. Voor elke familie hebben we de fasering onderzocht; d.w.z. of deze varianten zich op één (*in cis*) of twee (*in trans*) allelen bevonden en of ze individueel bijdragen aan D4Z4 hypomethylatie. In de eerste familie hebben we een splice site variant en een nonsense variant *in cis* in exon 21 gevonden. De splice site variant zorgt voor exon skipping van exon 21, wat als consequentie heeft dat de nonsense variant in exon 21 geneutraliseerd wordt. Echter, de skipping van exon 21 zorgt ervoor dat het open reading frame verstoord wordt met een premature stop codon in exon 22.

In de tweede familie is de proband drager van een missense variant in exon 28 en een splice site variant in exon 25. Hoewel van de missense variant in exon 28 voorspelt wordt dat die de functie van SMCHD1 verstoort, zorgt deze variant niet voor D4Z4 hypomethylatie in de moeder van de proband, die ook drager is van deze variant. Dit suggereert dat alleen de splice site variant in exon 25, een variant die in meer FSHD2 families voorkomt, SMCHD1 functie verstoort en bijdraagt aan het fenotype. Deze familie laat ook de voorspellende waarde van de analyse van D4Z4 methylatie zien voor het bepalen van de functionele consequenties van *SMCHD1* varianten.

De proband van de derde familie is drager van twee missense varianten in *SMCHD1* in exon 24 en exon 45 *in trans* en heeft sterke hypomethylatie van D4Z4. Ook in de vader en de moeder van de proband, beide drager van één van de twee missense varianten, wordt D4Z4 hypomethylatie gedetecteerd. Dit geeft aan dat beide missense varianten onafhankelijk bijdragen aan D4Z4 hypomethylatie. De proband en zijn zus, die drager zijn van beide varianten, hebben sterkere D4Z4 hypomethylatie, dan hun ouders die drager zijn van maar één van de varianten. Dit laat zien dat de combinatie van de missense varianten een sterker effect heeft op D4Z4 hypomethylatie. Zowel de moeder en de proband zijn drager van een permissief D4Z4 allel van 27 kopieën, maar alleen de proband met beide *SMCHD1* missense varianten heeft een FSHD fenotype. Dit suggereert dat de varianten ook een additief effect hebben op de presentatie van de ziekte. Deze familie laat ook zien dat een combinatie van twee missense varianten *in trans* niet lethaal is in mannen en vrouwen.

In hoofdstuk 3, hebben we een deletie van het volledige *SMCHD1* gen gedetecteerd op één van de chromosomen 18 in twee FSHD2 families, die eerder negatief getest waren voor exonische *SMCHD1* varianten. De deletie bevat ook een aantal naastgelegen genen, waarvan het niet bekend is dat ze sensitief zijn voor hemizygotie. Deze studie laat zien dat FSHD2 ook veroorzaakt kan worden door *SMCHD1* hemizygotie en dat dit resulteert in een haploinsufficiëntie mechanisme. Eenzelfde mechanisme wordt ook verwacht bij 1/3 van de FSHD2 patiënten die drager zijn van varianten die het open reading frame van *SMCHD1* verstoren.

SMCHD1 hemizygotie is ook aanwezig bij de meeste patiënten met 18p-deletie syndroom, hoewel er geen klinische kernmerken van FSHD zijn beschreven in deze patiënten. Analyse van de D4Z4 methylatie in 18p-deletie syndroom patiënten geeft aan dat de meeste van hen eenzelfde mate van D4Z4 hypomethylatie hebben als FSHD2 patiënten. Gebaseerd op de frequentie in de controle populatie van permissieve 4qA allelen in combinatie met een lengte van de D4Z4 repeat in de FSHD2 range (11-16 kopieën), is de inschatting dat ongeveer 1 op de 8 18p-deletie syndroom patiënten met *SMCHD1* hemizygotie een risico lopen om FSHD te krijgen.

In hoofdstuk 4 hebben we twee FSHD families beschreven met een intronische variant in *SMCHD1*. In de eerste familie is een variant gevonden 15 baseparen proximaal van exon 14 (c.1843-15A>G), die een splice acceptor site creëert. Deze variant resulteert in de inclusie van de laatste 14 nucleotiden van intron 13 in het transcript. Voor deze variant wordt voorspeld dat die het open reading frame verstoort en leidt tot een premature stopcodon. Deze variant is gevonden in de proband en zijn zus, en segregeert met hypomethylatie.

In de tweede familie hebben we een diep intronische variant in intron 34 van *SMCHD1* gevonden. Deze variant creëert een splice acceptor site (c.4347-236A>G), terwijl er al een cryptische splice donor site voorspeld wordt op positie c.4347-183 van *SMCHD1*. Deze variant zorgt voor de inclusie van 53 baseparen van intron 34 in het transcript. Hiervan wordt voorspeld dat dit het open reading frame zal verstoren met een premature stopcodon in exon 35 tot gevolg. In deze familie segregeert deze variant met D4Z4 hypomethylatie en versterkt het fenotype in de familieleden met een permissieve D4Z4 repeat array van 7 kopieën.

D4Z4 hypomethylatie kan niet in alle FSHD2 families verklaard worden door een variant in *SMCHD1*. In hoofdstuk 5 hebben we de identificatie van heterozygote missense varianten in *DNMT3B* in twee FSHD families beschreven. Deze *DNMT3B* varianten segregeren met D4Z4 hypomethylatie en verhoogde ziekte-ernst. We hebben *DUX4* expressie gedetecteerd in MyoD getransduceerde fibroblasten van een aangedaan familielid van een van deze families met een permissieve D4Z4 repeat van 13 kopieën, maar niet in een niet-aangedaan familielid van de andere familie met een permissieve D4Z4 repeat van 44 kopieën. Dit suggereert dat heterozygote *DNMT3B* varianten alleen tot *DUX4* expressie leiden in spiercellen wanneer dit gecombineerd is met kortere D4Z4 repeats.

Omdat homozygote of compound heterozygote *DNMT3B* varianten het autosomaal recessieve ICF1 syndroom veroorzaken, hebben we een klinische en epigenetische vergelijking gemaakt tussen ICF1 patiënten en FSHD2 families met een *DNMT3B* variant. Zowel ICF1 patiënten als FSHD2 patiënten hebben D4Z4 hypomethylatie. Ook een deel van de ouders van ICF1 patiënten, die net zoals deze FSHD2 patiënten drager zijn van een heterozygote *DNMT3B* variant, hebben D4Z4 hypomethylatie. Daarnaast hebben we ook *DUX4* expressie gevonden in myoblasten van een ICF1 patiënt en in MyoD getransduceerde fibroblasten van twee andere ICF1 patiënten. Een aantal van de epigenetische karakteristieken van ICF hebben we ook gedetecteerd in een drager van een heterozygote *DNMT3B* variant uit een van de FSHD2 families. Hoewel een aantal epigenetische karakteristieken overeenkomen, is er geen spierziekte beschreven in ICF1 patiënten en hun ouders, en geen immunodeficiëntie in FSHD2 families, in overeenstemming met ouders van ICF1 patiënten.

We concluderen dat het effect van *DNMT3B* varianten op *DUX4* expressie en FSHD ziekte presentatie, net zoals bij *SMCHD1* varianten, afhankelijk is van verschillende aspecten die geassocieerd zijn met het FSHD1 locus, zoals D4Z4 repeat lengte en de aanwezigheid van een *DUX4* polyadenylatie signaal. Deze studie suggereert dat meerdere factoren betrokken zijn bij de epigenetische status van D4Z4 en de regulatie van *DUX4* expressie in spiercellen.

Tot slot hebben we in hoofdstuk 6 het spectrum van ICF1 en ICF2 varianten uitgebreid. In zeven ICF1 patiënten van vier verschillende families hebben we in totaal zes missense varianten gevonden in het katalytische domein van DNMT3B. Deze varianten zijn niet eerder beschreven in ICF1 patiënten. Verder beschrijven we vijf ICF2 patiënten uit vijf verschillende families. Vier van deze ICF2 patiënten zijn drager van homozygote nonsense varianten in *ZBTB24*, waaronder één nonsense variant die niet eerder is geïdentificeerd in ICF2. De vijfde ICF2 patiënt heeft een homozygote deletie op chromosoom 6, inclusief *ZBTB24*. Deze observatie laat zien dat volledige afwezigheid van het ZBTB24 eiwit niet lethaal is en bevestigt de hypothese dat de meeste ICF2 patiënten functioneel ZBTB24 eiwit missen.

In totaal zijn er nu 77 ICF patiënten beschreven met een gen defect in één van de vier ICF genen die tot nu toe bekend zijn. 56% van de ICF patiënten heeft mutaties in *DNMT3B* (ICF1), 31% heeft mutaties in *ZBTB24* (ICF2), 7% heeft mutaties in *CDCA7* (ICF3) en 7% heeft mutaties in *HELLS* (ICF4). Alleen voor ICF2, hebben we een gender bias gevonden, met 79% mannelijke ICF2 patiënten. Dit suggereert dat *ZBTB24* varianten schadelijker zijn voor vrouwen.

LIST OF PUBLICATIONS

Expanding the mutation spectrum in ICF syndrome: Evidence for a gender bias in ICF2.
M.L. van den Boogaard, P.E. Thijssen, C. Aytekin, F.Licciardi, A.A. Kiykim, L. Spossito, V.A.S.H. Dalm, G.J. Driessen, R. Kersseboom, F. de Vries, M.M. van Ostaijen-ten Dam, A. Ikinciogullari, F. Dogu, M. Oleastro, E. Bailardo, L. Daxinger, E. Nain, S. Baris, M.J.D. van Tol, C. Weemaes, S.M. van der Maarel.
Clin. Genet. 2017 Oct;92(4):380-387

Integrating clinical and genetic observations in facioscapulohumeral muscular dystrophy.
K. Mul, M.L. van den Boogaard, S.M. van der Maarel, B.G. van Engelen.
Curr Opin Neurol. 2016 Oct;29(5):606-13

Mutations in *DNMT3B* Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy.
M.L. van den Boogaard, R.J. Lemmers, J. Balog, M. Wohlgemuth, M. Auranen, S. Mitsuhashi, P.J. van der Vliet, K.R. Straasheijm, R.F. van den Akker, M. Kriek, M.E. Laurense-Bik, V. Raz, M.M. van Ostaijen-Ten Dam, K.B. Hansson, E.L. van der Kooi, S. Kiuru-Enari, B. Udd, M.J. van Tol, I. Nishino, R. Tawil, S.J. Tapscott, B.G. van Engelen, S.M. van der Maarel.
Am J Hum Genet. 2016 May 5;98(5):1020-1029

Double *SMCHD1* variants in FSHD2: the synergistic effect of two *SMCHD1* variants on D4Z4 hypomethylation and disease penetrance in FSHD2.
M.L. van den Boogaard, R.J.F.L. Lemmers, P. Camaño, P.J. van der Vliet, N. Voermans, B.G.M. van Engelen, A. Lopez de Munain, S.J. Tapscott, N. van der Stoep, R. Tawil, S.M van der Maarel.
Eur J Hum Genet 2016 Jan;24(1):78-85

Hemizygosity for *SMCHD1* in Facioscapulohumeral Muscular Dystrophy Type 2: Consequences for 18p Deletion Syndrome.
R.J. Lemmers, M.L. van den Boogaard, P.J. van der Vliet, C.M. Donlin-Smith, S.P. Nations, C.A. Ruivenkamp, P. Heard, B. Bakker, S. Tapscott, J.D. Cody, R. Tawil, S.M van der Maarel.
Human Mutat 2015 Jul;36(7):679-83

Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4.
J. Balog, P.E. Thijssen, S. Shadle, K.R. Straasheijm, P.J. van der Vliet, Y.D. Krom, M.L. van den Boogaard, A. de Jong, R.J. Lemmers, R. Tawil, S.J. Tapscott, S.M van der Maarel.
Epigenetics 2015;10(12):1133-42.

Fusion of hIgG1-Fc to ^{111}In-anti-amyloid single domain antibody fragment VHH-pa2H prolongs blood residential time in APP/PS1 mice but does not increase brain uptake.
M. Rotman, M.M. Welling, M.L. van den Boogaard, L.G. Moursel, L.M. van der Graaf, M.A. van Buchem, S.M. van der Maarel, L. van der Weerd.
Nucl Med Biol 2015 Aug;42(8):695-702

Dysferlin regulates cell adhesion in human monocytes.
A. de Morrée, B. Flix, I. Bagaric, J. Wang, M.L. van den Boogaard, L. Grand Moursel, R.R. Frants, I. Illa, E. Gallardo, R. Toes, S.M. van der Maarel.
J Biol Chem 2013 May 17;288(20):14147-57

CURRICULUM VITAE

Marlinde van den Boogaard was born on June 2nd 1989 in Woerden, The Netherlands. She successfully completed pre-university education (VWO) at Het Ashram College in Alphen aan den Rijn in 2007 with a focus on Science and Health (natuur en gezondheid), supplemented with French and classical languages (Latin and Greek). In September 2007, she started her study Life Science and Technology at the University of Leiden and Delft University of Technology. To finish her Bachelor's programme she performed an internship at the department of Human Genetics at the Leiden University Medical Center (LUMC), focussing on identifying a potential role of dysferlin in cellular adhesion. This internship was supervised by Dr. Antoine de Morrée and Prof. Dr. Silvère M. van der Maarel. After graduating in February 2011, she enrolled in the master Life Science and Technology research and development. She performed her master research project at the department of Human Genetics at the LUMC, focussing on the modification of VHH small antibody fragments targeted against Amyloid β to increase their blood half-life. This research project was supervised by Dr. Maarten Rotman and Prof. Dr. Silvère M. van der Maarel. She received her master's degree with honours (cum laude) in April 2013. In June 2013 she started working as a Ph.D. student in the department of Human Genetics at the LUMC, under supervision of Prof. Dr. Silvère M. van der Maarel, Dr. Judit Balog and Dr. Richard Lemmers. Marlinde has worked as a post-doctoral researcher at the Princess Maxima Center for Pediatric Oncology in the group of Dr. Jan Molenaar, since October 2017.

DANKWOORD

Het is af! Het was een mooie tijd, waarin ik veel geleerd heb en waar ik met plezier op terug kijk. Mijn proefschrift had er nooit kunnen zijn zonder de hulp en steun van de vele mensen waar ik de afgelopen jaren mee samen heb gewerkt of die er voor me waren in mijn vrije tijd. Ik ben jullie daar allen zeer dankbaar voor!

Allereerst mijn promotor Silvère, na mijn stage in jouw groep wist ik al dat ik me op mijn plek zou voelen tijdens mijn promotietraject. Ondanks je drukke agenda was er toch genoeg tijd voor besprekingen en waren artikelen altijd zo nagekeken. Bedankt voor je vertrouwen, je eindeloze enthousiasme en de goede discussies. Richard and Judit, it was great to have two co-promoters who were always available for questions and discussions. Richard, bedankt voor al je hulp en advies bij het genetische werk. Ik heb genoten van je enorme passie voor en kennis van FSHD. Judit, thank you for all the guidance with cell culturing and ChIPs. I learned a lot from you and it was a pleasure to explore the MyoD transductions with you.

Wat is de FSHD groep een gezellige club om mee te werken! Amanda M, thank you for your enthusiasm and your enormous support, both in the lab and with presentations and writing. Maartje, ik weet nog goed hoe je me enthousiast kwam vertellen dat er een PhD positie in de FSHD groep zou komen. Dankjewel voor je steun en voor je goede ideeën. Het was gezellig om al die jaren (al vanaf mijn stages) bij in je in het kantoor te zitten. Patrick, dankjewel voor alle tips en tricks in het lab en je humor. Bianca en Kirsten, bedankt voor het beantwoorden van een hele hoop van mijn praktische vragen de afgelopen jaren en voor jullie gezelligheid en gekke verhalen. Yvonne F, moeder van het lab, dankjewel voor je hulp bij het FACSen en al je goede zorgen! Peter, dankjewel voor je mooie ideeën, levenswijsheden en goede gesprekken. My other roomies, Thomas, Yvonne K, Darina, David, thanks for all the good conversations, the funny stories and all the delicious treats! And Laure, Sandra K, Lucia, Jessica, Gijs, Remko, Kelly, Inge, Fatiha, Amanda P, Mara, Maja, Rob, Anita, Haoyu, Muriel, and the rest and former members of group Frants, thanks for the good suggestions and for a lot of fun during the last years! Maarten en Antoine, door mijn stages bij jullie is mijn liefde voor het onderzoek ontluikt, bedankt voor de leuke tijd. Kajal en Saskia, bedankt voor jullie inzet tijdens jullie stages, het was voor mij leuk en leerzaam om jullie te begeleiden. Anita, Majella, Ingrid en de rest van het secretariaat, bedankt voor alle hulp bij allerlei regeldingen. Babs, bedankt voor de ondersteuning bij de regels rondom het afronden van mijn proefschrift.

Naast de vele lieve collega's wil ik ook mijn vrienden en familie bedanken omdat jullie zo'n goed gezelschap zijn en een fijne afleiding voor het promoveren. Lisa, dankjewel voor je enorme steun en gezelligheid! Louise en Annemarie, fijn om met jullie lief en leed van promotieonderzoek te delen tijdens onze etentjes. Familie Wieser, dank voor de leuke spelletjesavonden, hier en in Frankrijk. Lotte, super om samen met

jou (tochten) te fietsen. Linthia en Rob, Bas en Dashne, Michiel en Amber, Anna, Lisa N, Eline, Janna en Ben, Stephan en Saartje, Melvin en Linda, Daniël en Karlijn, dank voor gezellige avonden. En altijd maar de vraag of het boekje al af is. Nu is het echt zover!

Mijn medetrainsters bij synchroonzwemmen: Amy, Lizzy, Angeline, Naomi, Manon, Kim en Marianne, en de andere trainsters, bedankt voor de gezelligheid! Sanne, dankjewel voor je steun de afgelopen jaren en ik vind het superleuk om weer samen training te geven! Recreatieteam van synchroonzwemmen, het is altijd vertrouwd om met jullie het water in te plonsen en zelf even lekker te zwemmen!

Wat is het fijn om een grote warme familie te hebben. Lieve oma, ooms en tantes, neven en nichten, dank jullie wel voor jullie betrokkenheid, zorgzaamheid en gezelligheid. Sjef en Kitty, Erwin en opa Fred, bedankt voor jullie interesse en warmte. Tim, jouw lach is zo waardevol voor mij. Roselien, fijn om zo'n lieve zus te hebben en om samen leuke dingen te doen! Pap en mam, wat ben ik jullie enorm dankbaar voor al jullie steun en liefde. Jullie zijn altijd geïnteresseerd in wat ik doe en staan altijd voor me klaar. Lieve Remco, wat ben ik je dankbaar voor je steun, je vertrouwen en voor al het plezier dat we samen hebben. Dankjewel dat je er altijd voor me bent!