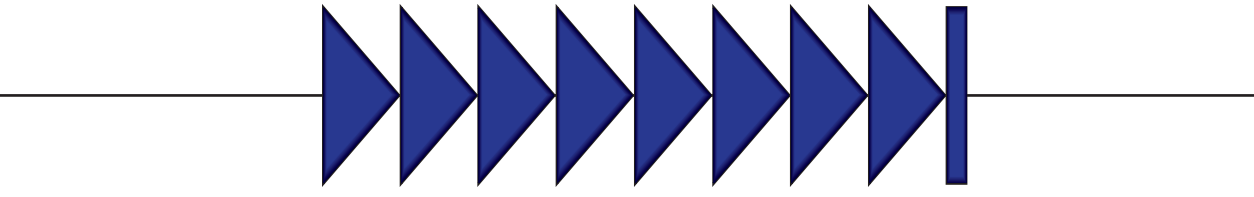

8 General Discussion

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

Repressing D4Z4: balancing repeat size and chromatin modifiers

In the majority of cases, FSHD is caused by contraction of the D4Z4 macrosatellite repeat on chromosome 4q35. Upon this contraction, a changed chromatin organization at the repeat leads to the derepression of the *DUX4* transcription factor in somatic cells. These two sentences summarize almost 20 years of scientific publications about the etiology of FSHD: from the linkage of FSHD to 4q35 and D4Z4 contraction in the early 1990's until the unifying disease mechanism in 2010[1-3]. In between, key publications described the specific association of FSHD with the 4qA haplotype, the involvement of epigenetic dysregulation at D4Z4 and the transcriptional activity of the D4Z4 repeat[4-11]. During this process, in absence of the proof for the unifying genetic mechanism of FSHD, alternative pathomechanisms were described, mainly involving proximal gene dysregulation[12-14].

Recurring throughout the years, and important to all the proposed disease mechanisms, is the epigenetic dysregulation of D4Z4 in FSHD. The partial deletion of the D4Z4 repeat (FSHD1), or mutations in chromatin modifiers of D4Z4 (FSHD2), lead to a changed chromatin conformation at the repeat in somatic cells, which was proposed to spread or loop proximally and affect proximal gene regulation through an *in cis* mechanism. Involvement of the transcriptional activation of *FRG1* and *FRG2* was often studied, however never unequivocally proven to play a role in the FSHD disease mechanism[12-18]. The work in this thesis strengthens a role for epigenetic dysregulation at D4Z4 and derepression of *DUX4* as essential contributors to FSHD pathology. Earlier reports showed reduced binding of the H3K9me3-HP1-Cohesin network to the D4Z4 repeat in FSHD[11]. Indeed, in **chapter 2** we firmly establish that there is a decreased chromatin compaction at D4Z4 (expressed as ChCS) in FSHD, either attributable to increased H3K4me2 or decreased H3K9me3 levels. However, our data in **chapter 4** could not confirm a causal role for H3K9me3 loss in the derepression of *DUX4* in a myogenic context. In patient derived myotubes, characterized by an increase in *DUX4* expression, the relative amount of H3K9me3:H3 at D4Z4 was similar to that observed in controls. Moreover, depletion of SUV39H1, shown to establish H3K9me3 at D4Z4 in HeLa cells and in immortalized human myoblasts[11, 19], or Cohesin subunits in control myotubes was not sufficient to activate *DUX4*. With this in mind, we attribute the decreased ChCS observed in **chapter 2** to an increased level of H3K4me2 in FSHD derived myoblasts, rather than a reduction in H3K9me3. This could simply reflect the increased expression

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

Chapter 4 highlights *SMCHD1* as the most potent known epigenetic regulator of D4Z4 to date as its ectopic expression reverses derepression of *DUX4* in both FSHD1 and FSHD2 derived myotubes. This is in line with genetic analyses showing 1) a causal role for *SMCHD1* mutations in FSHD2 and 2) a modifier effect of *SMCHD1* on disease severity in FSHD1[20-22]. Depletion of SMCHD1 in control myotubes mimics FSHD2 with *DUX4* becoming derepressed. In addition, we show in **chapter 4** that SMCHD1 is indeed partially lost from D4Z4 upon its ectopic depletion and results in increased levels of PRC2 components and the PRC2 associated histone marker H3K27me3 at the D4Z4 repeat. This is also reflected in FSHD2 patient derived myotubes, but not in FSHD1 derived cells. Moreover, chemical inhibition of EZH2, the catalytic subunit of PRC2, leads to increased DUX4 expression, but only in FSHD2 cells. These data indicate that, although highly similar, there are differences between FSHD1 and FSHD2 in the epigenetic regulation of D4Z4. Where in both forms of the disease SMCHD1 is a repressor of the repeat, involvement of the PRC2 complex in D4Z4 regulation could only be detected in FSHD2.

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

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

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

Recently, CRISPR-Cas9 technology was employed to target both a transcriptional activator and repressor to the D4Z4 repeat, leading to increased and decreased DUX4 expression, respectively[23]. This approach could be used to uncover proteins involved in regulating D4Z4 in a more unbiased (non-candidate driven) way. Upon targeting of an inactive Cas9 enzyme to the D4Z4 repeat, immunoprecipitation (IP) of the enzyme allows the enrichment of the D4Z4 chromatin template which can be subsequently analysed by proteomics and/or transcriptomics techniques. Together, these approaches can yield a more comprehensive and complete picture of the proteins and/or histone modifications present at D4Z4 and thereby facilitates the identification of new therapeutic targets and yet unknown FSHD2 disease genes[24].

In **chapter 2** we could not find a significant correlation between the derepression of D4Z4 and the age corrected clinical severity in FSHD patients, although a trend was observed in fibroblasts. In contrast, a correlation between CpG methylation levels at D4Z4 and clinical severity and/or penetrance has been demonstrated in multiple independent studies. In FSHD2 patients, the methylation level attributed to the shortest D4Z4 repeat of a single CpG in the proximal unit of D4Z4 (the FseI site) showed a significant correlation to disease severity[25]. In two other reports, CpG methylation levels throughout the D4Z4 repeat, measured by bisulfite sequencing of specific domains within D4Z4, were shown to be indicative for disease penetrance. Non-affected carriers of an FSHD-sized repeat on a 4qA allele showed methylation levels comparable to control individuals, but are however more susceptible to ectopic *DUX4* derepression than controls[26, 27]. These observations fit with other known genetic and epigenetic characteristics of D4Z4 in FSHD pathology. Firstly, FSHD1 shows the highest penetrance with residual repeat sizes of below 7 units and patients carrying only 1-3 residual repeat units are usually the most severely affected ones[28-33], although a recent study showed high phenotypic variability in this patient group as well[34]. Secondly, asymptomatic carriers of D4Z4 repeats between 7 and 10 units have a higher methylation level than expected based on repeat size[25]. Thirdly, the size of the residual D4Z4 repeat positively correlates with the CpG methylation levels at single CpGs proximal to the repeat[35]. Finally, FSHD2 patients typically carry D4Z4 repeats in the lower size range of controls.

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

units) (**Fig. 1**). This indicates that the ability of SMCHD1, and/or other yet unknown modifiers, to repress the D4Z4 macrosatellite increases with the number of repeat units. It also suggests that to rescue the effect of repeat contraction, modifiers require a minimal amount of repeat units to exert their effect. The insufficient repression by e.g. SMCHD1 with lower repeat numbers could be caused by impaired functionality of the protein at shorter repeats or with inefficient binding to short repeats. The latter would be consistent with the observed preferential binding of SMCHD1 at longer telomeres compared to short telomeres[36]. Although not experimentally proven, this model for penetrance and severity of FSHD must include the assumption that with decreasing capacity to repress D4Z4, the frequency of sporadic activation of *DUX4* and/or the expression levels upon activation increase, and that this leads an earlier onset and more progressive phenotype.

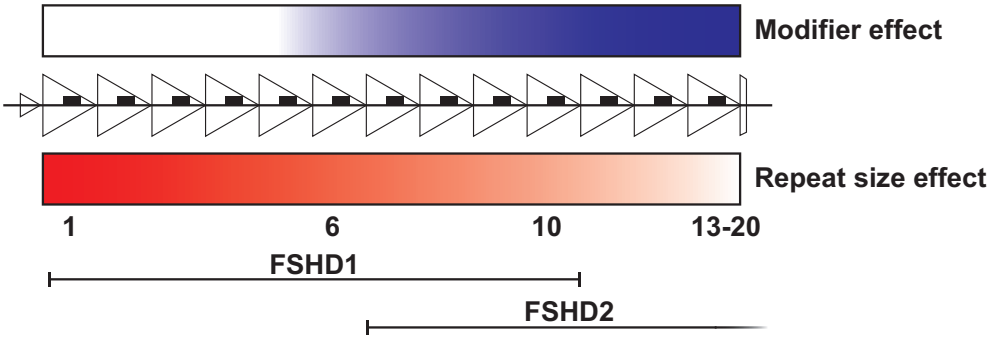

Figure 1: Schematic overview of the relative contribution of D4Z4 repeat size and chromatin modifiers in FSHD1 and FSHD2
Traingles represent D4Z4 repeat units. A D4Z4 repeat arrayn of less than ten units is diagnostic for FSHD1, however there is a considerable effect of modifiers on repression of repeats of more than six units in length. Conversely, in the higher repeat size range (>20 units), the effect of mutations in modifiers like *SMCHD1* does not necessarily lead to pathology.

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

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

These observations are in line with reports where inactivation of *Dnmt*'s, or the chromatin remodeler *Hells*, in murine model systems leads to a similar anti-correlation between CpG methylation and H3K27me3[37-39]. Loss of methylation, through knocking out *Dnmt1*, at already lowly methylated CpG islands in majority leads to decreased levels of H3K27me3 at these promoters[38]. In contrast, CpG poor regions are characterized by

increased presence of H3K27me3[37-39]. The mechanism behind this is unclear, but most likely relies on the affinity of PRC2 for unmethylated CpGs. The redistribution of this mark upon global loss of CpG methylation is proposed to be the effect of a random dilution though the increased genome wide abundance of unmethylated CpGs[38].

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

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

These results either suggest that, at D4Z4, CpG methylation reflects a more potent repressive mechanism than the PRC2 protein complex and its associated histone modifications. Alternatively, the proposed mechanism of PRC2 dilution may lead to a limited pool of PRC2 complexes which are “available” to repress D4Z4 upon loss of CpG methylation at the repeat. Both scenarios would fit with the occasional burst of DUX4 expression in sporadic myonuclei of patients. However, only the latter would explain the discrepancy between FSHD1 and FSHD2 with regard to the involvement of PRC2 and H3K27me3 at the repeat. Most likely, changes in the chromatin organization in FSHD1 cells are confined to the contracted D4Z4 repeat. In contrast, mutations in *SMCHD1* in FSHD2, may have a much broader genome-wide effect on chromatin organization, as suggested by the genome-wide transcriptional dysregulation and the failure of X-inactivation in mouse *Smchd1* knock out cells[40-42]. Although CpG hypomethylation at some other repeats was ruled out in FSHD2[43], studying the genome wide consequences of SMCHD1 loss on CpG methylation and H3K27me3 may reveal an inverse correlation as was observed for *Dnmt's* and *Hells*.

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

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

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

The mechanisms of proximal gene regulation revisited

Over the past decades, numerous studies have reported on potential deregulation of genes proximal to D4Z4 upon its contraction. Either deregulation was never unequivocally proven (*FRG1*), contribution to the FSHD pathomechanism was not clear (*FRG2*), or both (*FAT1*). Few studies have described a potential role for deregulation of *DUX4c* in the pathogenesis of FSHD. *DUX4c* is encoded by an inverted and incomplete D4Z4 repeat unit proximal of *FRG2*. It was reported to be deregulated in FSHD derived samples and to inhibit muscle cell differentiation by interfering with myogenic transcription factors like Myf5 and MyoD[44, 45]. However, in contrast to *DUX4*, its ectopic expression in *Xenopus laevis* did not reveal any developmental abnormalities[46]. Moreover, since it is lost on 4q35 in patients with proximally extended deletions of D4Z4, a major contribution to FSHD pathology is unlikely[47]. Detailed genetic analyses and the subsequent discovery of *DUX4* derepression, yet the most likely and best supported disease mechanism, further challenged the contribution of proximal genes to FSHD. The reported deregulation of both *FRG1* and *FRG2* was attributed to spreading of the derepression at D4Z4 and/or a changed higher order chromatin structure at 4q35. In **chapter 5** we have demonstrated that this is an unlikely explanation for the observed transcriptional upregulation of

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

More recently, a similar study showed that the same mechanism results in the upregulation of *FRG1* in FSHD[48]. The direct binding of DUX4 to the *FRG1* locus was again shown by ChIP and ectopic modulation of DUX4 levels support its role as transcriptional activator of *FRG1*. Intriguingly, whereas the DUX4 binding site at *FRG2* is upstream of the transcriptional start site (TSS), the DUX4 responsive peak in *FRG1* lies in intron 2 of the gene. At LTR repeats, but also at non repetitive targets like *ZSCAN4*, DUX4 has been shown to often create a new TSS at its exact binding site, leading to alternative transcripts from the adjacent target genes[49]. Unfortunately the exact transcriptional consequences of DUX4 binding at *FRG1* and the nature of the DUX4-induced FRG1 transcripts, as well as protein function, were not studied. Thus, both studies provide a central role for the DUX4 protein in the transcriptional activation of *FRG1* and *FRG2* in FSHD, however are not conclusive about the potential role for both genes in the pathology of FSHD.

So far, it is difficult to predict the possible role for DUX4 in regulating *FAT1*, the third gene upstream of D4Z4 which was implicated in FSHD pathology. A detailed study of the interaction between DUX4 and the *FAT1* locus will be pivotal to further evaluate an "D4Z4-independent" role for *FAT1* in the pathogenesis of FSHD. So far, FSHD samples were shown to have decreased FAT1 expression levels, which in a mouse model was shown to affect muscle development[50, 51]. Decreased FAT1 expression could be the consequence of DUX4 expression, although the data are currently inconclusive. On the one hand, DUX4 was shown to decrease FAT1 expression by at least twofold upon its overexpression in human myoblasts[50, 52]. DUX4 ChIP analysis revealed a strong binding peak in intron 2 of *FAT1*, suggesting a direct effect of DUX4 binding on *FAT1*. Intriguingly, if the effect would indeed be direct, repression of *FAT1* would be in disagreement with the currently known function of DUX4 as transcriptional activator. In contrast, inhibition of DUX4 expression through shRNA expression did not affect FAT1 expression levels, but resulted in decreased expression levels of well-established DUX4 target genes[51]. This discrepancy may be caused by the relatively easy way to detect activated target genes which are normally absent versus the diluted effect of restoring *FAT1* expression in only a minority of cells. A small number of individuals which do not display D4Z4 hypomethylation and/or D4Z4 contraction, but do shown an FSHD-like phen`otype were reported to carry possibly damaging *FAT1* variants[53]. These variants were often polymorphisms which are present in the normal population with low frequency[53]. DUX4 expression independent of D4Z4 contraction and/or hypomethylation should be analysed in these patients to exclude that the effect of *FAT1* deregulation is secondary to DUX4.

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

organization at the D4Z4 repeat in FSHD. Additionally, strong genetic evidence supporting the *in trans* model came from the identification of an FSHD1 patient carrying a contracted D4Z4 repeat array on chromosome 10, of which the distal end consists of chromosome 4 derived D4Z4 units[1]. If proximal gene deregulation would be mediated through an *in cis* effect, 10q26 genes would be affected. This was not studied, however if true it would argue against a causal role for *FAT1* and *FRG1* in the pathogenesis of FSHD as these genes are not present on 10q26. Although the deregulation of the upstream 4q35 genes is shown to be secondary to *DUX4* activation, a role in the pathology cannot be excluded and warrants further study.

Recent studies have shown an effect of telomere length on the regulation of 4q35 genes through a telomeric position effect (TPE). In mammals TPE is a poorly defined mechanism of chromatin spreading: telomeric heterochromatin leads to the silencing of genes in close proximity to the telomere and this silencing is lost upon progressive telomere shortening[54]. Expression of *DUX4* and *FRG2*, and more recently *SORBS2*, was shown to be increased with decreasing telomere length in FSHD derived cells[55, 56]. In **chapter 5** we have shown that the observed effect on *FRG2* is most likely secondary to increased DUX4 expression as a result of telomere shortening. *SORBS2* is located ~4.5 Mb upstream of D4Z4, is expressed in skeletal muscle and so far described as a signal transduction and/or structural protein in cardiac muscle cells[55]. The increase of SORBS2 expression is mediated through a long range interaction of *SORBS2* with *FRG1*, which is lost upon extensive telomere shortening in FSHD cells only[55]. Interestingly, several DUX4 binding peaks were identified near or in *SORBS2*, however there is no other evidence for an *in trans* effect of DUX4 on *SORBS2*, as observed for the previously described upstream genes. Overall, there may be a contribution of telomere shortening on the expression of 4q35 genes, including *DUX4* itself, however this effect is small in comparison to the effect of either D4Z4 contraction and/or mutations in modifiers like *SMCHD1*. For example, carriers of residual repeats of 1-3 units are often severely affected in early childhood where extensive telomere shortening is not yet expected. To better establish a role for TPE in FSHD, the results obtained in cellular models in which telomere length was ectopically modified should be translated to in vivo studies or large patient populations. This may prove difficult because the subtelomeric position of these genes may be primate specific and because of the large variation in telomere length in the general population.

SORBS2 is one of the four genes reported so far which are deregulated through disturbed long range interactions with extensive telomere shortening. This mechanism has been described as “TPE over long distances” (TPE-OLD) and differs from the classic TPE. In TPE-OLD, the 3D conformation of the genome rather than individual gene location determines the transcriptional response to extensive telomere shortening[57]. It remains unclear at this point if *DUX4* transcription is affected in a similar mechanism or is affected through classic TPE. The independent observations that SMCHD1 preferentially localizes at longer telomeres and that it is the major repressor of D4Z4 may suggest its involvement in TPE and/or TPE-OLD and might link telomere length to *DUX4* expression[36].

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

In any case, the possible activation of *DUX4* with decreased telomere length and/or senescence may form the molecular basis for the progressive nature of FSHD. With age, average telomere length declines variably but significantly. Moreover, there is build-up of senescent cells throughout aging tissues[62]. Both scenarios may lead to increased transcription of *DUX4* with age and would thereby contribute to the progressive nature and general late onset observed in FSHD. In this view, telomere length could be considered as a modifier of disease, but to support this claim larger studies in patient cohorts and even in specific tissues are necessary.

FSHD disease models: fighting against evolution

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

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

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

The D4Z4-12.5 mouse line carries a D4Z4 repeat in the size range of those observed in FSHD2 individuals. Crossbreeding this mouse line with a mouse line carrying mutations in *Smchd1* essentially generates offspring with an FSHD2 genotype. Moreover, the same approach in the D4Z4-2.5 mouse creates a genotype in which the genetic requirements of both FSHD1 and FSHD2 are met. In humans this genotype leads to an aggravated phenotype and led to the identification of *SMCHD1* as a disease modifier in FSHD1[22]. The generation of these models would yield more mechanistic insight in the effect of SMCHD1 on the D4Z4 repeat, but also offers additional models to study therapeutic interventions aiming at *SMCHD1*. However, it first remains to be determined whether these mice display a similar interaction between *Smchd1* and D4Z4 as observed in humans.

Although genetic, epigenetic and transcriptional features of FSHD are recapitulated in our models, the D4Z4-2.5 mouse does not display a clinically relevant muscle phenotype. Approximately half of the D4Z4-2.5 mice do develop an abnormal eye phenotype of yet unclear etiology. This is in contrast to a published animal model relying on ectopic expression of DUX4. Intramuscular delivery of DUX4-expressing adeno-associated viruses (AAV) resulted in profound local muscle damage through the induction of p53 dependent apoptosis[64]. Mice developed muscle weakness, but also showed quick recovery after these ectopic bursts of DUX4 expression. The expression of DUX4 in this model is localized to the site of injection and does not reflect the typical pattern of sporadic nuclei expressing DUX4. Moreover, the distinct regulatory mechanism including chromatin derepression of a repeat array is not recapitulated in this model[64]. Similar detrimental effects of DUX4 expression, muscle specific or body wide, were observed in *Danio rerio*, highlighting the in vivo toxicity of DUX4[64, 65]. Gross developmental muscle abnormalities and the muscle degeneration seen in *D. rerio* and AAV injected mouse muscle are not representing the muscle histology observed in FSHD patients[66], suggesting that the observed phenomena are generated through

species-specific molecular pathways.

More recently, a doxycycline inducible transgenic mouse model was generated in which the *DUX4* transgene was inserted in an euchromatic region of the mouse genome, upstream of the ubiquitously expressed *Hprt* gene on the X chromosome[67]. Unexpectedly, leaky expression in non-doxycycline-treated animals led to early lethality in male mice only. Females displayed several non-muscular phenotypes, including a striped pattern of scaly skin, possibly as a consequence of random X-inactivation. Rarely surviving male carriers displayed runting, delayed muscle development and had a homogenous scale skin. DUX4 was detectable in several tissues including testis, retina and brain, but was hardly detectable in muscle samples[67]. In concordance with the D4Z4-2.5 mice described in **chapter 6**, cultured muscle cells showed sporadic bursts of DUX4 expression. Strikingly, this mouse model confirms our observation of absence of an obvious muscular muscle phenotype upon systemic *DUX4* transgenesis. What is also in common is the detrimental effect of DUX4 on the eye, which may be related to the observed extra-muscular phenotype of FSHD patients. Again, although low levels of DUX4 expression can be tolerated in the mouse, the virtually absent muscle phenotype suggests different molecular consequences of DUX4 expression between mouse and man.

The absence of a muscle phenotype in the two *DUX4* transgenic mouse models published to date may be a consequence of the evolutionary distance between mouse and man. *DUX4* has no orthologue in the mouse, however a paralogue, *Dux*, has been identified to be present in a tandem array in the mouse. DUX4 in human cells has the propensity to bind repetitive elements, in particular specific subclasses of LTRs, which have expanded specifically in the primate genome. This correlates well with the presence of *DUX4/DUXC* in primates while absent in rodents. This suggests co-evolutionary events in which *DUX4*, normally expressed in germline cells, has a specifically evolved set of target genes to exert its yet unknown normal function. This selective pressure on DUX4 binding sites is absent in the mouse. Therefore, the transcriptional consequences of DUX4 in mouse cells have, are far less obvious than in human cells, as was reported in **chapter 6** and elsewhere[68]. Consequently, the small overlap in transcriptional targets of DUX4 between mouse and man is a plausible explanation for the absence of muscle phenotypes and the manifestation of abnormalities not seen in FSHD patient.

The recent observation that *FRG1* is a DUX4 target gene in human cells, but not in the mouse, was suggested to explain the absence of a muscle phenotype in the D4Z4-2.5 mouse[48]. Although it may have some relevance, the transcriptional activation of *FRG1* in response to DUX4 expression in human cells is low compared to the minimum levels of FRG1 associated with a muscle phenotype in the previously published transgenic *FRG1* mouse[13]. Generating the FSHD2 mouse model described above could reveal a possible role for the activation of human *FRG1* in recapitulating FSHD like muscle symptoms in the mouse, as the transgene in the D4Z4-12.5 mouse model includes *FRG1*. As *FRG1* is only one of the many target genes discordant between mouse and man, and it is currently not known which of the pathways deregulated in humans by DUX4 are causal

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

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

Instead of expressing the *DUX4* transgene in mouse muscle, Zhang *et al*. decribed a completely different approach by xenografting human control and FSHD patient derived muscle biopsies into immunodeficient recipient mice[69]. This approach is similar to our previous work using isogenic muscle cell clones discordant for D4Z4 repeat size[70]. In both models, the human derived muscle cells succeeded to form new myofibers in the mouse muscle and specifically FSHD patient derived material was shown to express DUX4. Although the evolutionary distance between mouse and man is overcome in these models, they have their own limitations. First of all, in these models a contribution of the immune system to FSHD pathology cannot be analysed. More importantly, the xenografting procedure does not allow high-throughput studies and relies on the availability of donor material. Additionally, scoring performance of these animals in functional tests is uninformative with only single muscles being affected by the grafting. In conclusion, every conceivable disease model relying on *DUX4* expression in non-primate species will suffer from the primate specific evolution of DUX4 and its targets, while more humanized models come with their own constraints. With these limitations in mind, the scientific community should adjust their demands and expectations regarding animal models for FSHD.

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

In **chapter 7** we have described the identification of two new disease genes underlying ICF syndrome. ICF syndrome is characterized by a triad of seemingly non-related phenotypes. Firstly, patients suffer from recurring infections of the gastro-intestinal and respiratory tracts due to a- or hypogammaglobunemia, in the presence of B-cells. Secondly, cultured blood cells from patients display chromosomal instability at the centromeres of chromosomes 1, 9 and 16, which is correlated to CpG hypomethylation of centromeric repeats. Finally, almost all patients display a distinct but variable set of facial dysmorphisms, often including hypertelorism, flat nasal bridge and epicanthus. So far, mutations in four different genes - *DNMT3B*, *ZBTB24*, *CDCA7* and *HELLS* - have been identified to underlie the syndrome. Although the phenotype can vary between patients, it cannot be used to classify the patients into one of the genetic subtypes a priori[71]. The overlap between the different patients with regard to the three phenotypic features suggests that the four causative genes functionally converge at one or multiple

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

For two of the four genes, *DNMT3B* and *HELLS*, a clear functional connection has been previously reported. The chromatin remodelling activity of HELLS is required for proper functioning of the de novo methyltransferase DNMT3B during early development[72]. Loss of *Hells* in murine cells mainly leads to defects in CpG methylation at repetitive elements, but also affects a significant number of genic CpG sites[37, 73-75]. These effects were reported to rely on the role of HELLS in enabling the proper establishment of CpG methylation through interacting with DNMT3A and DNMT3B, however the involvement of HELLS in maintaining CpG methylation is unclear yet. Arguing against a role for *Hells* in maintenance of CpG methylation is that the methylation status of episomal DNA can be maintained in absence of HELLS[76]. In contrast, HELLS associates with late replicating DNA, interacts with DNMT1 and *Hells* deficiency was initially reported to affect CpG methylation at an imprinted region, all implicating a role for HELLS in maintaining CpG methylation patterns[77-79]. In **chapter 7** we provide additional evidence for a role of HELLS in maintaining CpG methylation at murine satellite repeats. siRNA mediated knockdown of HELLS in mouse embryonic fibroblasts (MEFs), in which the establishment of DNA methylation is completed, led to decreased CpG methylation of satellite DNA. Considering the molecular function of HELLS, a chromatin remodeler necessary for DNMT functioning, it may not be surprising that it promotes both the establishment and maintenance of CpG methylation patterns.

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

By homology, *ZBTB24* belongs to a family of BTB-domain transcription factors of which some are involved in lymphocyte development[80]. The presence of 8 tandem zinc finger domains suggests that ZBTB24 has DNA binding capacity and can likely act as a transcription factor. However, we have recently discovered that ZBTB24 promotes the repair of DNA double strand breaks during immunoglobulin class switching in B-cells and that this depends on its ZNF domain (unpublished observations). During this process, ZBTB24 binds and stabilizes poly (ADP-ribose) chains on the DNA damage signaling protein poly (ADP-ribose) polymerase 1 (*PARP1*) and thereby promotes repair. Through identification of the molecular function of *ZBTB24* these data for the first time mechanistically explain the immunodeficiency in ICF2 syndrome. Given the high overlap in the immunological phenotype of all ICF patients, it can be anticipated that *DNMT3B*, *CDCA7* and *HELLS* converge with *ZBTB24* at some point during development of antibody

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

So far, no molecular function for *CDCA7* has been established, although it has been implicated in several processes. CDCA7 has been shown to interact with Myc and thereby plays a role in neoplastic transformation[81]. More recently, *Cdca7* was identified in a screen for Notch target genes involved in hematopoietic stem cell emergence[82]. *CDCA7* contains a 4-CXXC zinc finger domain, which is conserved but only shared with its close homologue *CDCA7L*. It has been shown that *CDCA7L* acts as a transcriptional repressor for monoamine oxidases (MAO) and that this likely depends on the DNA binding capacity of the 4-CXXC zinc finger[83]. Based on the high conservation of this domain, a similar function could be expected of *CDCA7*, however the localization of CDCA7L to chromatin depends on its N-terminal p75 binding domain which is not present in *CDCA7*[84].

To address the etiology of ICF syndrome it will be of great interest to better characterize the molecular functions of *ZBTB24* and *CDCA7*. At the same time, it is pivotal to delineate how the absence of these functionalities, and that of *DNMT3B* and *HELLS*, leads to all phenotypic characteristics of ICF syndrome. One way to address these questions is by generating mouse models harboring ICF-like genotypes. Mouse models in which *Dnmt3b* or *Hells* were knocked out, were proven very useful to characterize the molecular function of both proteins, however early lethality in both models greatly impairs studying immunological and developmental features characteristic of ICF syndrome[85, 86]. Moreover, the immunological phenotype observed in ICF mouse models, based on patient derived missense mutations in *Dnmt3b*, involves impaired T-cell function, rather than impaired B-cell function. Based on the genetics observed in ICF patients, a "classic" knockout model for *Zbtb24* should recapitulate ICF2. By using CRISPr-Cas9 technology, the identified missense mutations at conserved residues in *CDCA7* could be introduced into the mouse genome to model ICF3. To overcome (possible) lethality in mouse models for ICF syndrome, an alternative would be to isolate fetal liver cells from developing embryo's deficient for one of the four genes and transplant those into immune compromised recipient mice to specifically follow the development of the immune cell repertoire[87].

In conclusion, the identification of four different genes to underlie ICF syndrome offers a great opportunity to better study the mechanisms underlying the disease. The shared phenotype between all patients calls for the identification of the molecular functions and especially the spatio-temporal expression of the disease genes. Only functions and/or (redundant) pathways shared by all four genes are likely to be disease causing. For example, it remains to be determined whether loss of *ZBTB24*, *CDCA7* and *HELLS* results in similar genome wide CpG hypomethylation as observed in ICF1 derived patient material[88]. Overlapping these shared functions, pathways and phenotypes will likely result in shared defective pathways leading to disease and filter out effects mediated by

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

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

The notion that there is the global defect in CpG methylation patterns in ICF1 patients, including CpG hypomethylation of the D4Z4 repeat, puts those patients at risk for developing FSHD. Indeed, while not extensively studied in ICF2-4 syndrome, ICF1 patients show a similar degree of D4Z4 hypomethylation as FSHD patients[89]. Moreover, a recent report describes two families in which heterozygous mutations in *DNMT3B* segregate with D4Z4 hypomethylation and result in the development of FSHD2 or aggravation of FSHD1 (van den Boogaard et al., 2016). Although limited in the number of patients and families, these data confirm *DNMT3B* as a modifier of D4Z4 chromatin structure and, more importantly, as an FSHD2 disease gene, similar to what is observed for *SMCHD1*. Since heterozygous carriers of ICF1-like *DNMT3B* mutations are apparently at higher risk for developing FSHD, ICF patients and related heterozygous carriers should be carefully examined for the presence of FSHD associated symptoms, which was previously reported not to be the case[90]. However, given the early onset and severity of ICF syndrome, the clinical signs of FSHD, usually with later onset, may have been missed. Conversely, it will be important to evaluate potential comorbidities in FSHD2 families with *DNMT3B* mutations, as the CpG hypomethylation may not be confined to the D4Z4 repeat array. This was suggested in the study of Van den Boogaard et al., as some individuals showed CpG hypomethylation of other repetitive elements in the genome. In concordance with a normal immune-phenotype in ICF1 carriers, this study did not reveal (subclinical) immunological defects in *DNMT3B* carriers from these FSHD2 families. The effect of mutations in *ZBTB24*, *CDCA7* or *HELLS* on genome wide, and more specifically D4Z4, chromatin organization is currently unclear. With similar approaches described in **chapter 4**, the involvement of these genes in D4Z4 repression could be studied to mechanistically support co-occurrence of FSHD and ICF syndrome.

Concluding remarks

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

References

1. Lemmers,R.J. et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329, 1650-1653 (2010).
2. Wijmenga,C. et al. Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. Genomics 9, 570-575 (1991).
3. Wijmenga,C. et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat. Genet. 2, 26-30 (1992).
4. Bodega,B. et al. Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. BMC. Biol. 7, 41 (2009).
5. de Greef,J.C. et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Hum. Mutat. 30, 1449-1459 (2009).
6. Lemmers,R.J. et al. Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. Am J Hum Genet 75, 1124-1130 (2004).
7. Lemmers,R.J. et al. Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am. J Hum. Genet 81, 884-894 (2007).
8. Snider,L. et al. RNA Transcripts, miRNA-sized Fragments, and Proteins Produced from D4Z4 Units: New Candidates for the Pathophysiology of Facioscapulohumeral Dystrophy. Hum. Mol. Genet 18, 2414-2430 (2009).
9. Snider,L. et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS. Genet. 6, e1001181 (2010).
10. van Overveld,P.G. et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat. Genet 35, 315-317 (2003).
11. Zeng,W. et al. Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats in facioscapulohumeral dystrophy (FSHD). PLoS. Genet 7, e1000559 (2009).
12. Gabellini,D., Green,M., & Tupler,R. Inappropriate Gene Activation in FSHD. A Repressor Complex Binds a Chromosomal Repeat Deleted in Dystrophic Muscle. Cell 110, 339-248 (2002).
13. Gabellini,D. et al. Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. Nature 439, 973-977 (2006).
14. Rijkers,T. et al. FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. J Med Genet 41, 826-836 (2004).
15. Jiang,G. et al. Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. Hum Mol. Genet 12, 2909-2921 (2003).
16. Klooster,R. et al. Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. Eur. J. Hum. Genet. 17, 1615-1624 (2009).
17. Winokur,S.T. et al. The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. Chromosome Res 2, 225-234 (1994).
18. Yao,Z. et al. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. Hum. Mol. Genet.(2014).
19. Zeng,W. et al. Genetic and Epigenetic Characteristics of FSHD-Associated 4q and 10q D4Z4 that are Distinct from Non-4q/10q D4Z4 Homologs. Hum. Mutat. 35, 998-1010 (2014).
20. Larsen,M. et al. 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 (2015).
21. Lemmers,R.J. et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat. Genet. 44, 1370-1374 (2012).
22. Sacconi,S. et al. The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. Am. J. Hum. Genet. 93, 744-751 (2013).
23. Himeda,C.L., Jones,T.I., & Jones,P.L. CRISPR/dCas9-mediated Transcriptional Inhibition Ameliorates the Epigenetic Dysregulation at D4Z4 and Represses DUX4-fl in FSH Muscular Dystrophy. Mol. Ther.(2015).
24. Fujita,T. & Fujii,H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation

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

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

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