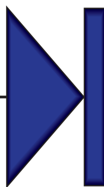
Cover Page

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

Author: Thijssen, P.E.
Title: Genetics and epigenetics of repeat derepression in human disease
Issue Date: 2016-09-01



1

General introduction

Epigenetic regulation of the genome

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

Chromatin, histones and their post-translational modifications

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

In general, euchromatin is characterized by high levels of acetylation on lysine residues in histones (**Fig. 1B**). The chromatin structure is directly affected by histone acetylation as it neutralizes the positive charge of lysine residues on nucleosomes, thereby interfering with the interaction of the nucleosome with negatively charged DNA and increasing the binding possibility of transcription factors to DNA[4]. Both eu- and hetero-chromatic regions are enriched for lysine methylation, which can have different degrees and functionalities depending on the number of methyl groups added to the substrate: mono, di or trimethylation (**Fig. 1B-D**). Both the degree of methylation and the specific histone tail residue used as substrate are associated with different chromatin contexts. For example, active promoters are typically marked by high levels of Histone 3 lysine 4 di- and tri-methylation (H3K4me2/3) (**Fig. 1B**), whereas long distance enhancers are usually marked by H3K4me1[5]. Methylation of H3K4 is thus considered

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

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

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

Lysine methylation can be "read" by a versatile group of protein domains, including the PHD zinc finger and the chromodomain[9]. As for acetylation, "reading" methylation can create a positive feedback loop. H3K9me creates a binding site for the chromodomain of Heterochromatin protein 1 (HP1) which recruits the H3K9me "writer" SUV39H1 (**Fig. 1C**)[10, 11]. Similarly, the WD40 domain of the PRC2 component embryonic ectoderm development (EED) binds H3K27me3 and thereby promotes more H3K27me3 (**Fig. 1D**)[8, 12].

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

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

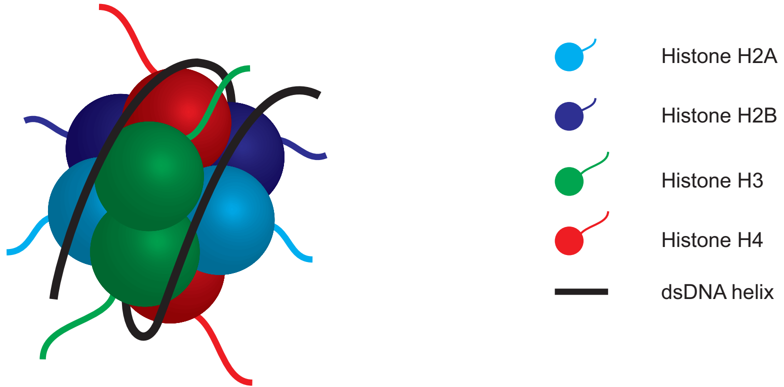
Epigenetic regulation on the DNA backbone: CpG methylation

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

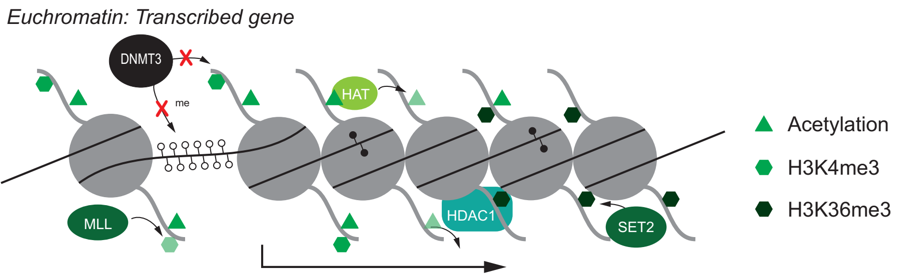
Methylation of CpGs is "written" by DNA methyltransferases (DNMTs). *DNMT1* primarily acts on hemi-methylated DNA and thereby is pivotal for maintaining CpG methylation patterns during DNA replication[17, 18]. DNMT1 is targeted to DNA replication foci by its interaction with proliferating cell nuclear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

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

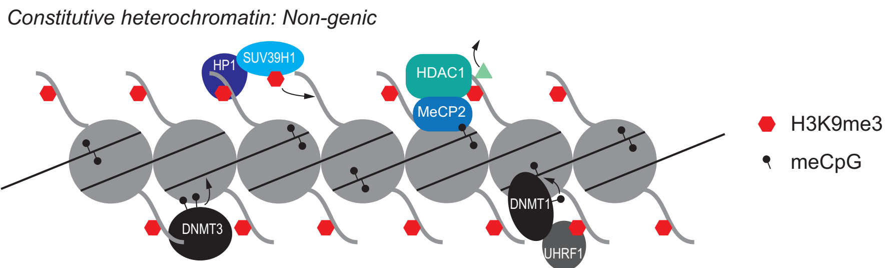
A



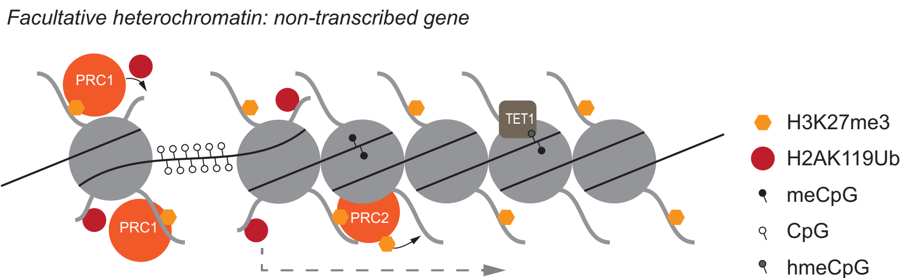
B *Euchromatin: Transcribed gene*



C *Constitutive heterochromatin: Non-genic*



D *Facultative heterochromatin: non-transcribed gene*



DNMT3A and *DNMT3B* encode de novo methyltransferases which, together with the non-catalytic *DNMT3L* co-factor, establish the genome wide pattern of DNA methylation during early development[17, 19]. Establishment of DNA methylation in mammals is at least in part dependent on crosstalk with histone modifications (or the lack thereof). DNMT3A/B enzymes contain an ATRX-DNMT3-DNMTL (ADD) domain which efficiently binds unmethylated H3K4[17]. However, H3K4me3, highly enriched at promoters of actively transcribed genes, inhibits binding of DNMT enzymes and as a consequence promoter CGIs are protected from de novo methylation (**Fig. 1B, D**). In contrast, methylation of H3K9 has a strong positive correlation with CpG methylation. At a subset gene promoters, which are silenced during differentiation, H3K9me (in) directly recruits DNMT3A and/or DNMT3B and thereby promotes CpG methylation. De novo methylation at sporadic, non-genic CpG sites can occur either dependent or independent of H3K9me machineries, reliant on the genetic context. CpG methylation at these sites is an important mechanism to maintain genomic integrity and preserve the heterochromatic conformation of non-transcribed loci[17].

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

More recently, a class of enzymes was discovered that can “erase”, or better “edit”, CpG methylation. Active removal of CpG methylation is carried out through stepwise oxidation of the methyl group to hydroxymethyl, formyl and carboxyl which finally can be removed and subsequently repaired. This oxidation, and removal of meCpG, is carried out by ten eleven translocation 1 (TET1), TET2 and TET3 proteins (**Fig. 1D**)[21]. Next to active removal of CpG methylation, TET enzymes create another layer of possible epigenetic regulation: the intermediates formed by the TET enzymes may have biological roles themselves[22]. In support of this, for example, is the observed stable and persistent enrichment of hydroxymethylation at euchromatic regions in cells of the neuronal lineage, which positively correlates with gene expression[22]. In summary, CpG methylation is established and maintained by DNMTs, interpreted by MBD containing proteins and removed by TET enzymes. It correlates with histone modification patterns and together these epigenetic systems dictate the organization of the chromatin template and create a platform to maintain and use genetic information in order to establish heritable patterns of gene expression, which identify cell identity.

Epigenetics and disease

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

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

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

The example of fragile X syndrome shows that a gene mutations can have an epigenetic effect *in cis* which leads to disease. The list of disorders where genetic mutations lead to an epigenetic phenotype *in trans* is considerably larger. Mutations in numerous "writers", "readers" and "erasers" have been identified to underlie syndromes, often characterized by developmental problems and intellectual disability[26]. An intriguing example of an *in trans* disorder is Kabuki syndrome, characterized by intellectual

disability, facial dysmorphisms and short stature. Kabuki syndrome is caused by mutations in *MLL2* or *KDM6A*, an H3K4 KMT and H3K27 KDM respectively[28, 29]. By modulating lysine methylation on histones MLL2 promotes chromatin relaxation whereas KDM6A inhibits chromatin repression. This essentially results in the same: a shifted balance of gene expression at target genes of these machineries, which is supported by the indistinguishable phenotype of both patient groups[26].

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

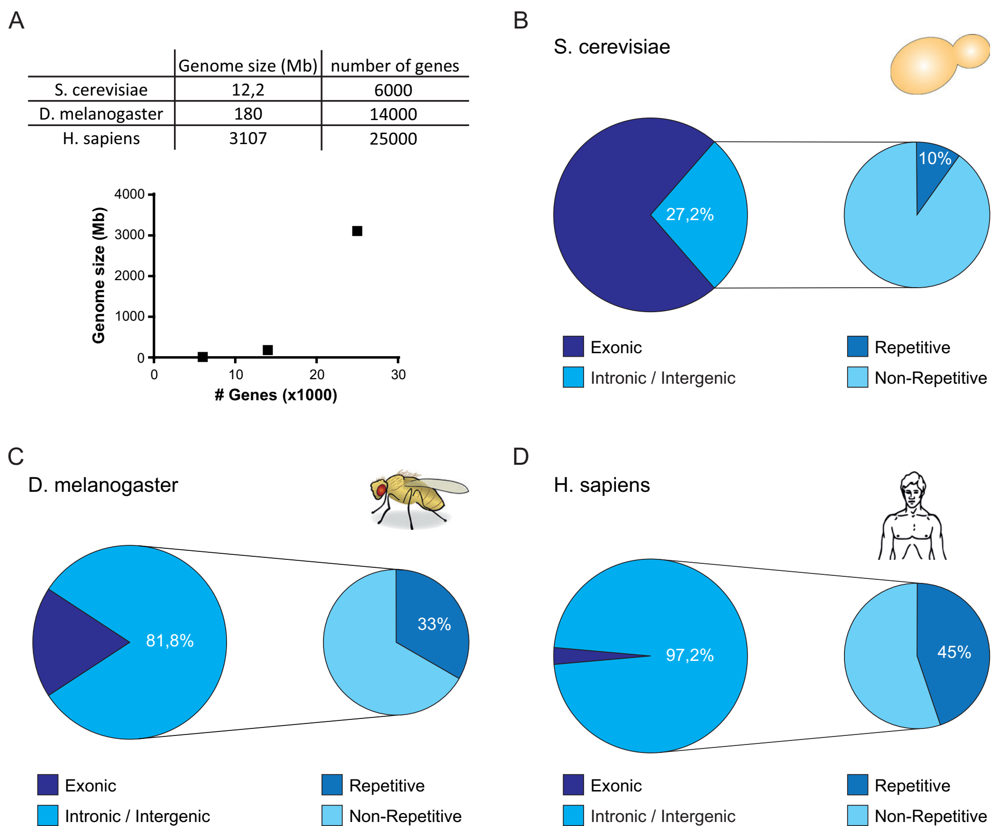
The repetitive genome: expand and silence.

With the completion of the human genome project at the beginning of the century, early estimates of the total number of genes in the human genome (around 100.000) were proven wrong[30]. In fact, the latest numbers indicate that the human genome contains less than 25.000 genes. Compared to the number of genes identified in lower, less complex, eukaryotes like *Saccharomyces cerevisiae* and *Drosophila melanogaster*, it becomes clear that increased organismal complexity does not solely depend on the number of genes (**Fig. 2A**)[31, 32]. Besides, the number of identified genes in these organisms does not reflect the size of their genomes (**Fig. 2A**). In other words: the complexity of human life, compared to that of budding yeast or fruit fly, cannot be simply explained by an increase in the number of genes.

Rather than the coding part of the genome, the steep increase in non-coding DNA sequences underlies the dramatic expansion of the human genome compared to that of other eukaryotes. With increasing genome sizes in *Saccheromyces cerevisiae*, *Drosophila melanogaster* and *Homo sapiens*, there is a concomitant decrease in the percentage of protein coding basepairs (**Fig. 2B-D**)[33]. The vast majority of the human genome, more than 97%, is actually non-protein coding and was referred to as non-functional “junk DNA”[34]. The publication of the “Encyclopedia of DNA Elements” (ENCODE), however, revealed that many of these junk DNA regions are actually functional, e.g. by acting as distant gene expression enhancer sites, and contribute to the regulation of gene expression patterns[35]. The increase in non-coding regulatory elements thus creates an additional layer of transcriptional regulation and thereby contributes to organismal complexity.

Apart from the size of the genome, the fraction of repetitive DNA positively correlates to organismal complexity (**Fig. 2B-D**). Up to 45% of the non-coding part of the human genome is repetitive of nature and is typically packed into constitutive heterochromatin. Repeated sequences include large stretches (10-300 kb) of duplicated sequence blocks known as segmental duplications. However, the majority of repetitive DNA is comprised of two main classes of highly repetitive elements: interspersed and tandem repeats[36].

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



	Genome size (Mb)	number of genes
S. cerevisiae	12,2	6000
D. melanogaster	180	14000
H. sapiens	3107	25000

Figure 2: The size of the genome and the fraction of repetitive DNA correlate with organismal complexity.
A) The genome size, rather than the number of encoded genes, correlates with increasing organismal complexity in *S. cerevisiae*, *D. melanogaster* and *H. sapiens*. **B-D)** The relative amount of non-coding DNA (intronic and intergenic) and the relative distribution between non-repetitive and repetitive DNA in *S. cerevisiae* **(B)**, *D. melanogaster* **(C)** and *H. sapiens* **(D)** shows a correlation between organismal complexity and the amount of repetitive DNA.

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

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

FSHD: derepression of a macrosatellite repeat

FSHD (OMIM 158900/158901) is a progressive muscular dystrophy first described by Landouzy and Dejerine, with recent estimates to affect approximately 1 in 8000 individuals[38, 39]. Patients suffer from progressive weakening of the facial, shoulder and proximal limb muscles and often show asymmetric involvement of muscles[40]. With disease progression, also other muscles may become affected. FSHD mostly shows an age at onset in the second decade of life, but is however characterized by a high inter- and intra-familial variability in onset, progression and severity[40]. Extreme cases show muscle weakness at birth, whereas some individuals remain asymptomatic throughout life. Eventually, 20% of FSHD patients above the age of 50 years become wheelchair bound. A minority of patients shows respiratory and cardiac involvement (atrial arrhythmia), of which the latter is rarely symptomatic. Extra muscular symptoms have been reported and mainly involve retinal vasculopathy and progressive hearing loss[40].

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

In most cases, FSHD is inherited in an autosomal dominant manner, with a high frequency (10-30%) of de novo cases[39, 41]. In the early nineties, linkage analysis revealed that FSHD segregates with marker loci in the subtelomere of chromosome 4q35, which harbours the D4Z4 macrosatellite repeat (**Fig. 3A**)[42-44]. Each D4Z4 repeat unit is 3.3 kb in size and the number of repeats per allele is highly polymorphic. The D4Z4 array consists of 1 to over 100 units, leading to a possible size difference of more than three megabasepairs between individual alleles (**Fig. 3A**)[37]. The 4q subtelomere exists in two equally frequently occuring haplotypes (4qA and 4qB), and FSHD uniquely associates with the A variant[45-48]. Using restriction enzyme analysis, it was found that partial deletion of D4Z4 on 4qA alleles, resulting in a repeat array of less than 11 but more than 1 units, leads to the development of FSHD type 1 (FSHD1)[49-52]. The number of residual repeats shows a rough positive correlation with age at onset and wheelchair use[53, 54]. Only contraction of 4qA alleles is pathogenic since D4Z4 repeat arrays of less than 10 units in the control population can be observed on 4qB chromosomes[46].

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

haplotype, as all FSHD2 patients carry at least one such allele[55].

The D4Z4 macrosatellite repeat is located in the subtelomere of chromosome 4q35, which is immediately adjacent to the intact telomeric [TTAGGG] repeats. Subtelomeres are characterized by the presence of repetitive DNA and segmental duplications and are packed in a constitutive heterochromatin structure like the adjacent telomeres[57, 58]. Subtelomeric segmental duplications have occurred both intra- and inter-chromosomally and the duplicons can also be identified in non-subtelomeric regions of the genome, such as pericentromeres[57]. Indeed, the subtelomere of chromosome 4q, including the D4Z4 repeat array, is duplicated to the subtelomere of chromosome 10q (**Fig. 3A**), but contractions of the 10q copy of D4Z4 are typically not pathogenic[59, 60]. Additionally, single, often incomplete, D4Z4 copies can be found dispersed throughout the genome, but were never linked to pathogenicity[61-63].

Together, genetic analyses put the partial deletion of the D4Z4 macrosatellite repeat at 4q35 at the centre of FSHD pathology. Each D4Z4 unit encodes a copy of the *DUX4* retrogene, a member of the double homeobox transcription factor gene family which has only been identified in placental mammals[64]. *DUX4* is most likely a retrotransposed copy of the ancestral and intron containing *DUXC* gene which is lost in the primate lineage[64, 65]. *DUX4* does not have a rodent orthologue, however a paralogue has been identified: the rodent specific Dux array identified in mouse and rat suggests divergent evolutionary events leading to conservation of a tandemly repeated *Dux* gene[64-66]. Remarkably, the organization of *DUXC/DUX4/Dux* like genes into a tandem repeat array is conserved in mammals[66]. By ectopic expression, *DUX4* was found to be a pro-apoptotic protein and an inhibitor of muscle cell differentiation, however its expression or dysregulation in FSHD muscle could for a long time not be established[67-70]. The non-detectable dysregulation of DUX4, together with the fact that only partial deletion of the heterochromatic D4Z4 repeat causes FSHD, suggested an epigenetic component in FSHD disease aetiology[56].

The complex interplay of chromatin regulators at D4Z4

The D4Z4 repeat, as most macrosatellites, is transcriptionally silenced and organized into heterochromatin in somatic cells. D4Z4, characterized by a high density of CpG dinucleotides, is highly but inhomogeneously methylated and it is marked by H3K9me3 in somatic cells, consistent with its heterochromatic nature (**Fig. 3B**)[55, 56, 71-75]. Remarkably, histone markers for euchromatin (acetylation), as well as facultative heterochromatin (H3K27me3) were also found to be enriched at D4Z4 (**Fig. 3B**)[75, 76]. In FSHD individuals, the chromatin organization is disrupted in somatic cells as CpG methylation levels are reduced at D4Z4 (**Fig. 3B**)[55, 56, 72]. Moreover, using primary myoblasts and fibroblasts, it was shown that H3K9me3, mediated by SUV39H1, is decreased at D4Z4 in FSHD patient derived cell lines compared to healthy controls or patients suffering from other muscular dystrophies (**Fig. 3B**)[75]. Furthermore, the downstream “readers” of H3K9me3, HP1γ and Cohesin, were shown to be reduced at D4Z4[75]. Together this shows that the heterochromatin organization at D4Z4 is disrupted in patients, leading to partial relaxation of the locus.

The epigenetic changes observed at D4Z4 are common to FSHD1 and FSHD2. In fact, in FSHD2 individuals the chromatin changes are observed on both D4Z4 repeat arrays on chromosome 4 as well as the 10q copies, whereas in FSHD1 individuals the effects are restricted to the contracted pathogenic repeat[55, 56, 75]. FSHD1 is an *in cis* epigenetic disorder: the contraction of the repeat leads to a change in local chromatin structure, similar to fragile-X syndrome. In contrast, FSHD2 is an *in trans* epigenetic disorder as >80% of the FSHD2 patients carry mutations in the structural maintenance of chromosomes flexible hinge domain containing 1 (*SMCHD1*) gene, which is underlying the changes in D4Z4 chromatin structure[77].

SMCHD1 is structurally related to the SMC protein superfamily, which constitutes core proteins of the Cohesin complex, and was first identified in a screen to identify epigenetic modifiers of variegated expression in a murine model[78]. *Smchd1* has been shown to play a role in X-chromosome inactivation, an epigenetic process ensuring dosage compensation in females by silencing one of the two X chromosomes. A hallmark of X-chromosome inactivation is the expression of a long non-coding RNA (lncRNA) known as *Xist*. *Xist* covers the X-chromosome *in cis* and recruits the PRC2 complex to ensure gene silencing throughout the inactive X-chromosome, with the exception of some genes that escape this process[79]. In female *Smchd1* knockout mice, X-chromosome inactivation is perturbed, with promoter hypomethylation of CpG islands and concomitant upregulation of clustered transcripts normally subject to X-chromosome inactivation, showing a role for Smchd1 in establishment and/or maintenance of CpG methylation[78, 80-82]. Furthermore, it has been shown that SMCHD1 is involved in the higher order compaction of the inactivated X-chromosome by interacting with *Xist* and H3K27me3[82]. Next to its role in X-chromosome inactivation, Smchd1 is involved in the silencing of several mono-allelically expressed autosomal genes, among which the clustered protocadherin genes on mouse chromosome 18[81, 83].

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

D4Z4 chromatin changes in FSHD lead to the derepression of DUX4

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

genes. This hypothesis relied either on proximal spreading of the altered chromatin structure at D4Z4 and/or changes in higher order chromatin organization and long range interactions. 4q35, in contrast to 10q26, preferentially localizes to the nuclear periphery[88]. This is likely mediated through interactions with the nuclear matrix, which is disturbed upon D4Z4 contraction[89]. Next to disturbed interactions with the nuclear matrix, D4Z4 contractions also lead to an altered higher order chromatin structure at

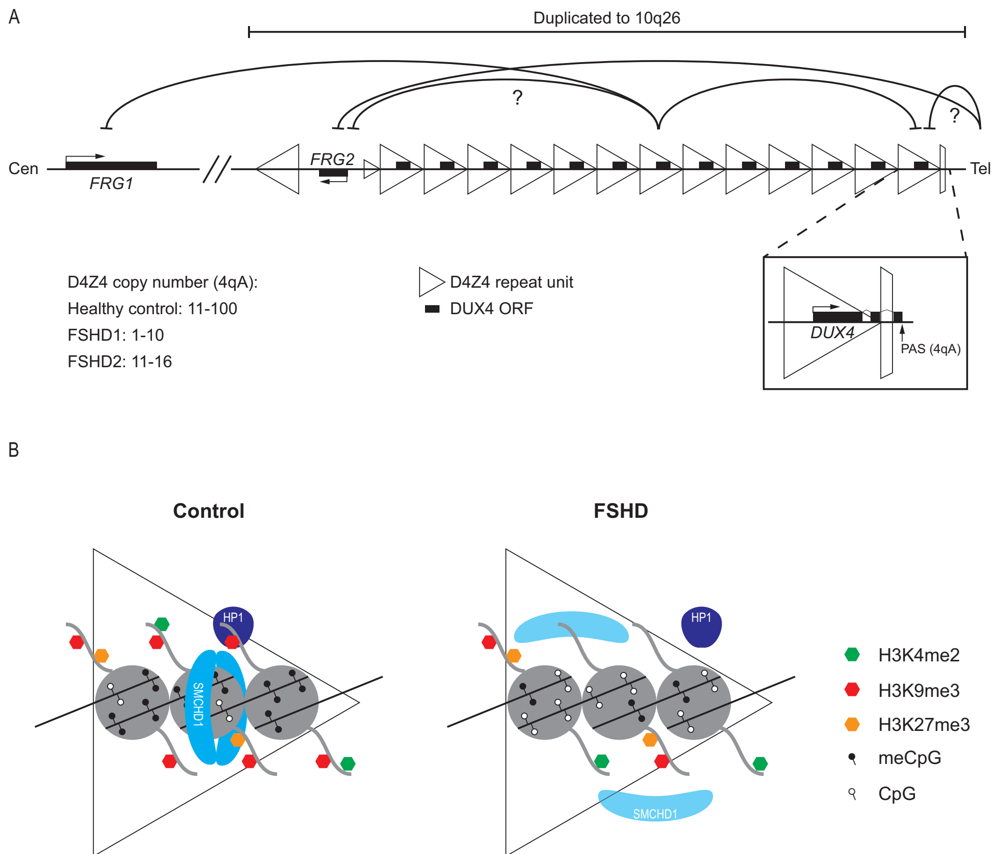


Figure 3: schematic representation of the genetic and epigenetic features of the FSHD locus on chromosome 4q35.
A) The D4Z4 macrosatellite (triangles) and its flanking sequences, including the proximal *FRG1* and *FRG2* genes, is contracted in FSHD1 patients. All patients (FSHD1 and FSHD2) carry the 4qA variant of the p-LAM sequence element distal to the D4Z4, encoding a non-canonical poly adenylation signal allowing the formation of stable DUX4 transcripts. The D4Z4 repeat and some flanking sequences are duplicated to chromosome 10q26. Arches depict the reported long range interactions and/or position effects of the chromatin affected in FSHD. Inset: overview of the DUX4 transcript produced from the most distal D4Z4 unit and the p-LAM sequence. The full *DUX4* open reading frame is included in the first exon and is therefore present in each repeat unit. B) The D4Z4 chromatin structure is characterized by the presence of methylation markers for both eu- and hetero-chromatin (hexagons) and high levels of CpG methylation. D4Z4 compaction is further established by binding of HP1 and SMCHD1. In FSHD patients, D4Z4 is decompacted evidenced by a loss of CpG methylation and H3K9me3 with a concomitant loss of SMCHD1 and HP1 binding.

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

Two candidate genes were identified in the region flanking D4Z4 proximally: FSHD region gene 1 (*FRG1*) and *FRG2*, of which the latter is also present on chromosome 10 (**Fig. 3A**)[93, 94]. Both *FRG1* and *FRG2* were reported to be upregulated in FSHD, suggesting a mechanism of long range interactions and/or spreading of chromatin derepression from the D4Z4 repeat (**Fig. 3A**)[63, 93, 95-97]. Overexpression of FRG1, an Actin bundling and mRNA processing protein, induces a muscular dystrophy phenotype in different animal models, however most follow up studies failed to confirm the upregulation of *FRG1* in FSHD patient material[96-112]. In contrast, *FRG2* activation is a robust and reproducible hallmark of FSHD cells, however its function is unknown and overexpression in mice did not lead to a muscle phenotype[93, 97, 101].

More recently, few studies revealed a possible link between deregulation of FAT atypical cadherin 1 (*FAT1*), located 3.6 Mb upstream of D4Z4, and FSHD pathology[113-115]. Mice in which *Fat1* was genetically ablated developed asymmetric muscle wasting reminiscent of FSHD, and genetic analysis of *FAT1* in human patients may suggest a secondary or indirect involvement of *FAT1* in FSHD pathology[113-115]. Altogether, these studies highlight that genes proximal to the D4Z4 repeat may be deregulated in FSHD. Their involvement in the disease mechanism and the mechanisms behind their deregulation remain unclear at this point.

With the identification of several transcripts produced from D4Z4, including an mRNA encoding the full length DUX4 protein, efforts to identify the FSHD disease mechanism focused on *DUX4* again[116]. The key to the FSHD disease mechanism lies within the unique association of FSHD with the 4qA haplotype distal to the D4Z4 repeat[45-47]. This sequence element (pLAM) encodes 1) an additional *DUX4* exon distal to the last repeat unit and 2) a non-canonical *DUX4* polyadenylation signal (PAS). Both these elements are absent in 4qB alleles, while the PAS is absent from 10q alleles[68, 117]. It was shown that the presence of this PAS can lead to the formation of a stable full length DUX4 transcript and genetically unifies all FSHD patients (**Fig. 3B**)[117]. Additionally, full length DUX4 was shown to be abundantly expressed in sporadic myonuclei of FSHD derived proliferating myoblasts and differentiated myotubes, but at low or even undetectable levels in control derived material[118, 119].

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

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

Upon expression, DUX4 acts as a potent transcriptional activator. In muscle cells, DUX4 activates a specific set of genes, through direct binding to a double homeodomain DNA motif[120]. In response to DUX4 activation, genes involved in in germline biology, early stem cell development and innate immunity are deregulated[120]. Furthermore, DUX4 binds and activates retroelements, mainly of the ERV/MaLR type, which can lead to the formation of alternative transcriptional start sites for flanking genes[120, 121]. Many of the DUX4 targets identified by overexpression are deregulated in patient derived material, including fetal muscles, and account for the majority of transcriptional changes between FSHD and control muscle cells and/or biopsies[112, 120, 122]. More recently, a reporter based approach, allowing transcriptome analysis of individual muscle cells expressing endogenous *DUX4*, confirmed many of these targets and highlighted a role for disrupted RNA metabolism in FSHD pathology[123].

Apart from initiating aberrant transcriptional programs in muscle cells, DUX4 expression has other detrimental effects in various model systems which may or may not depend on its function as a transcriptional activator. In rhabdomyosarcoma cells it was shown that DUX4 induces cell cycle arrest in a P21 dependent fashion, possibly impacting muscle regeneration[124]. Besides, DUX4 expression in murine ES cells leads to reduced pluripotency and an imbalance in the formation of the three germlayers upon differentiation[125]. Expression of DUX4 in mesenchymal stromal cells promotes their differentiation into osteoblasts by an unknown mechanism[126]. DUX4 was also shown to inhibit RNA degradation through nonsense mediated decay (NMD). Expression of DUX4 leads to the degradation of the NMD factor UPF1, thereby creating a positive feedback loop as DUX4 itself is a substrate for NMD[127]. Altogether, this likely accounts for the observed toxicity of DUX4 expression in skeletal muscle cells.

Although the consequences of DUX4 expression are extensively studied, the mechanisms underlying the sporadic activation of DUX4 are not so well understood. What is driving the sporadic bursts of expression? Why is DUX4 expression increased during myotube differentiation? Few studies have focused on these aspects and how the sporadic expression of DUX4 leads to the progressive and variable phenotype characteristic of FSHD. For example, *DUX4* activation is repressed by active Wnt/β-catenin signalling[128]. Next to that, *DUX4* activation was linked to the activity of two enhancers proximal to D4Z4, which show myogenic activity in controls and patients[129]. *DUX4* was also shown to be controlled by a telomere position effect (TPE), a chromatin mediated regulation similar to what was proposed for proximal gene regulation by D4Z4 (**Fig. 3A**)[130]. It was shown that the expression levels of *DUX4*, as well as *FRG2*, inversely correlate with the length of the adjacent 4q telomere. As telomere length naturally declines with (cellular) age and was shown to influence the epigenetic regulation of the adjacent subtelomeres[131, 132], this study possibly links the expression levels of *DUX4* to the progressive nature of FSHD.

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

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

Next to new avenues for mechanistic and molecular studies, the firm establishment of the FSHD disease mechanism also paves the road for the generation of animal models to initiate more translational research. The generation of faithful animal models is however challenged by the fact that the D4Z4 macrosatellite and the *DUX4* gene do not have a homologue in a similar genomic context in other, non-primate species[65]. Recapitulation of the FSHD phenotype therefore has to rely on the ectopic expression of DUX4 in animal models, with the potential pitfall that the transcriptional targets of DUX4, and thereby its molecular effects, may not be conserved between different species.

ICF syndrome: an epigenetic disorder in trans

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

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

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

the missense type, resulting in reduced methyltransferase activity of DNMT3B[136, 142, 143]. ICF1 patients carry at least one partially functional *DNMT3B* copy, as nonsense alleles are only identified in combination with missense alleles in compound heterozygotes[136]. This is in line with the observed phenotypes of mouse models for ICF1. Whereas *Dnmt3b*$^{-/-}$ mice show embryonic lethality, hypomorphic mouse models carrying (patient derived) missense mutations in *Dnmt3b* present with CpG hypomethylation at centromeric satellite repeats, craniofacial abnormalities, runting and an impaired immune system characterized by increased levels of apoptosis in T-cells[144-146]. Although some features of ICF syndrome are clearly recapitulated in these models, the most prominent difference is that none of the available models displays impaired B-cell functionality.

In line with a defect in one of the two de novo methyltransferases, the genome of ICF1 patients is characterized by global hypomethylation at both coding and non-coding regions[147-150]. Moreover, higher order chromatin structure organization is altered in ICF1 patients, exemplified by a changed nuclear localization of genes on the inactived X-chromosome[151]. ICF patients show hypomethylation at various types of repetitive elements throughout the genome, including interspersed LINE repeat elements and tandem repeats like centromeric satellites and the subtelomeric D4Z4 macrosatellite[139, 147, 152, 153]. CpG hypomethylation at subtelomeres in ICF patients correlates with extensive shortening of the adjacent telomeres[154]. Telomeres are transcriptionally active, as they have been shown to produce telomere repeat containing RNAs (TERRA)[155, 156]. In ICF1 derived patient cell lines an increase in expression levels of TERRA lncRNAs has been observed, most likely in some way linked to the extensive telomere shortening in these cells[154]. How these observations contribute to the disease mechanism remains unclear at this point. In general, analyses of methylation at genic and non-genic regions and transcriptional changes in ICF1 patient derived cell lines only showed partial correlations and have not revealed a comprehensive disease mechanism yet[147-150, 157, 158].

A second group of patients, negative for mutations in *DNMT3B*, shares all epigenetic and phenotypic characteristics with ICF1 patients, however with additional hypomethylation of alpha-satellite DNA, a centromeric macrosatellite[135, 159]. Additionally, specific germline genes display hypomethylation and concomitant transcriptional activation in ICF1 derived patient material only[150]. Genetic analyses of these DNMT3B mutation negative patients revealed that the majority has mutations in zinc finger and BTB domain containing protein 24 (*ZBTB24*; ICF2; OMIM 614064)[160-163]. In contrast to what has been observed for *DNMT3B*, mutations in *ZBTB24* do not localize to a confined domain of the gene[136]. In fact, mutations in *ZBTB24* are almost exclusively of the nonsense type, most likely leading to complete absence of the full length protein in patients[136]. Thus far, no molecular function has been described for *ZBTB24*, although by homology it is member of the ZBTB family of (hematopoietic) transcription factors[164]. The BTB domain, found in the N-terminus of ZBTB24, mediates homo- or hetero-dimerization and may facilitate additional protein-protein interactions[164]. The C-terminal zinc finger (ZNF) of ZBTB proteins is thought to mediate the localization to specific DNA sequences. Based hereon, *ZBTB24* is functionally unrelated to *DNMT3B* and discovering its function is of great importance to understand ICF pathology.

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

Outline of the thesis

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

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

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

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

Finally, with regards to ICF syndrome, **chapter 7** describes the identification of mutations in the cell division cycle associated 7 (*CDCA7*) and the helicase, lymphoid specific (*HELLS*) genes in previously unexplained cases. By doubling the number of ICF disease genes, this work highlights the genetic heterogeneity of the disorder and leaves only few cases unexplained. Molecular characterization of these genes will help to decipher the pathomechanism of ICF syndrome.

References

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

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

50. van Deutekom,J.C. et al. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum. Mol. Genet. 2, 2037-2042 (1993).
51. Wijmenga,C., Brouwer,O.F., Padberg,G.W., & Frants,R.R. Transmission of de-novo mutation associated with facioscapulohumeral muscular dystrophy. Lancet 340, 985-986 (1992).
52. Wijmenga,C. et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet 2, 26-30 (1992).
53. Lunt,P.W. et al. Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). Hum Mol Genet 4, 951-958 (1995).
54. Tawil,R. et al. Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. The FSH-DY Group. Ann Neurol 39, 744-748 (1996).
55. de Greef,J.C. et al. Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Hum. Mutat. 30, 1449-1459 (2009).
56. van Overveld,P.G. et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat. Genet 35, 315-317 (2003).
57. Ambrosini,A., Paul,S., Hu,S., & Riethman,H. Human subtelomeric duplicon structure and organization. Genome Biol. 8, R151 (2007).
58. Mefford,H.C. & Trask,B.J. The complex structure and dynamic evolution of human subtelomeres. Nat Rev Genet 3, 91-102 (2002).
59. Bakker,E. et al. The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle Nerve 2, 39-44 (1995).
60. Deidda,G. et al. Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the Facioscapulohumeral muscular dystrophy locus on chromosome 4qter. Eur J Hum Genet 3, 155-167 (1995).
61. Hewitt,J.E. et al. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum Mol Genet 3, 1287-1295 (1994).
62. Lyle,R., Wright,T.J., Clark,L.N., & Hewitt,J.E. The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. Genomics 28, 389-397 (1995).
63. Winokur,S.T. et al. The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. Chromosome Res 2, 225-234 (1994).
64. Clapp,J. et al. Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. Am. J. Hum. Genet. 81, 264-279 (2007).
65. Leidenroth,A. & Hewitt,J.E. A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. BMC. Evol. Biol. 10, 364 (2010).
66. Leidenroth,A. et al. Evolution of DUX gene macrosatellites in placental mammals. Chromosoma 121, 489-497 (2012).
67. Bosnakovski,D. et al. DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. Exp. Neurol. 214, 87-96 (2008).
68. Dixit,M. et al. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proc. Natl. Acad. Sci. U. S. A 104, 18157-18162 (2007).
69. Kowaljow,V. et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord 17, 611-623 (2007).
70. Wallace,L.M. et al. DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. Ann. Neurol. 69, 540-552 (2011).
71. Gaillard,M.C. et al. Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. Neurology 83, 733-742 (2014).
72. Hartweck,L.M. et al. A focal domain of extreme demethylation within D4Z4 in FSHD2. Neurology 80, 392-399 (2013).
73. Huichalaf,C., Micheloni,S., Ferri,G., Caccia,R., & Gabellini,D. DNA methylation analysis of the macrosatellite repeat associated with FSHD muscular dystrophy at single nucleotide level. PLoS. ONE. 9, e115278 (2014).
74. Jones,T.I. et al. Individual epigenetic status of the pathogenic D4Z4 macrosatellite correlates with

disease in facioscapulohumeral muscular dystrophy. Clin. Epigenetics. 7, 37 (2015).
75. Zeng,W. et al. Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats in facioscapulohumeral dystrophy (FSHD). PLoS. Genet 7, e1000559 (2009).
76. 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).
77. 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).
78. Blewitt,M.E. et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat. Genet. 40, 663-669 (2008).
79. Zhao,J., Sun,B.K., Erwin,J.A., Song,J.J., & Lee,J.T. Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. Science 322, 750-756 (2008).
80. Gendrel,A.V. et al. Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev. Cell 23, 265-279 (2012).
81. Gendrel,A.V. et al. Epigenetic functions of smchd1 repress gene clusters on the inactive x chromosome and on autosomes. Mol. Cell Biol. 33, 3150-3165 (2013).
82. Nozawa,R.S. et al. Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nat. Struct. Mol. Biol. 20, 566-573 (2013).
83. Mould,A.W. et al. Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. Epigenetics. Chromatin. 6, 19 (2013).
84. Larsen,M. et al. Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. Eur. J. Hum. Genet. 23, 808-816 (2015).
85. 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).
86. Cabianca,D.S. et al. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell 149, 819-831 (2012).
87. Lim,J.W. et al. DICER/AGO-dependent epigenetic silencing of D4Z4 repeats enhanced by exogenous siRNA suggests mechanisms and therapies for FSHD. Hum. Mol. Genet.(2015).
88. Masny,P.S. et al. Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? Hum Mol. Genet 13, 1857-1871 (2004).
89. Petrov,A. et al. Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. Proc. Natl. Acad. Sci. U. S. A 103, 6982-6987 (2006).
90. Petrov,A. et al. A nuclear matrix attachment site in the 4q35 locus has an enhancer-blocking activity in vivo: implications for the facio-scapulo-humeral dystrophy. Genome Res.(2007).
91. Bodega,B. et al. Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. BMC. Biol. 7, 41 (2009).
92. Pirozhkova,I. et al. A functional role for 4qA/B in the structural rearrangement of the 4q35 region and in the regulation of FRG1 and ANT1 in facioscapulohumeral dystrophy. PLoS. ONE. 3, e3389 (2008).
93. Rijkers,T. et al. FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. J Med Genet 41, 826-836 (2004).
94. van Deutekom,J.C.T. et al. Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. Hum Mol Genet 5, 581-590 (1996).
95. 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).
96. Jiang,G. et al. Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. Hum Mol. Genet 12, 2909-2921 (2003).
97. Klooster,R. et al. Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. Eur. J. Hum. Genet. 17, 1615-1624 (2009).
98. Arashiro,P. et al. Transcriptional regulation differs in affected facioscapulohumeral muscular dystrophy patients compared to asymptomatic related carriers. Proc. Natl. Acad. Sci. U. S. A 106, 6220-6225 (2009).

99. Celegato,B. et al. Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre differentiation and a general deregulation of MyoD-dependent genes. Proteomics. 6, 5303-5321 (2006).
100. Cheli,S. et al. Expression profiling of FSHD-1 and FSHD-2 cells during myogenic differentiation evidences common and distinctive gene dysregulation patterns. PLoS. ONE. 6, e20966 (2011).
101. Gabellini,D. et al. Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. Nature 439, 973-977 (2006).
102. Hanel,M.L., Wuebbles,R.D., & Jones,P.L. Muscular dystrophy candidate gene FRG1 is critical for muscle development. Dev. Dyn. 238, 1502-1512 (2009).
103. Liu,Q., Jones,T.I., Tang,V.W., Brieher,W.M., & Jones,P.L. Facioscapulohumeral muscular dystrophy region gene-1 (FRG-1) is an actin-bundling protein associated with muscle-attachment sites. J. Cell Sci. 123, 1116-1123 (2010).
104. Masny,P.S. et al. Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei. Eur. J. Hum. Genet. 18, 448-456 (2010).
105. Osborne,R.J., Welle,S., Venance,S.L., Thornton,C.A., & Tawil,R. Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. Neurology 68, 569-577 (2007).
106. Pistoni,M. et al. Rbfox1 downregulation and altered calpain 3 splicing by FRG1 in a mouse model of Facioscapulohumeral muscular dystrophy (FSHD). PLoS. Genet. 9, e1003186 (2013).
107. Sancisi,V. et al. Altered Tnnt3 characterizes selective weakness of fast fibers in mice overexpressing FSHD region gene 1 (FRG1). Am. J. Physiol Regul. Integr. Comp Physiol 306, R124-R137 (2014).
108. Sun,C.Y. et al. Facioscapulohumeral muscular dystrophy region gene 1 is a dynamic RNA-associated and actin-bundling protein. J. Mol. Biol. 411, 397-416 (2011).
109. Tsumagari,K. et al. Gene expression during normal and FSHD myogenesis. BMC. Med. Genomics 4, 67 (2011).
110. Winokur,S.T. et al. Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. Hum Mol. Genet 12, 2895-2907 (2003).
111. Xu,X. et al. DNaseI hypersensitivity at gene-poor, FSH dystrophy-linked 4q35.2. Nucleic Acids Res. 37, 7381-7393 (2009).
112. Yao,Z. et al. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. Hum. Mol. Genet.(2014).
113. Mariot,V. et al. Correlation between low FAT1 expression and early affected muscle in facioscapulohumeral muscular dystrophy. Ann Neurol. 78, 387-400 (2015).
114. Puppo,F. et al. Identification of variants in the 4q35 gene FAT1 in patients with a facioscapulohumeral dystrophy-like phenotype. Hum. Mutat. 36, 443-453 (2015).
115. Caruso,N. et al. Deregulation of the protocadherin gene FAT1 alters muscle shapes: implications for the pathogenesis of facioscapulohumeral dystrophy. PLoS. Genet. 9, e1003550 (2013).
116. Snider,L. et al. RNA Transcripts, miRNA-sized Fragments, and Proteins Produced from D4Z4 Units: New Candidates for the Pathophysiology of Facioscapulohumeral Dystrophy. Hum. Mol. Genet 18, 2414-2430 (2009).
117. Lemmers,R.J. et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329, 1650-1653 (2010).
118. Jones,T.I. et al. Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. Hum. Mol. Genet. 21, 4419-4430 (2012).
119. Snider,L. et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS. Genet. 6, e1001181 (2010).
120. Geng,L.N. et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Dev. Cell 22, 38-51 (2012).
121. Young,J.M. et al. DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. PLoS. Genet. 9, e1003947 (2013).
122. Ferreboeuf,M. et al. DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. Hum. Mol. Genet. 23, 171-181 (2014).
123. Rickard,A.M., Petek,L.M., & Miller,D.G. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Hum. Mol.

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

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

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