Cis and trans modifiers in facioscapulohumeral muscular dystrophy

Sikrová, D.

Citation

Sikrová, D. (2022, December 14). *Cis and trans modifiers in facioscapulohumeral muscular dystrophy*. Retrieved from https://hdl.handle.net/1887/3497752

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3497752

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

CHAPTER 1

General Introduction

Epigenetic regulation of mammalian genome

The functional output of our genome at any given time is not only determined by the information encoded in its genetic layer, i.e. the DNA sequence itself, but also by different *epigenetic layers* which help in its interpretation. Epigenetic layers mean factors or modifications controlling the stability and inheritance of gene expression patterns across different cell divisions or generations which are not the result of changes in the DNA sequence itself. This genome-wide epigenetic information is known as the *epigenome*. A growing number of diseases stem from mutations that alter different parts of the epigenome. Such mutations can affect chromatin configuration *in cis* or alter either the abundance or activity of epigenetic modifiers leading to epigenetic changes *in trans,* ultimately affecting gene expression. This thesis focuses on one such disease called Facioscapulohumeral muscular dystrophy (FSHD), in which epigenetic deregulation of a specific macrosatellite repeat array favors the expression of its embedded gene. Therefore, the first part of the Introduction is devoted to describing different epigenetic mechanisms involved in transcriptional regulation of the genome, mainly focusing on the epigenetic silencing of repetitive elements. The second part of the Introduction will discuss the genetic and epigenetic etiology of FSHD.

DNA methylation

One of the epigenetic layers that regulates the genome is the direct modification of DNA bases. The most studied DNA modification in mammals is methylation of the 5th carbon of cytosine (*5mC*), generally referred to as DNA methylation. The existence of another form of methylation in genomic DNA, that of the 6th carbon of adenine (*6mA*), while being prevalent in prokaryotes [1], remains disputable in mammals [2]. In the mouse and human genome, around 5% of all cytosines are methylated [3] making 5mC a relatively abundant modification which is therefore sometimes referred to as the fifth DNA base (next to adenine, guanine, cytidine and thymine). The 5mC is usually found in a CpG dinucleotide context resulting in two 5mCs positioned diagonally to each other on opposite DNA strands [4]. The occurrence of 5mC in this symmetrical CpG context allows for faithful reproduction of the methylation information during DNA replication as a 5mC on the mother strand serves as a template for the methylation machinery to methylate the cytosine on a newly replicated daughter strand [5]. In most cell types, except for specific stages during embryogenesis [6] and gametogenesis [7], around 80% of CpGs are methylated [8]. These are typically isolated CpGs, while the remaining unmethylated CpGs are usually clustered in *CpG islands* (*CGIs*), genomic regions which contain a higher density of CpGs than one would expect by chance [9]. These unmethylated CGIs are predominantly associated with promoters of active genes. The exceptions to this are promoter CGIs of three classes of genes for which life-long stable silencing mediated by promoter methylation in somatic tissues is crucial. These include genes on the inactive X chromosome [10], imprinted genes [11] and germline genes [12]. In addition, CGIs found within gene bodies (intragenic) or between genes (intergenic), collectively termed as 'orphan' CGIs, can also become methylated during development or are methylated in a tissue-specific manner [13]. Therefore, the lack of methylation at CGIs is often associated with active transcriptional start sites (TSSs) while their

methylation is associated with gene silencing [14]. In contrast to hypomethylated CGIs of active promoters, the bodies of actively transcribed genes are enriched with methylated CpGs which prevent spurious intragenic transcription initiation events [15,16]. Interestingly, around 20% of gene-associated CGIs in the human genome are absent from the homologous mouse genes and further analysis suggested that both humans and mice are losing CGIs over evolutionary time [17]. This can be explained by the hypermutability of 5mC since it is prone to spontaneous deamination resulting in C to T transitions in the genome, therefore resulting in progressive loss of CpGs through acquired transitions [18].[1]

The DNA methylation patterns are generally stably maintained in somatic cells. However, the somatic epigenome poses a major barrier to sexual reproduction and preparation for a next generation requires its resetting. The reconfiguration of genome-wide DNA methylation patterns happens in two steps during specific developmental time windows (reviewed here [19]). First, somatic methylation signatures are removed in the primordial germ cells (PGCs) and germ cell-specific as well as sex-specific signatures are established during later stages of germ cell development enabling meiotic maturation and subsequent fertilization [20]. After fertilization, the epigenome of a newly formed zygote becomes reprogrammed during subsequent cell divisions to erase gamete-specific signatures inherited from the oocyte and the sperm [21,22]. The DNA methylation erasure is completed at the pre-implantation blastocyst stage after which it is ready for the initiation of the embryonic developmental program and setting up lineage specific methylation profiles.

The life-cycle of DNA methylation is carried out by a collection of enzymes which can be considered based on their action either as the *writers* of this mark (DNA methyltransferases, DNMTs) or *erasers* (ten-eleven translocation enzymes, TETs). In mammals, writers belong to a family of DNMTs consisting of four catalytically active members (DNMT1, DNMT3A, DNMT3B and rodent-specific Dnmt3c) and one catalytically inactive member (DNMT3L), each of which evolved to perform largely non-overlapping functions [23].[2]

DNMT1 was the first DNMT identified [24] and for a long time recognized as a canonical maintenance DNMT because of its high affinity for hemi-methylated DNA [25,26] and its role in re-establishing CpG methylation patterns after DNA replication. However, this longstanding view has been challenged over time as some studies reported that it can also act *in vitro* on unmethylated DNA substrates, albeit with lower efficiency [27,28] and its *de novo* methylation activity was reported in oocytes outside the context of DNA replication [29] as well as during replication-coupled methylation maintenance [30,31]. Whether this *de novo* methylation

1 5mC does not exist in genomes of several widely used model organisms such as *Caenorhabditis elegans*, fission yeasts and bakers' yeasts and is found at very low levels only during early stages of embryonic development of *Drosophila*.

2 After identification of DNMT1, the second candidate mammalian DNMT gene was found which shared a sequence homology with DNMT1 and was named DNMT2 [318]. However, it turned out that it does not have properties that can be expected of a DNA methyltransferase as it has very low affinity towards double stranded DNA and it primarily localizes to the cytoplasm instead of the nucleus. Indeed, it was demonstrated that it is an RNA methyltransferase responsible for methylating the 38th cytosine residue in anticodon loop of certain tRNAs [319].

potential of DNMT1 is biologically relevant or creates only aberrant unspecific byproducts was addressed recently when it was demonstrated that murine Dnmt1 displays *de novo* methylation activity targeted at specific classes of retrotransposons [32]. Similarly, the strict classification of DNMT3A and DNMT3B as purely DNA methylation establishing DNMTs requires fine-tuning. Both DNMT3A and 3B are highly expressed during early embryonic development as well as in mouse embryonic stem cells. Upon differentiation, their expression dramatically declines which is in line with the assumption that they are then dispensable [33–35]. Nevertheless, they are essential for the long-term maintenance of DNA methylation imprints at least in mouse embryonic stem cells [36] and somatic inactivating mutations in DNMT3A have been reported in hematologic malignancies [37]. Moreover, DNMT3B isoforms without catalytic activity can act as accessory factors aiding DNMT1 activity in somatic cells [38]. This suggests that DNMTs could indeed work cooperatively to maintain methylation fidelity and that both DNMT3A and DNMT3B are also important in (some) somatic cells.

Dnmt3c and Dnmt3l, the two most recent evolutionary additions to the family of DNMTs, are involved in mammalian reproduction. *Dnmt3c* arose through a tandem duplication of the *Dnmt3b* gene specifically in the *Muroidea* lineage and is expressed only in male germ cells where it selectively methylates the promoters of evolutionarily young transposable elements thus ensuring their repression. This specialized Dnmt3c activity is required for male fertility [39]. Dnmt3l is a catalytically inactive cofactor that stimulates methyltransferase activities of Dnmt3a and Dnmt3b [40,41]. Similarly to Dnmt3c, its loss leads to male sterility due to the reactivation of certain classes of retrotransposons [42]. In addition, it is also required for proper oogenesis by helping Dnmt3a to establish maternal methylation imprints [43].

As mentioned before, mammalian genomes undergo two rounds of epigenomic resetting during which the majority of 5mC marks are removed. This can be accomplished by passive loss of DNA methylation through replication [44] or by its active removal by the TET family of proteins [45–47]. In mammals, the TET family consists of three members, TET1, TET2 and TET3, all of which catalyze the erasure of the 5mC modification in three sequential oxidation steps by generating 5-hydroxymethylcytosine (5hmC) which is further converted to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [47]. The last two products can be excised from DNA by the base excision repair pathway thus re-installing in unmodified cytosine bases [48,49].

Post-translational modifications of core histone tails

Another epigenetic layer is achieved by organizing DNA into a higher order structure known as *chromatin* which, amongst others, serves as a docking platform for other regulatory molecules. The smallest unit of chromatin is the *nucleosome*, which consists of 147 base pairs of DNA wrapped around two copies of four different histone proteins, usually H2A, H2B, H3 and H4 [50]. Other chromosome region-specific histone subvariants may occur such as centromeric protein A (CENPA) which is a centromere-associated H3 histone variant required for kinetochore assembly and for proper chromosome segregation during cell

division [51]. The N-terminal tails of histones extrude from the nucleosome and can undergo a variety of reversible and dynamic post-translational modifications (PTMs), predominantly at lysine or arginine residues [52]. Specific modifications can either directly influence the chromatin accessibility by changing the local charge or serve to recruit chromatin factors that either condense (repress) or relax (activate) chromatin. Furthermore, the local chromatin composition is a major determinant of the transcriptional activity of a locus.

Chromatin was initially divided into euchromatin and heterochromatin based on a different cytological staining density during interphase, where less compact and brighter stained regions were termed euchromatin and more compact densely stained regions were termed heterochromatin [53]. Nowadays, from a molecular perspective we recognize at least two types of heterochromatin, facultative and constitutive, which have distinct regulatory functions and are enriched for different proteins and protein modifications, but both result in transcriptional attenuation. Facultative heterochromatin is marked by Polycomb Group (PcG) proteins which exist in two separate protein complexes, PRC1 and PRC2, catalyzing monoubiquitination of lysine 119 on histone H2A (H2AK119Ub1) [54] or trimethylation of lysine 27 on histone H3 (H3K27me3) [55], respectively. In mammals, facultative heterochromatin regulates primarily the spatiotemporal expression of developmental genes [56,57] and the formation of the inactive X chromosome in females [58,59]. In contrast, constitutive heterochromatin is mainly formed at gene-poor and repeat-rich regions. A histone mark typical for constitutive heterochromatin is the trimethylation of lysine 9 on histone H3 (H3K9me3), which can be recognized by heterochromatin protein 1 (HP1) homologues promoting further chromatin compaction [60,61]. In mammals, deposition of H3K9me is catalyzed by at least six (five in humans) different H3K9 lysine methyltransferases (KMT) forming three distinct enzymatic systems, namely Suv39h1/Suv39h2, Eset1/Eset2 (Eset1 corresponds to human SETDB1) and G9a/Glp (also known as Ehmt2 and Ehmt1, respectively). Each of them targets different genomic regions. Suv39h1 and Suv39h2 are functionally redundant and primarily responsible for the deposition of H3K9me3 at centromeric and pericentromeric repeats [62,63]. Eset1 is important for silencing of endogenous and newly introduced retroviruses [64,65] and for the establishment of X inactivation [66,67], while G9a/Glp are important for early lineage commitment and permanent silencing of genes driving pluripotency [68–70]. Compound loss of all six H3K9 KMTs in mouse embryonic fibroblasts leads to complete dissolution of heterochromatin, transcriptional de-repression of nearly all families of repeat elements and genomic instability [71], marking the importance of H3K9me3 for transcriptional silencing and maintaining genome integrity. On the other hand, euchromatin represents an accessible chromatin state and contains transcriptionally active genes together with their regulatory elements. Promoters of actively transcribed genes are typically enriched for trimethylated lysine 4 on histone H3 (H3K4me3), while the bodies of actively transcribed genes are enriched for trimethylated lysine 36 on histone H3 (H3K36me3) [72,73].

Crosstalk between DNA methylation and histone modifications

The existence of bidirectional crosstalk between histone modifications and DNA methylation was initially hypothesized based on the observation of genome-wide colocalization of particular histone modifications with DNA methylation [74]. For example, DNA methylation is generally excluded from promoters of actively transcribed genes, whereas the bodies of actively transcribed genes are enriched for DNA methylation. These gene elements are also distinctly enriched for specific histone modifications such as H3K4me3 and H3K36me3, respectively. A mechanistic explanation was subsequently provided for these observations. Both DNMT3A and 3B enzymes contain apart from their methyltransferase domain also two chromatin reader domains, namely the ADD and PWWP domain, which allow for a direct readout of the H3 histone tail and thus help to regulate the deposition of DNA methylation in a chromatin state-aware manner. Specifically, the ADD domain recognizes unmodified H3K4 but is repelled by the increasing number of methyl moieties at K4 [75,76] and thus H3K4me3 acts as a shield against DNA methylation deposition. On the other hand, the PWWP domain directly binds to H3K36me3 and in this way the DNMT3 enzymes are targeted to the bodies of actively transcribed genes [77,78].

A peculiar example is the relationship between DNA methylation and H3K27me3. While they have been shown to co-occupy many CpG-poor regions, they are mutually exclusive at the CGI promoters of PcG target genes [79,80], whose promoters are co-marked by H4K4me3 and H3K27me3 in embryonic stem cells and at the E6.5 epiblast stage before lineage differentiation. This *bivalent* active/repressive state allows for these genes to be readily activated or repressed during lineage specification [81,82]. Furthermore, DNA methylation was shown to interfere with the PRC2 recognition of unmodified as well as H3K27me3-modified nucleosomes *in vitro* [83]. Consistent with this, the loss of DNA methylation results in genome-wide redistribution of H3K27me3 to regions which would be otherwise DNA methylated [80,84], while titrating it away from its native targets leading to their insufficient repression [84]. However, the fact that some regions can adopt both DNA methylation and H3K27me3 suggests that their coexistence might be context-dependent and that under certain unknown circumstances, the avoidance behavior of PRC2 towards methylated DNA sites can be overcome.

Perhaps the tightest cooperative relationship is between DNA methylation and H3K9me3 [74,85]. Together, these modifications enforce a more stable silenced chromatin state and aid each other in its initial establishment and its mitotic propagation. For example, H3K9me3 controls the maintenance of DNA methylation. DNMT1 recognizes H3K9me3 both directly via its RFTS domain [86] as well as indirectly through cooperation with its interacting factor UHRF1 [87,88]. These additional mechanisms, next to hemimethylated DNA itself, boost the fidelity of maintaining DNA methylation patterns in the H3K9me3 context. In addition, DNA methylation at major satellites, which form pericentric heterochromatin in mouse embryonic stem cells, is dependent on Suv39h1/2-mediated deposition of H3K9me3, which is in turn recognized by HP1 proteins facilitating the recruitment of *Dnmt3b* [89]. An earlier study using

immunofluorescence as a readout for H3K9me3 occupancy at major satellites claimed that H3K9me3 is retained at pericentric regions upon loss of DNA methylation in mouse embryonic stem cells lacking all three Dnmts (Dnmt1, Dnmt3a and Dnmt3b) [90]. However, a later study discovered a significant reduction in H3K9me3 in the same cells using quantitative mass spectrometry [91]. The reliance of H3K9me3 on DNA methylation became even more apparent when studying cells lacking DNMT1 which show reduced levels of H3K9me3 at pericentric regions [92]. However, it should be noted that pericentric heterochromatin represents a specific example of crosstalk between H3K9me3 and DNA methylation which cannot be automatically translated to other heterochromatic regions co-enriched for these two marks.

Epigenetic regulation of repetitive elements

Repetitive elements, which comprise over half of the human genome [93], can have a profound effect on gene regulation [94,95], chromosome (in)stability (reviewed here [96]), human health (reviewed here [97]) and can even drive species-specific adaptations [98]. Yet, their detailed annotation in human genome assemblies was lacking for a long time due to their repetitiveness. Recent advances in long-read sequencing technologies inspired a new consortium to follow in the footsteps of the Human Genome Project, which mission is to deliver gapless telomere-to-telomere chromosomes assemblies at base pair resolution (hence the name telomere-to-telomere or T2T consortium) and to generate the first complete assembly of the human genome since its first draft was published over 20 years ago [99,100]. The majority of these gaps are comprised of repetitive elements and several pre-prints are already starting to appear delivering comprehensive genetic and epigenetic annotations of previously known as well as newly discovered repetitive elements [93,101,102].

Classification of repetitive elements in mammalian genome

Based on the genomic organization, eukaryotic repeats can be classified into two classes: interspersed repeats and tandem repeats.

Interspersed repeats typically comprise transposable elements (TE) which can be further subdivided based on their mode of moving in the genome. Class I elements or retrotransposons work in a “copy and paste” mechanism in which they replicate themselves by reverse transcription and insert a new copy at the target site. Therefore, their copy number amplifies over time. In contrast, class II elements work in a “cut and paste” mechanism when a specialized enzyme, a transposase or a recombinase usually encoded by the TE itself, mediates its excision from the current position followed by insertion into a new genomic location. Size-wise, TEs are relatively short sequences (50 bp to 12 kb), however, it is their sheer number that can in some extreme cases make up almost 85% of the genome such in the case of wheat ((IWGSC) et al., 2018).[3] Although most TEs have lost their ability

3 One of the largest Class II TEs (up to 100 kb) was recently discovered in a model fungus *Podospora anserina* [320]. The authors whimsically name the new TE *“Enterprise”* as its transported “cargo” is a block of meiotic driver genes termed *Spoks (spore killing)*.

to mobilize further, some of these elements are still capable of "hopping" around causing insertional mutagenesis which can yield a neutral, deleterious or even advantageous outcome (reviewed here [104]). However, there is growing evidence that their main function in the genome is rather their capacity to influence the expression of neighbouring genes. Such function might look selfish at the first glance, however, there are specific instances when the host took advantage of this phenomenon and co-opted it into its own gene regulatory network. One of the most studied occurrences of transposon-mediated regulation of gene expression is during a zygotic genome activation when the embryonic transcriptional program is kickstarted [105]. One particular type of TEs, the murine endogenous retrovirus with leucine tRNA primer (MERVL), has been discovered as being central to this process in mice [106] with its human counterpart human ERVL (HERVL) serving the same function [107,108].

In contrast to interspersed repeats, tandem repeats are comprised of repetitive units which are usually organized in head-to-tail orientation and include multi-copy gene families (such as ribosomal DNA) and satellite repeats.[4] Depending on the length of the satellite unit, satellites can be classified in micro- (2-6 bp), mini- (10-100 bp) or macrosatellites (up to several kb). Tandem repeats often form structural elements of chromosomes which are important for genomic stability such as centromeric [109] and telomeric regions [109] or represent a boundary element driving higher-order chromosome architecture such as the DXZ4 macrosatellite repeat at the inactive X chromosome [110,111]. Furthermore, they show a high degree of polymorphism in their sequence, structure and their copy number [112] all of which can contribute to inter- as well as intra-species phenotypic variation, especially when a tandem repeat in question is formed by gene duplications [113–115]. However, copy number variation of some tandem repeats can also negatively impact human health if they alter the coding region or influence gene expression *in cis*. The most notable examples are microsatellite expansion disorders, in which the microsatellite copy-number increases in successive generations and once it reaches a certain threshold becomes unstable. Over 50 genetic disorders have been linked so far to such repeat expansions (reviewed here [97]). In addition, a reduction in tandem repeat copy number can also be detrimental as is the case for Facioscapulohumeral muscular dystrophy [116] which will be further discussed in the second part of this Introduction.

Regulators of the repeats' epigenetic state

Already in the early 90s, it was observed that integrating an increasing number of gene copies in tandem in plant genomes does not yield higher transcriptional output as compared to the single copy integration event [117,118], a surprisingly counterintuitive result as one would expect. Moreover, multiple tandem insertions are associated with higher DNA methylation, a mark that was as capable of modulating gene expression [119,120]. This phenomenon, when

4 The term satellite DNA was first coined by Pech et al. [321] and was referring to a DNA component that produces a specific *satellite* band that separates from the main DNA band during a caesium chloride density gradient centrifugation. As the density of DNA is a function of its base composition and highly homogeneous or repetitive sequences have this base ratio skewed, this will result in a different migration pattern along the density gradient compared to bulk DNA. The satellite DNA from the Pech paper was later confirmed to belong to centromeric AT-rich alpha satellites.

repetitive regions trigger *cis* heterochromatinization in a copy number-dependent manner, was termed *repeat-induced gene silencing* (RIGS) and was later also confirmed to operate in mammals [112]. Initially, RIGS was proposed to evolve as a protective mechanism of eukaryotic genomes against integration-prone foreign DNA elements such as viruses or transposons [121]. However, RIGS was later also recognized as a natural mechanism for regulation of expression of nearby genes [112] thus representing a case of position effect variegation (PEV) [122].

PEV refers to a phenomenon when a gene is placed (intentionally or by chance) in proximity to or within a heterochromatic region resulting in its stochastic transcriptional silencing (i.e. variegated expression) due to heterochromatin spreading into the juxtaposed locus. The pioneer of this field was Hermann Joseph Muller in the early 20th century who derived several *Drosophila* mutant lines with different variegated phenotypes due to X-ray induced chromosomal rearrangements [123]. Muller's discovery of PEV kickstarted new studies focusing on how gene expression is influenced by its chromatin environment. Numerous studies revealed many trans-acting modifiers which influence the probability of heterochromatin spreading and thus gene silencing (reviewed here [124]). Factors that increase the mutant phenotype were termed enhancers of variegation, while factors that decrease the mutant phenotype were coined suppressors of variegation. Later, these modifiers have been defined as either structural components of heterochromatin, enzymes that modify chromatin or as nuclear structural components and many of the identified factors were found to be conserved also in mammals [125].

Similarly to genetic screens to identify modifiers of PEV in *Drosophila*, analogous approaches were used to identify factors involved in RIGS in mammals using loci which show variegated phenotypes under genetic homogeneity thus allowing for uncovering factors whose mutation would skew the phenotypic spectrum one or the other way. Such loci, whose epigenetic state can intergenerationally switch from active to repressed, were termed *metastable epialleles* and were studied to capture both 1) the epigenetic basis of the phenotypes associated with these alleles and 2) the stochasticity of their epigenetic state.

The most relevant screen for this thesis is the one conducted in the Emma Whitelaw lab to search for modifiers of variegated multicopy transgene expression [126,127]. This screen used a transgenic inbred mouse line (GFP1 line) carrying a random integration of a transgene array consisting of ~11 copies of a construct in which the α-globin promoter and enhancer drive expression of a GFP reporter resulting in its variegated expression in red blood cells. Importantly, the variegated expression of this transgene is stable throughout generations culminating at around 55% of red blood cells being GFP positive [128]. A shift in the percentage of GFP-expressing red blood cells was used as a read-out in the offspring born to ENU-treated males and mutant alleles which showed enhanced or suppressed variegated expression were designated as *Modifiers of Murine Metastable Epialleles* (*Mommes*) [126]. This screen yielded more than 40 of such dominant mutant alleles (termed MommeDX or ^{MDX}, where "D" denotes a dominant screen and "X" a number referring to an allele in order

in which it was identified) and revealed previously known (e.g. *Dnmt1*, *Dnmt3b*, *Setdb1*, *Suv39h1*) as well as novel genes (e.g. *Smchd1*, *Rlf*, *D14Abb1e, Morc3*) and even genes (e.g. *eIF3h, Hbb*) without a clear link to epigenetic processes (full gene list reviewed here [125]). Therefore, the interpretation of the results should be carried out in light of confounding factors inherent to the screen design such as transgene integration site, tissue-specific phenotypic read-out (potential identification of genes affecting hematopoiesis in this case), the introduction of a foreign DNA sequence which potentially triggers similar host genome responses as retrotransposons or integration-prone viruses, genetic background (i.e. mouse strain) in which the screen was conducted, parent-of-origin effects (screening progenies of ENU-treated males) or the actual structure of the transgene (tandem repeat in this case).

Indeed, several commonalities between retrotransposon and transgene silencing were pointed out previously [129,130]. In line with that, several *MommeD* alleles were found to also modulate the Agouti viable yellow (*A*^{vy}) locus, in which a spontaneous insertion of an intracisternal A particle (IAP), belonging to a Class II endogenous retrovirus (ERV) family, was shown to modulate the expression of *in cis Agouti* gene responsible for, among others, coat colour [131]. *Agouti* is normally expressed only transiently from a hair cycle-specific promoter and is responsible for the deposition of yellow and black pigment during mouse hair growth [132]. The inserted IAP creates a cryptic promoter that drives continuous expression of *Agouti* leading to a completely yellow coat. However, partial or full silencing of this IAP by e.g. DNA methylation leads to mottled or wild-type-like brown fur color. Specifically, S*mchd1*^{MD1}, *Dnmt1*^{MD2}, *Trim28*^{MD9} and *Setdb1*^{MD13} alleles resulted in a shift to a yellow fur (i.e. failure to repress the IAP), while *Smarca5*^{MD4}, *Rlf*^{MD8} and *Wiz*^{MD30} alleles resulted in a shift to a brown fur [126,127,133]. Interestingly, the resulting phenotypic shifts in the coat color due to these alleles were concordant with their effect on GFP transgene expression suggesting that they play the same role, either repressing or activating, at these two loci. Furthermore, the phenotypic outcome of the coat color and thus *Agouti* gene expression reversely correlated with the DNA methylation status at the 5’ long terminal repeat (5’ LTR) of the inserted IAP [134]. Similar observation was made also for the methylation status and expression of the GFP transgene and when combined with concrete *MommeD* alleles, namely S*mchd1*^{MD1}, *Rlf*^{MD8}, *Dnmt3b*^{MD14}, *Dnmt1*^{MD32} and *Nrf1*^{MD46} [127,135–137]. However, some *MommeD* alleles such as *Hdac1*^{MD5}, *Baz1b*^{MD10}, *Wiz*^{MD30} and *Rif1*^{MD18} showed no changes in DNA methylation of the GFP transgene and yet showed changes in expression suggesting that these factors are involved in layers of epigenetic regulation unrelated to DNA methylation [127,136].

Follow-up studies employing reverse genetics approaches uncovered that genes underlying *Momme* alleles are involved in epigenetic regulation of diverse endogenous loci including different types of repeats. For instance, Dnmt3b seems to be particularly specialized in the establishment of DNA methylation at pericentromeric [138] and subtelomeric repetitive regions [139] and is also responsible for silencing genes on the inactive X chromosome [140]. Similarly, Suv39h1/Suv39h2 mediate deposition of H3K9me2/me3 at pericentromeric [63,141] and subtelomeric repeats [142]. In contrast, a trio of Mommes (Morc3, Trim28 and Setdb1) is involved in the repression of IAP elements [143–145].

Mutations in *Momme* genes have also been linked to diverse human diseases and syndromes. The most worthy to mention in the context of this thesis are mutations in two genes, *DNMT3B* and *SMCHD1*, as their heterozygous mutations are associated with Facioscapulohumeral muscular dystrophy [146,147]. In addition, biallelic mutations in *DNMT3B* cause the rare Immunodeficiency, Centromeric region instability and Facial anomalies type 1 (ICF1) syndrome [148]. Similarly to *DNMT3B*, mutations in *SMCHD1* can also yield a pleiotropic phenotypic outcome since they are also causative of Bosma Arhinia Microphthalmia Syndrome (BAMS), a very rare condition, with less than 50 patients being reported, characterized by nasal, ocular and reproductive defects [149].

Facioscapulohumeral muscular dystrophy

The FSHD locus

The road to elucidating the root cause of the disease took over 100 years since its first description as a distinct disease entity as FSH type muscular dystrophy by the French physicians Louis Landouzy and Joseph Dejerine in 1884.[5] Studies in the early 90s helped to narrow down the search for the FSHD locus by linking the disease to an *EcoRI* genomic fragment which was polymorphic in length and detected by a DNA probe (p13E-11) mapping to 4q35 [150–152]. Specifically, *EcoRI* fragments usually larger than 28 kb were detected in non-affected individuals, while shorter fragments between 14 – 28 kb co-segregated with FSHD [151]. Interestingly, even after 30 years, a slightly modified approach is being used to this day for routine FSHD diagnostics (Figure 1A) [153]. Soon after, it was shown that the locus in question contains a tandemly repeated sequence dubbed D4Z4[6] which consists of copies of a 3.3 kb repeat unit defined by a *KpnI* restriction site (Figure 1A). Similar repeat sequences map to other locations in the human genome [154,155] with a highly homologous tandem repeat present at 10q26 that can vary between 1-100 units in the population [156,157]. However, the reason why shortening of this particular repeat only on chromosome 4 causes FSHD remained elusive for a long time. The initial hypothesis to explain the chromosome 4 specificity of the disease was inspired by the PEV mechanism and proposed that longer D4Z4 repeats tend to adopt a more heterochromatic structure which would spread *in cis.* In FSHD, due to the reduced D4Z4 copy-number, this heterochromatinization would be partially lost leading to inappropriate expression of nearby gene(s) [154].

---

5 Initially, FSHD was referred to as Landouzy-Dejerine muscular dystrophy, however, some disputes were raised over who should be acknowledged for the priority of describing this disease as a separate clinical entity [322] as the very first description of the disease was done by the French neurologist Duchenne de Boulogne. The peculiar pattern of muscle weakness first affecting distal leg muscles while skipping proximal leg muscles was first recognized by German neurologist Wilhelm Heinrich Erb. However, it was Landouzy and Dejerine who 'absorbed' prior clinical descriptions of Duchenne and Erb together with observations from their casuistry into one FSH type of muscular dystrophy.

6 The name 'D4Z4' is derived from a nomenclature system which was used for DNA regions of unknown significance during the human genome project: D stands for DNA, 4 stands for chromosome 4, Z indicates a repetitive sequence and 4 is a assigned serial number based on the submission order. Hence, the homologous repeat on chromosome 10 cannot be truly termed D4Z4 and was unfortunately never assigned a D10Z serial number.

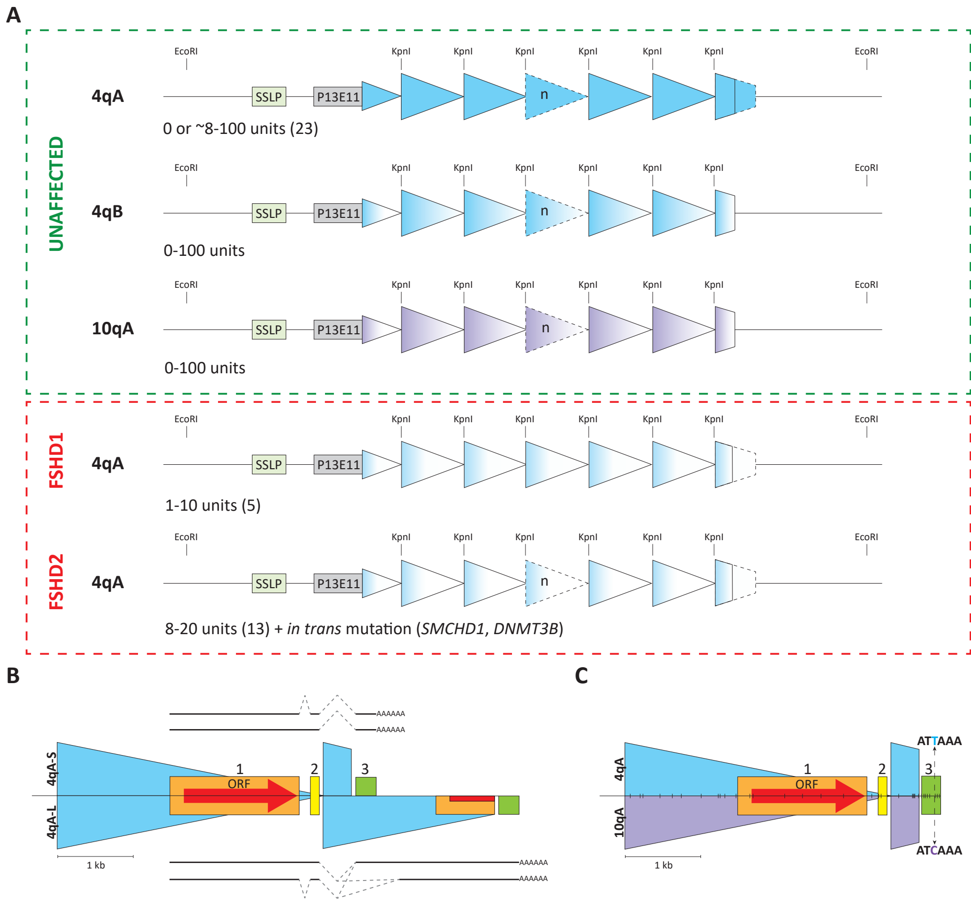

Figure 1. Genetics and epigenetics of D4Z4 in FSHD. A) Schematic representation of homologous D4Z4 repeats present on the long arms of chromosomes 4 and 10 as illustrated by blue (4q) and violet (10q) triangles organized in tandem. Each triangle represents one repeat unit as defined by a *KpnI* restriction site and the first and the last repeat units are incomplete. Proximal to the repeat are two sequence elements that are utilized for the assessment of D4Z4 haplotype (SSLP) and for the determination of copy-number by Southern blot (P13E11). The 4qA/B typing is performed with Southern blotting using probes that hybridize further downstream of the distal *EcoRI* site and are thus not depicted. Unaffected individuals either carry variably sized 4qB D4Z4 repeats, whose epigenetic state is not relevant for FSHD or D4Z4 repeats on the 4qA background whose length is sufficient for proper epigenetic silencing of the repeat (usually more than 8 units). Similarly to 4qB D4Z4 repeats, any copy-number and its associated epigenetic state of 10qA D4Z4 repeats is irrelevant for FSHD pathogenesis. Epigenetic dysregulation of 4qA D4Z4 repeats is caused either by its reduction in copy-number (FSHD1; 1-10 units) or by co-inheritance of intermediately-sized (8-20 units) repeat together with a mutation in (at least) one of its *trans* modifiers (FSHD2). Numbers in the brackets next to the designated repeat ranges refer to a median size of the repeat based on their prevalence in the European population. The color gradient of triangles represents varying levels of 4q/10q D4Z4 epigenetic repression found in healthy and FSHD individuals (the lighter the color the lower the DNA methylation levels and thus repression). **B)** Specifically the 4qA161 D4Z4 haplotype, which is the most frequent haplotype in FSHD individuals, can end in two different forms (4qA-S or 4qA-L) depending on the break-point in the most distal incomplete repeat unit. The two forms give rise to different *DUX4* mRNA isoforms differing in their 3'UTR. The *DUX4* mRNA isoforms are further diversified by the optional splicing of exon 2. Three *DUX4* exons are represented by brackets of different colors (orange, yellow and green) with the DUX4 ORF being fully contained within exon

1 is represented by the thick red arrow. **C)** Comparison of the terminal 4qA (blue) and 10qA (violet) repeat unit depicted as a mirror image. The position of the *DUX4* PAS in exon 3 is depicted by a dashed arrow line highlighting the 4qA/10qA SNP in the *DUX4* PAS. Other distinguishing SNPs between 4qA and 10qA D4Z4 are marked on the line that separates 4q from 10q.

In line with this hypothesis, it was shown that D4Z4 repeat contractions leads to DNA hypomethylation of this locus [158] which is accompanied by reduced levels of H3K9me3 [159] and thus possibly affecting the regulation of several candidate FSHD genes proximal to 4q D4Z4 repeat, including *ANT1, FRG1*, *TUBB4Q* and *FRG2* [160–163]. However, while one study documented upregulation of some candidate genes in FSHD muscle biopsies [160], other studies reported no changes in mRNA expression of these genes between FSHD and control cases arguing against the PEV hypothesis [164–167]. Furthermore, it was shown that at least one D4Z4 unit is required for disease development [168] suggesting that FSHD is tightly associated with the D4Z4 repeat itself rather than its surrounding chromosomal region. Indeed, every D4Z4 unit was found to contain an open reading frame encoding for a putative double homeobox protein termed *DUX4* [169–171]. Thus, another hypothesis was put forward suggesting that the epigenetic de-repression of contracted D4Z4 repeats leads to the expression of this repeat-encoded *DUX4* gene [171]. But it was not until 2010 that a unifying mechanism for FSHD-associated *DUX4* expression was presented, which confirmed the latter hypothesis (Figure 1B) [172]. Furthermore, the possible involvement of other candidate genes on chromosome 4 was challenged by describing FSHD individuals having atypical D4Z4 rearrangements. This includes cases with large proximal deletions occurring *in cis* to the contracted 4q D4Z4 repeat sometimes encompassing *FRG2* and *TUBB4Q* [173,174], as well as FSHD cases with interchromosomal rearrangements between 4q35 and 10q26 resulting in a hybrid, contracted D4Z4 repeat at 10q26 and leading to a physical separation of the contracted D4Z4 repeat partially of 4q origin and other 4q FSHD candidate genes [175].

Nowadays, two genetically distinct forms of FSHD are recognized, FSHD1 (OMIM #158900) and FSHD2 (OMIM #158901), however, both involve epigenetic de-repression of the 4q D4Z4 repeat associated with *DUX4* expression in skeletal muscle. They do, however, arise by distinct genetic mechanisms (Figure 1A). While in FSHD1 it is the contraction of the D4Z4 repeat that causes loss of its heterochromatinization, in FSHD2, it is mutation(s) *in trans* in genes which play a role in establishing or maintaining the heterochromatic state of D4Z4 [146,147,176]. Another notable difference is that in the latter case, the chromatin state of the 10q26 D4Z4 repeat is also affected whereas in FSHD1 the chromatin changes are constricted to the contracted 4q allele only [158,177].

Clinical presentation

FSHD is regarded one of the more common muscular dystrophies in adults with an estimated prevalence ranging between 0.8 and 4.6 per 100,000 [178]. From a clinical perspective, FSHD1 and FSHD2 cases are indistinguishable [179,180]. Age at onset as well as clinical severity varies extensively from patient to patient with one-fifth of individuals with an FSHD-sized D4Z4

repeat within FSHD1 families remaining asymptomatic [181]. However, this number might be lower as an important factor for disease presentation is age. While FSHD affects both sexes, some FSHD1 family studies reported that females are more likely to be less severely affected or asymptomatic than males despite carrying an identical D4Z4 repeat array [182–184]. Evidence for such sex bias in FSHD2 families is weaker, but this can be also due to the relatively small sample sizes as compared to studies in FSHD1 families [179,180]. In "classical" FSHD cases, the first symptoms become apparent in the second decade. Being a slowly progressive disease, individuals are often diagnosed relatively late in life with a median age at diagnosis of around 40 [185]. Although FSHD patients typically have a normal life expectancy, their fitness decreases over time with almost one-fourth of cases requiring a wheelchair by the age of 50 [186,187]. In addition, there is an infantile form of the disease, representing around 4% of all FSHD cases, with more severe symptoms and faster progression [188,189].

The typical clinical presentation of FSHD includes early involvement of the muscles of the face, shoulder girdle, and upper arms, often in an asymmetric manner. As the disease progresses, lower extremities can also become affected, starting with the distal muscles and later involving more proximal leg muscles. Apart from muscle involvement, extramuscular manifestations have been reported in FSHD. These include ophthalmological abnormalities [190,191], high-frequency hearing loss [192–194] and CNS abnormalities like epilepsy and mental retardation. These extramuscular manifestations are more prominent in the more severe infant onset form of the disease [189,195,196].

Cis modifiers in FSHD

FSHD1 is due to a contraction of at least one 4q D4Z4 repeat to a size of 1 – 10 units. However, for a 4q D4Z4 contraction to result in FSHD, it needs to occur on a specific 4q subtelomeric genetic background [197,198]. Based on sequence variations immediately distal to D4Z4, 4q subtelomeres were initially subgrouped into two main allelic variants, 4qA and 4qB, which are equally common in the European population [199]. Interestingly, contractions of D4Z4 on 4qB alleles have not been observed to cause FSHD suggesting that some 4qA-specific sequences underlie 4qA pathogenicity or that 4qB alleles contain protective genetic elements [197,200]. The most noteworthy difference between 4qA and 4qB alleles is the presence of a 260 bp sequence immediately distal to D4Z4 on the 4qA background, termed pLAM which is followed by a β-satellite repeat [200]. Furthermore, the 10q subtelomere shows a high degree of sequence homology (98%) to 4qA [199] and thus is referred to as 10qA. Yet, D4Z4 contractions at 10qA are not pathogenic [156]. In addition, even though all FSHD D4Z4 alleles are of the 4qA type, not all contracted 4qA D4Z4 alleles result in FSHD [198]. A worldwide population study further revealed nine subtelomeric 4qA haplotypes based on several sequence polymorphisms found within and flanking the repeat [201]. One of the main sequence features defining the haplotype is a simple sequence length polymorphism (SSLP) located approximately 3 kb proximal to both 4q and 10q D4Z4 repeats (Figure 1A). All haplotypes are thus defined by their chromosomal location (4 or 10), distal variant (A or B) and SSLP (between 157 and 182 bp). Out of nine defined 4qA haplotypes, only three (4A159, 4A161

and 4A168) have been indisputably associated with FSHD [172], while the classification of 4A166 as FSHD-permissive haplotype remains inconclusive due to contradicting findings [198,202,203]. The most predominant haplotype found in FSHD in Europe is unsurprisingly 4A161 [198] as it is also the most frequent FSHD-permissive haplotype found in the control European population [201]. In addition, this haplotype shows another degree of variability in its distal end. The most distal unit in 4qA haplotypes is incomplete and usually formed by a proximal 0.33 kb of the D4Z4 unit. However, the 4A161 haplotype can instead of this short (S) end, also terminate with a longer incomplete unit of 1.6 kb which is then referred to as the long (L) variant (Figure 1B) [172,204]. Nevertheless, in regards to FSHD, contractions of either 4A161 variant (4A161S or 4A161L) are disease-causing [172,204]. A near-perfect explanation for the 4qA linkage of FSHD came with a seminal study providing a functional explanation for the pathogenicity of certain 4qA haplotypes [172]. Two earlier studies already showed that it is almost exclusively the most distal full-length D4Z4 unit that can express fully processed and stabilized *DUX4* mRNA by using a polyadenylation signal (PAS) in the pLAM region (Figure 1B) [165,205]. Extending on that, Lemmers et al. showed that the 10qA pLAM region contains a SNP in the sequence corresponding to the *DUX4* PAS sequence found on disease-permissive 4qA haplotypes (4qA: AT**T**AAA -> 10qA: AT**C**AAA) corrupting its functionality (Figure 1C) [172]. The only haplotype that currently remains unresolved is 4A166 as its disease association remains unclear. While it does contain a functional 4qA *DUX4* PAS sequence, the majority of other SNPs in its pLAM region are more 10qA-like [172]. Therefore, more exhaustive population studies, as well as functional dissection of the effects of 4qA/10qA sequence polymorphisms on *DUX4* expression, are required to further fine-tune our understanding of the genetic predisposition to FSHD.

One of the most important *cis* modifiers in FSHD is the D4Z4 copy number itself as even inheritance of a contracted allele on a disease-permissive haplotype is not 100% predictive of disease penetrance. This is a key aspect of FSHD, i.e. that the phenotypic outcome is on a continuous quantitative scale rather than categorized by simple binary qualitative groups (non-affected vs affected) as the clinical severity is often inversely correlated with D4Z4 copy-number [206–210]. Individuals with shorter alleles typically have an early onset whereas carriers of FSHD alleles in the upper size range (7 – 10 units) present with milder symptoms or even remain life-long asymptomatic [181,208]. The latter cases make prenatal diagnosis and genetic counselling challenging as it is associated with high levels of uncertainty [211]. Furthermore, differences have been observed in the D4Z4 length distribution in non-affected as well as in FSHD1 cohorts of ethnically different populations. Particularly, the median size of 4q D4Z4 repeats in unaffected Asian populations is smaller than in the Caucasian population [212] and such difference in distribution is also observed for the size of the contracted allele in FSHD1 cohorts from Asia and Europe (median of 3 – 4 units vs 5 – 6 units, respectively) [213–215]. Therefore, it seems that Asian populations are less permissive to FSHD. The factors behind this reduced permissiveness, either of environmental or genetic origin or both, remain to be elucidated but could be instrumental to our understanding of *DUX4* expression regulation. However, it should be noted that we only operate with the assumption that shorter D4Z4 alleles yield higher DUX4 levels since no larger-scale correlative studies have been conducted regarding a relationship between D4Z4 copy-number and *DUX4* expression.

Other *cis* modifiers proximal to D4Z4 have been proposed to be involved in FSHD by influencing the muscle-specific phenotype [216] and D4Z4 de-repression [217]. For example, two *DUX4* myogenic enhancers (DME1 and DME2) have been described upstream of D4Z4 which by looping to the *DUX4* promoter are supposed to enhance its expression specifically in skeletal muscle cells but not in skin fibroblasts [216]. Another D4Z4 proximal genetic element was found to give rise to a long non-coding RNA (DBE-T) which was shown to be upregulated in FSHD and to be responsible for the recruitment of the Trithorax group protein ASH1L to the D4Z4 repeat causing chromatin remodeling with subsequent *DUX4* de-repression [217]. However, the identification of individuals presenting with FSHD carrying proximally extended D4Z4 deletions encompassing aforementioned *cis* sequence elements challenged their relevance for FSHD pathology [174,218–220].

Trans **modifiers in FSHD**

The first indication about the existence of possible *trans* modifiers in FSHD came with the recognition that around 5% of FSHD cases do not carry a contracted D4Z4 allele and yet show DNA hypomethylation of the repeat [158,177,179,221]. These cases were classified as contraction-independent type 2 FSHD (FSHD2). In these FSHD2 individuals, both 4q and 10q D4Z4 repeats were found to be hypomethylated as opposed to only the contracted repeat in FSHD1, suggestive of the involvement of a *trans* factor affecting D4Z4 methylation [146,158]. The introduction of whole-exome sequencing (WES) into clinical genetics practice accelerated the identification of heterozygous mutations in the *SMCHD1* gene, which co-segregated with D4Z4 hypomethylation and, if combined with 4qA allele, resulted in FSHD [146]. The spectrum of *SMCHD1* mutations identified in FSHD2 include nonsense, missense, splicing-affecting mutations and even larger genomic deletions encompassing the entire *SMCHD1* locus resulting in SMCHD1 hemizygosity (detailed overview of SMCHD1 mutations is reviewed here [222]). Therefore, the current consensus is that D4Z4 hypomethylation in FSHD2 is due to reduced amounts of functional SMCHD1 protein. In addition, the nature of SMCHD1 mutations correlates with residual DNA methylation level at D4Z4. Specifically, heterozygous *SMCHD1* mutations which preserve the open reading frame (usually missense mutations) show more pronounced D4Z4 hypomethylation and thus seem to be more deleterious than heterozygous *SMCHD1* mutations which disrupt the SMCHD1 open reading frame and result in lower SMCHD1 protein levels [223]. One possible explanation for this observation is that SMCHD1 forms homodimers [224,225] and thus haploinsufficiency of SMCHD1 would reduce the number of functional WT SMCHD1 homodimers to 50% as compared to the WT situation, while the dominant negative effect of missense mutations would lead to only 25% of WT functional SMCHD1 homodimers if we assume that the mutant SMCHD1 monomer can form heterodimers with WT SMCHD1 monomer (25% WT:WT, 50% WT:MUT, 25% MUT:MUT). Furthermore, missense mutations positioned at the N-terminus of the protein were shown to have a greater effect on D4Z4 methylation than those at the C-terminus [223]. Similarly to the previously observed rough inverse correlation between the length of contracted D4Z4 repeat, its methylation and clinical severity in FSHD1 individuals, a significant correlation was also found for the length of the shortest 4qA D4Z4 allele and its DNA methylation in FSHD2

individuals [223], suggesting that also in FSHD2 cases the repeat length plays a modifying role for its epigenetic state. This is further supported by the observation that the median size of a 4qA D4Z4 repeat in European FSHD2 individuals was found to be shorter (13 units) than the median size in the control European population (23 units) [223], which would suggest that SMCHD1 mutation carriers with longer permissive alleles do not develop FSHD or its manifestation is very mild [226], creating a reservoir of *SMCHD1* mutations in the population. At that time, an enigmatic exception was a group of FSHD2 individuals with longer 4qA D4Z4 repeats (up to 70 units). However, closer genetic examination revealed that the majority of these cases have a duplication allele consisting of the earlier diagnosed longer repeat array followed by a smaller FSHD-sized repeat array duplication that is likely permissive to *DUX4* expression [227]. Mutations in SMCHD1 have been also reported to modify the disease penetrance as well as severity in FSHD1 cases [180,226,228–230]. Interestingly, only individuals with upper-sized D4Z4 repeats (7 – 10 units) in combination with an *SMCHD1* mutation were described, which prompts the question if a combination of a shorter D4Z4 repeat with *SMCHD1* mutation is under negative selection pressure and incompatible with life, or that it is the sheer rareness of this combination that has prevented its reporting thus far. Alternatively, such bias in the findings could be due to the existing FSHD diagnostics practice, when cases suspected of FSHD are first undergoing D4Z4 sizing and if a contracted 4qA allele is identified no further screening for *SMCHD1* mutations is undertaken. On the other hand, enough comparative methylation studies between FSHD1 and FSHD2 were reported [223,231,232] that would potentially reveal D4Z4 hypomethylation outliers in FSHD1 cohorts sparking the motivation for identifying possible *in trans* mutations in these individuals.

Even almost 10 years after the first description of the association of *SMCHD1* mutations with FSHD2, we still do not have a clear mechanistic explanation of how germline SMCHD1 loss-of-function relates to the observed D4Z4 hypomethylation in somatic cells. SMCHD1 is expressed in somatic cells where it binds to D4Z4 and its binding is reduced in somatic cells derived from FSHD2 individuals [146]. Furthermore, depletion of SMCHD1 either in FSHD1 or FSHD2 skeletal muscle cells leads to further *DUX4* transcriptional de-repression suggesting that it aids in D4Z4 silencing also in somatic cells with an already compromised D4Z4 chromatin state [226,233]. In addition, increasing SMCHD1 levels in FSHD1 and FSHD2 muscle cells, either by its overexpression or by its mutation correction in the case of FSHD2, was shown to result in significant *DUX4* downregulation [230,233]. Although complete transcriptional repression of *DUX4* was not achieved, low levels of DUX4 have been detected also in unaffected relatives of FSHD subjects [234], thus absolute *DUX4* somatic silencing might not be necessary to achieve clinical benefit. Such observations inspired a discussion about the possibility of modulating SMCHD1 levels as a general therapeutic strategy for FSHD.

SMCHD1 is encoded by 48 exons giving rise to a 2005 aa-long protein in humans whereas the mouse ortholog of SMCHD1 is 2 aa longer. It contains two main functional domains: an N-terminal ATPase and a C-terminal hinge domain which are connected by a flexible linker [235]. Both the hinge domain as well as the ATPase domain are required for SMCHD1 homodimerization [224,236,237]. The hinge domain was further shown to interact with nucleic

acids [236,238]. More recently, two extra domains flanking the ATPase module were characterized, namely the N-terminal ubiquitin-like (UBL) and transducer domains, which aid in ATPase dimer stabilization during ATP hydrolysis [225,237]. Mouse Smchd1 was initially identified, as mentioned before, in the dominant screen for modifiers of murine metastable epialleles and follow up studies showed that its homozygous loss-of-function results in female-specific embryonic lethality due to a failure in X inactivation [235]. In contrast to the active X (Xa), which is in its higher-order structure more similar to autosomes by being partitioned into smaller topologically associated domains (TADs), the inactive X (Xi) is folded into two megadomains with limited short-range intra-chromosomal interactions both in mouse and humans [239–241]. Smchd1 is a key factor in this folding process as its loss results in increased short-range interactions over the Xi due to enhanced CCCTC-binding factor (CTCF) binding leading to Xi decompaction [242–244]. A similar gain of Ctcf has been observed also at clustered protocadherins and Hox gene cluster in the absence of Smchd1 [238,242]. Both clusters were already known to be transcriptionally sensitive to the loss of Smchd1 [238,245,246]. In addition, Smchd1 has been shown to regulate the expression of monoallelically expressed genes such as selected genes within the *Snrpn* cluster [238,245,246]. Furthermore, Smchd1 was recently shown to act as a maternal effect gene in the mouse, when the maternal *Smchd1* allele is the only source for Smchd1 production until at least the 32-cell stage and is required for the imprinted expression of 10 genes [247]. However, whether human SMCHD1 expression is regulated similarly during human pre-implantation development remains to be elucidated. But even if so, it might be of little relevance for SMCHD1-mediated D4Z4 epigenetic regulation as both maternal and paternal transmission of an *SMCHD1* mutation has been documented in FSHD2 families with no apparent methylation or clinical differences between the sexes [146].

Nowadays, it is estimated that at least 85% of FSHD2 cases are explained by mutations in SMCHD1 [146,223]. This number is likely higher as mutations in cases suspected of FSHD2 are typically identified by WES or *SMCHD1* exon sequencing and thus potential deep intronic *SMCHD1* mutations go unnoticed. Indeed, one such FSHD2 family has been reported recently [230]. Nevertheless, further studies into other *trans* modifiers identified two families in which a heterozygous mutation in *DNMT3B* was co-segregating with D4Z4 hypomethylation and was shown to modify the disease penetrance in family members carrying a relatively short permissive D4Z4 repeat [147]. Identifying *DNMT3B* mutations was not surprising as recessive mutations in *DNMT3B* were previously shown to cause ICF1 syndrome, in which the D4Z4 repeat is also hypomethylated [248]. Interestingly, despite SMCHD1 and DNMT3B both converging at the epigenetic regulation of D4Z4, other repeats which are hypomethylated in ICF1 individuals, such as pericentromeric satellite repeat types II and III and the NBL2 macrosatellite repeat, are not hypomethylated in FSHD2 individuals with *SMCHD1* mutations [221] suggesting that these two factors do not always co-regulate the same genomic regions or alternatively, that aforementioned repeats are less sensitive to SMCHD1 than to DNMT3B dysfunction.

The epigenetic makeup of D4Z4 in somatic cells consists of high levels of DNA methylation and H3K9me3 which both ensure *DUX4* repression as treating cells either with 5-aza-2'-

deoxycytidine (Aza), a deoxycytidine analogue which cannot be methylated by DNMTs, or chaetocin, a non-specific inhibitor of histone methyltransferases, results in *DUX4* transcriptional de-repression [159,217,249,250]. Complementary experiments to reduce these marks by lowering the protein levels of both DNMT1 and DNMT3B or SUV39H1 confirmed their importance in somatic D4Z4 silencing [250,251]. Furthermore, both DNA methylation and H3K9me3 are reduced at D4Z4 in FSHD1 and FSHD2 not only in skeletal muscle cells but also in cells of other tissues derived from different germ layers [159,202,232,251,252], suggesting that either the establishment of these repressive marks was impaired before the multi-lineage differentiation or the mechanism of their maintenance is impaired in all tissues. Apart from this, additional changes in histone marks have been documented such as increase in H3K4me2/3 in both FSHD1 and FSHD2 [253,254] and a specific increase in H3K27me3 in FSHD2 individuals with SMCHD1 mutations [233]. In addition, several studies identified other factors of D4Z4 chromatin in somatic cells which contribute to the repression of this locus. First, HP1γ (CBX3) and the cohesin complex were found to be associated with 4q and 10q D4Z4 repeats in control somatic cells and their recruitment was shown to be dependent on H3K9me3 [159]. A follow-up study also explored the heterochromatin state of D4Z4-like sequences which are present at other chromosomes, mainly at acrocentric chromosomes [250]. Interestingly, neither DNA methylation, H3K9me3, HP1γ nor the cohesin complex was affected at those regions in both FSHD1 and FSHD2 cells with *SMCHD1* mutation. This prompts the question of why the function of SMCHD1 is restricted to 4q and 10q D4Z4 repeats and to what degree D4Z4-like sequences on other chromosomes are different from 4q/10q D4Z4 repeats in their chromatin regulation [250]. Limited data is available on these sequences but one noticeable difference is that different from 4q and 10q, these repeat sequences do not seem to form homogeneous tandem repeat arrays [155]. This also raises another concern regarding chromatin studies that employ PCR amplification to investigate the association of specific chromatin factors with D4Z4 as our findings and conclusions about FSHD-relevant chromatin changes are only as good as the specificity of the primers or probes we use.

Recently, an unbiased proteomic study identified 261 proteins as being enriched at D4Z4 in control myoblasts, including components of the NuRD and CAF1 complexes and interestingly also several *Momme* factors, namely PBRM1, RIF1, SMARCA4, SMARCA5, UHRF1, HDAC1, SETDB1 and TRIM28 [255]. It remains to be investigated if all of these protein factors act in a parallel or redundant fashion, if and how they contribute to disease penetrance, and if any of these repressive components can be employed for future therapeutic strategies aiming at re-repression of D4Z4 in FSHD skeletal muscle cells.

Lastly, as mentioned earlier, the epigenetic changes to D4Z4 are not specific for skeletal muscles of FSHD individuals but are also present in other somatic tissues [159,202,232,251,252] Therefore, the apparent predominant muscle phenotype in FSHD raises the question why other tissues are not affected. Either other tissues are somehow resistant to DUX4 toxicity or more likely, they do not even express *DUX4*. We know that for example cultured fibroblasts derived from skin biopsies of FSHD individuals do not express *DUX4* at all and that *DUX4*

expression can only be detected after their forced transdifferentiation into myotubes [147]. In addition, neither *DUX4* nor its transcriptional signature was detected in the RNA-seq study of whole blood from FSHD individuals [256], although EBV-transformed peripheral blood leucocytes derived from FSHD individuals do recapitulate both D4Z4 epigenetic as well as *DUX4* transcriptional changes of FSHD myoblasts [202,254,257]. Even more peculiar are the inter-muscular differences as some muscle groups seem to be more prone to *DUX4* expression than others, which could explain their differential involvement in FSHD [258,259]. Furthermore, as *DUX4* expression increases during myogenic differentiation [233], it seems that the epigenetic changes at D4Z4 only create an environment permissive for *DUX4* expression and that muscle-specific factors or intracellular changes during myogenesis or EBV-transformation are required to initiate *DUX4* transcription. Previously, it has been shown that protein levels of SMCHD1 decrease during myogenic differentiation [233] and therefore, one could hypothesize that reduced availability of some D4Z4 repressors might contribute to this muscle-restricted misexpression of *DUX4.*

Conservation of *DUX4* and consequences of its expression in skeletal muscle

DUX4 belongs to the *DUX* gene family, which includes among others also the intronless *Dux* gene present in the mouse genome [260]. Both *DUX4* and *Dux* are hypothesized to have arisen independently by a retrotransposition-related expansion of an ancestral *DUXC* gene and are organized in a tandem array, although not at a syntenic location. Furthermore, the single repeat unit of the *Dux*-forming macrosatellite is longer than that of D4Z4 (4.9 kb vs 3.3 kb) [261]. DUX4, as well as Dux, contain two highly homologous N-terminal homeodomains as well as a conserved C-terminal transcriptional transactivation domain [262]. Only recently it has been shown that they are indeed functional homologs by regulating the zygotic genome activation (ZGA), a process after fertilization during which the transcription of newly combined genetic material starts for the very first time [107,108,263]. Expression of *DUX4* mRNA was shown to peak during the 4-cell cleavage stage, whereas *Dux* mRNA expression peaks already at the 2-cell stage, both corresponding to their species-specific ZGA timepoints [107,108]. Both Dux and DUX4 activate the transcription of ZGA genes by directly binding to their promoters through their homeodomains [263]. Furthermore, both proteins bind also to a specific family of retrotransposons (MERVL in mice and HERVL in humans), which serve as alternative promoters of some cleavage-specific genes during ZGA [108,263]. However, how *Dux*/*DUX4* expression itself is so swiftly regulated during this short time window is still poorly understood. It is also not known whether a failure in *DUX4* silencing in FSHD individuals begins at this point (although it should be noted that it has not been established if the cleavage-specific *DUX4* transcripts are specifically of only 4q D4Z4 origin).

Interestingly, certain culturing conditions allow mESCs to fluctuate between pluripotent (ICM-like state of blastocyst) and totipotent state (2-cell blastomere-like cleavage stage) and at any given moment around 1% of the mESC population is in this 2-cell-like stage [106]. These 2C-like cells recapitulate many attributes of the 2-cell stage blastomeres including their transcriptome which is characteristic of the ZGA phase [106,264], chromatin accessibility

landscape [108], high core histone mobility [265] and the capacity to contribute to extra-embryonic tissues [106]. The conversion of mESCs to 2C-like cells is regulated both by a variety of chromatin factors [264,266–275]. Furthermore, induction of 2C-like cells was shown to strongly depend on Dux as ectopic expression of *Dux* forces mESCs into 2-cell like cells [108]. In line with this, *Dux* knock-out prevents mESCs from their conversion to 2-cell like cells [107]. However, follow up studies challenged the notion of Dux being an essential driver of ZGA, since *Dux* zygotic knock-out embryos can give rise to viable pups, albeit with decreased developmental potential due to delayed ZGA onset [276–278]. Therefore, Dux seems to help in synchronizing the ZGA, but probably other yet unidentified factors in addition to Dux are involved in the onset and propagation of the ZGA process *in vivo*. In contrast, efficient silencing of *Dux* past the 2-cell stage seems to be of bigger importance for proper embryonic development as its sustained expression impedes the 2-cell exit and causes embryonic arrest [272,279]. The emerging recent model suggests that *Dux* repression is achieved by tethering its genomic locus to the perinucleolar heterochromatin space by the LINE1/Nucleolin/Trim28 complex both in mESCs and early embryos [272,280,281]. Nucleoli are membrane-less nuclear organelles, whose boundaries are thought to be defined by liquid-liquid phase separation and are a place for rRNA and ribosome biogenesis (reviewed here Lafontaine et al., 2020). Both 2C-like cells and 2-cell blastomeres possess yet immature more compact nucleoli sometimes referred to as nucleolar precursor bodies (NPBs) which exhibit low rRNA transcriptional output [282]. Following fertilization, the rRNA levels sharply increase from the 2-cell stage onwards to the blastocyst stage to cope with the embryonic need for a sufficient amount of translational apparatuses [280]. This rRNA transcriptional change is associated with nucleolar maturation and with the formation of perinucleolar heterochromatin. Thus, it seems that the embryonic need for increased translational output and the termination of the ZGA phase were naturally co-opted into one regulatory mechanism during early genome spatial reorganization when activation of rRNA synthesis shuts down expression of *Dux* for cells to continue into the next cleavage stages. This also explains a prior counterintuitive observation that the LINE1/Nucleolin/Trim28 complex while positively regulating rRNA expression negatively regulates expression of *Dux* [272]. It remains to be investigated if a similar mechanism also operates in *DUX4* silencing during human embryonic development. Interestingly, other D4Z4-like sequences are present on the short arms of acrocentric chromosomes [155,250] which are responsible for the nucleolar organization (reviewed here McStay, 2016) and similarly, also the 4q D4Z4 repeat has been observed to preferentially localize either to the nuclear or nucleolar periphery in somatic cells [284–286]. Despite that, the nuclear localization of contracted 4q D4Z4 was not changed in cells from FSHD1 individuals which could otherwise explain the sporadic transcriptional activation of *DUX4*.

Overexpression of DUX4 in cultured myoblasts elicits a transcriptional response similar to what was identified during human ZGA, including upregulation of specific retroelements and cleavage-specific genes, which are also misexpressed in FSHD cultured muscle cells as well as in FSHD biopsies [287–290]. Endogenous *DUX4* expression is a rather rare event in FSHD 2D muscle cell cultures, with only around 1:200-1000 of nuclei expressing *DUX4* at any

given moment, depending on the differentiation stage, culture conditions and donor [291–294]. However, since mononuclear muscle precursor cells fuse during myogenic differentiation to form multinucleated myofibers in which they eventually share their cytoplasmic space, even one nucleus expressing *DUX4* can "infect" its neighboring nuclei with DUX4 protein upon its translation in the cytoplasm. This can be visualized by staining for DUX4 protein in differentiated muscle cells, typically creating a DUX4 staining gradient across clustered nuclei that is getting weaker with the distance of the acceptor nucleus from the donor *DUX4* expressing nucleus (Figure 2). Since DUX4 is a transcription factor, the consequence of this is that even transcriptomes of nuclei that do not express *DUX4* themselves will be rewired by DUX4, thus explaining the observed easier mRNA detection of DUX4 target genes than *DUX4* itself [289]. This was also confirmed by single-nucleus RNA-seq (snRNA-seq), when many more nuclei show expression of DUX4 target genes while *DUX4* mRNA itself is in majority of cases not detectable in them [295]. For this reason, some of the DUX4 target genes have been considered as potential biomarkers instead of direct detection of *DUX4* [296]. Interestingly, during the course of differentiation, *DUX4* expression and expression of its target genes become discordant, when nuclei can remain expressing DUX4 target genes even after the nucleus is no more DUX4 protein positive [297]. One plausible explanation for this phenomenon is that DUX4 initiates expression of, among others, a cascade of transcription factors including its gene orthologue *DUXA*, which can then contribute to their perduring expression [295,297]. In addition, DUX4 was shown to induce changes in the chromatin landscape of its target genes by at least two distinct mechanisms, which sensitize these genes for their re-activation or sustained expression. First, DUX4 was shown to recruit the p300/CBP H3K27 acetyltransferase complex to its target DNA sites via its C-terminal transactivation domain, which helps chromatin opening of these loci for transcription [298]. Indeed, treating DUX4-expressing cells with a selective p300 inhibitor was sufficient to attenuate transcriptomic changes known to be elicited by DUX4 [299]. Second, DUX4 induces expression of two histone variants, namely H3.X and H3.Y, which get incorporated into gene bodies of DUX4 target genes resulting in a more relaxed chromatin configuration [300]. As some evidence suggests that endogenous *DUX4* expression occurs in bursts [292], after initial DUX4-mediated re-setting of the chromatin, following bursts of *DUX4* expression can lead to enhanced activation of its target genes as their chromatin is already more accessible for transcription [300].

Apart from the DUX4-induced transcriptional changes, DUX4 has been linked to other disruptive processes which might contribute to its myopathic effect. High levels of DUX4 can cause apoptosis in skeletal muscle cells via distinct mechanisms including activation of caspase 3/7- [301] and p53-mediated apoptotic pathways [302], induction of hypoxia signaling [303], increasing sensitivity to oxidative stress [304,305], upregulation of the pro-apoptotic factor MYC [306] and/or activation of the double-stranded RNA (dsRNA) response pathway [306,307]. On the other hand, low expression of DUX4 in myogenic cells was shown to negatively affect their myogenic differentiation potential *in vitro* [308]. Homeodomains of DUX4 display high amino acid sequence homology to a homeodomain of the muscle specific transcription factor PAX7 [309]. PAX7 is strictly expressed in myogenic precursor satellite cells and is required

for their proliferative capacity, thus ensuring a regenerative potential of skeletal muscle tissue [310,311]. Because of this homology, it was hypothesized that DUX4 might interfere with the transcriptional program activated by PAX7 thus leading to impaired myogenesis [309]. In line with this competitive inhibition model, it was demonstrated that overexpression of Pax7 in murine C2C12 myogenic cells counteracts the DUX4-induced cytotoxic effect in a dose-dependent manner [312]. However, DUX4 and PAX7 have non-overlapping expression patterns during normal myogenic differentiation, which argues against this competitive model [313]. Despite that, a recent analysis of different gene expression studies from muscle biopsies showed that PAX7 downstream genes (so-called PAX7 target gene score) are indeed repressed in FSHD samples compared to controls [314]. Intriguingly, the PAX7 score was proposed to be a more robust discriminator of FSHD-affected muscles than the expression of *DUX4* or its target genes [315] and it was shown that this score is a good biomarker for FSHD progression over a period of at least 1 year [316], therefore offering a possibility of being utilized for monitoring of FSHD development in future clinical trials as a reliable biomarker is still missing. A more recent transcriptomic study conducted on FSHD muscle biopsies suggested that DUX4 and PAX7 expression signatures might rather mark different stages of the disease (van Den Heuvel et al., 2022).

Scope of the thesis

Research presented in this thesis focuses both on *cis* and *trans* contributors to Facioscapulohumeral muscular dystrophy. In **chapter 2**, we employ a genome editing tool termed adenine base editing to efficiently mutate the somatic polyadenylation signal of *DUX4,* an important *cis* modifier in FSHD to test this approach as a possible future FSHD gene therapy. In **chapter 3**, we describe a proband with clinical symptoms consistent with FSHD that carries a homozygous out-of-frame deletion in exon 2 of the *LRIF1* gene combined with a disease permissive D4Z4 allele of 13 units. We confirmed that the D4Z4 epigenetic profile in the proband's cells exhibits perturbations as described for FSHD2 cases and we detect also the expression of *DUX4* itself in the proband's cells, thus uncovering a novel *trans* modifier in FSHD. We further extend this finding in **chapter 4**, where we study the action of LRIF1 together with its interacting partner SMCHD1 in D4Z4 repression in human somatic cells with distinct D4Z4 chromatin contexts. And lastly, in **chapter 5**, we explore the role of all three FSHD2 genes by performing loss of function studies in mESCs and we uncover the assistance of Lrif1 in the repression of mouse *Dux*, which is a functional homologue of human *DUX4*.

References

1. VANYUSHIN, B. F., BELOZERSKY, A. N., KOKURINA, N. A. & KADIROVA, D. X. 5-Methylcytosine and 6-Methylaminopurine in Bacterial DNA. *Nat. 1968 2185146* **218**, 1066–1067 (1968).
2. Li, X. *et al.* The exploration of N6-deoxyadenosine methylation in mammalian genomes. *Protein Cell 2021* 1–13 (2021) doi:10.1007/S13238-021-00866-3.
3. Nowialis, P. *et al.* Catalytically inactive Dnmt3b rescues mouse embryonic development by accessory and repressive functions. *Nat. Commun.* **10**, (2019).
4. Doskočil, J. & Šorm, F. Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. *Biochim. Biophys. Acta* **55**, 953–959 (1962).
5. Bird, A. P. Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. *J. Mol. Biol.* **118**, 49–60 (1978).
6. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nat. 2012 4847394* **484**, 339–344 (2012).
7. Seisenberger, S. *et al.* The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Mol. Cell* **48**, 849–862 (2012).
8. Ehrlich, M. *et al.* Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.* **10**, 2709 (1982).
9. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nat. 1986 3216067* **321**, 209–213 (1986).
10. Norris, D. P., Brockdorff, N. & Rastan, S. Methylation status of CpG-rich islands on active and inactive mouse X chromosomes. *Mamm. Genome 1991 12* **1**, 78–83 (1991).
11. SanMiguel, J. M. & Bartolomei, M. S. DNA methylation dynamics of genomic imprinting in mouse development. *Biol. Reprod.* **99**, 252–262 (2018).
12. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet. 2010 4212* **42**, 1093–1100 (2010).
13. Zeng, J., Nagrajan, H. K. & Yi, S. V. Fundamental diversity of human CpG islands at multiple biological levels. *http://dx.doi.org/10.4161/epi.27654* **9**, 483–491 (2014).
14. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010 (2011).
15. Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. **520**, 243–247 (2015).
16. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. **543**, 72–77 (2017).
17. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11995 (1993).
18. Shen, J. C., Rideout, W. M., 3rd & Jones, P. A. The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* **22**, 972 (1994).
19. Reik, W., Dean, W. & Walter, J. Epigenetic reprogramming in mammalian development. *Science (80-.).* **293**, 1089–1093 (2001).
20. Guibert, S., Forné, T. & Weber, M. Global profiling of DNA methylation erasure in mouse primordial germ cells. *Genome Res.* **22**, 633–641 (2012).
21. Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nat. 2014 5117511* **511**, 611–615 (2014).
22. Hill, P. W. S. *et al.* Epigenetic reprogramming enables the transition from primordial germ cell to gonocyte. *Nat. 2018 5557696* **555**, 392–396 (2018).
23. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol. 2019 2010* **20**, 590–607 (2019).
24. Bestor, T., Laudano, A., Mattaliano, R. & Ingram, V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells: The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.* **203**, 971–983 (1988).

25. Hermann, A., Goyal, R. & Jeltsch, A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase Methylates DNA Processively with High Preference for Hemimethylated Target Sites *. *J. Biol. Chem.* **279**, 48350–48359 (2004).
26. Pradhan, S. *et al.* Baculovirus-mediated expression and characterization of the full-length murine DNA methyltransferase. *Nucleic Acids Res.* **25**, 4666–4673 (1997).
27. Vilkaitis, G., Suetake, I., Klimašauskas, S. & Tajima, S. Processive Methylation of Hemimethylated CpG Sites by Mouse Dnmt1 DNA Methyltransferase. *J. Biol. Chem.* **280**, 64–72 (2005).
28. Fatemi, M., Hermann, A., Pradhan, S. & Jeltsch, A. The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J. Mol. Biol.* **309**, 1189–1199 (2001).
29. Li, Y. *et al.* Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nat. 2018 5647734* **564**, 136–140 (2018).
30. Wang, Q. *et al.* Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance. *Nat. Genet. 2020 528* **52**, 828–839 (2020).
31. Ming, X. *et al.* Kinetics and mechanisms of mitotic inheritance of DNA methylation and their roles in aging-associated methylome deterioration. *Cell Res. 2020 3011* **30**, 980–996 (2020).
32. Haggerty, C. *et al.* Dnmt1 has de novo activity targeted to transposable elements. *Nat. Struct. Mol. Biol. 2021 287* **28**, 594–603 (2021).
33. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases [1]. *Nature Genetics* vol. 19 219–220 (1998).
34. Watanabe, D., Suetake, I., Tada, T. & Tajima, S. Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mech. Dev.* **118**, 187–190 (2002).
35. Lei, H. *et al.* De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195–3205 (1996).
36. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
37. Yan, X. J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat. Genet. 2011 434* **43**, 309–315 (2011).
38. Duymich, C. E., Charlet, J., Yang, X., Jones, P. A. & Liang, G. DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. *Nat. Commun.* **7**, 11453 (2016).
39. Molaro, A., Malik, H. S. & Bourc'his, D. Dynamic Evolution of De Novo DNA Methyltransferases in Rodent and Primate Genomes. *Mol. Biol. Evol.* **37**, 1882–1892 (2020).
40. Hata, K., Okano, M., Lei, H. & Li, E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129**, 1983–1993 (2002).
41. Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H. & Tajima, S. DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. *J. Biol. Chem.* **279**, 27816–27823 (2004).
42. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nat. 2004 4317004* **431**, 96–99 (2004).
43. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science (80-.).* **294**, 2536–2539 (2001).
44. Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science (80-.).* **334**, 194 (2011).
45. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (80-.).* **324**, 930–935 (2009).
46. Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nat. 2010 4667310* **466**, 1129–1133 (2010).
47. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (80-.).* **333**, 1300–1303 (2011).
48. Maiti, A. & Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: POTENTIAL IMPLICATIONS FOR ACTIVE DEMETHYLATION OF CpG SITES. *J. Biol. Chem.* **286**, 35334–35338 (2011).

49. He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science (80-.).* **333**, 1303–1307 (2011).
50. Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science (80-.).* **184**, 868–871 (1974).
51. Fukagawa, T. & Earnshaw, W. C. The Centromere: Chromatin Foundation for the Kinetochore Machinery. *Dev. Cell* **30**, 496 (2014).
52. Smith, B. C. & Denu, J. M. Chemical mechanisms of histone lysine and arginine modifications. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1789**, 45–57 (2009).
53. Heitz, E. *Das heterochromatin der moose*. (Bornträger, 1928).
54. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nat. 2004 4317010* **431**, 873–878 (2004).
55. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in polycomb-group silencing. *Science (80-.).* **298**, 1039–1043 (2002).
56. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nat. 2006 4417091* **441**, 349–353 (2006).
57. Lee, T. I. *et al.* Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells. *Cell* **125**, 301–313 (2006).
58. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science (80-.).* **300**, 131–135 (2003).
59. de Napoles, M. *et al.* Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Dev. Cell* **7**, 663–676 (2004).
60. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nat. 2001 4106824* **410**, 120–124 (2001).
61. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nat. 2001 4106824* **410**, 116–120 (2001).
62. Peters, A. H. F. M. *et al.* Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. *Mol. Cell* **12**, 1577–1589 (2003).
63. Peters, A. H. F. M. *et al.* Loss of the Suv39h Histone Methyltransferases Impairs Mammalian Heterochromatin and Genome Stability. *Cell* **107**, 323–337 (2001).
64. Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nat. 2010 4647290* **464**, 927–931 (2010).
65. Karimi, M. M. *et al.* DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell* **8**, 676–687 (2011).
66. Keniry, A. *et al.* Setdb1-mediated H3K9 methylation is enriched on the inactive X and plays a role in its epigenetic silencing. *Epigenetics Chromatin 2016 91* **9**, 1–20 (2016).
67. Minkovsky, A. *et al.* The Mbd1-Atf7ip-Setdb1 pathway contributes to the maintenance of X chromosome inactivation. *Epigenetics Chromatin 2014 71* **7**, 1–15 (2014).
68. Feldman, N. *et al.* G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat. Cell Biol. 2006 82* **8**, 188–194 (2006).
69. Tachibana, M. *et al.* Histone methyltransferases G9a and GLP form heteromeric complexes and are both crucial for methylation of euchromatin at H3-K9. *Genes Dev.* **19**, 815–826 (2005).
70. Tachibana, M. *et al.* G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16**, 1779–1791 (2002).
71. Montavon, T. *et al.* Complete loss of H3K9 methylation dissolves mouse heterochromatin organization. *Nat. Commun. 2021 121* **12**, 1–16 (2021).
72. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).
73. Bernstein, B. E. *et al.* Genomic Maps and Comparative Analysis of Histone Modifications in Human and Mouse. *Cell* **120**, 169–181 (2005).

74. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nat. 2008 4547205* **454**, 766–770 (2008).
75. Zhang, Y. *et al.* Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.* **38**, 4246–4253 (2010).
76. Otani, J. *et al.* Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. *EMBO Rep.* **10**, 1235–1241 (2009).
77. Baubec, T. *et al.* Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nat. 2015 5207546* **520**, 243–247 (2015).
78. Dhayalan, A. *et al.* The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation. *J. Biol. Chem.* **285**, 26114–26120 (2010).
79. Statham, A. L. *et al.* Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res.* **22**, 1120–1127 (2012).
80. 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).
81. Bernstein, B. E. *et al.* A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315–326 (2006).
82. Pan, G. *et al.* Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells. *Cell Stem Cell* **1**, 299–312 (2007).
83. Bartke, T. *et al.* Nucleosome-Interacting Proteins Regulated by DNA and Histone Methylation. *Cell* **143**, 470–484 (2010).
84. Reddington, J. P. *et al.* Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. *Genome Biol. 2013 143* **14**, 1–17 (2013).
85. Fu, K., Bonora, G. & Pellegrini, M. Interactions between core histone marks and DNA methyltransferases predict DNA methylation patterns observed in human cells and tissues. *https://doi.org/10.1080/15592294.2019.1666649* **15**, 272–282 (2019).
86. Ren, W. *et al.* Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. *Proc. Natl. Acad. Sci.* **117**, 18439–18447 (2020).
87. Liu, X. *et al.* UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat. Commun. 2013 41* **4**, 1–13 (2013).
88. Rothbart, S. B. *et al.* Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat. Struct. Mol. Biol. 2012 1911* **19**, 1155–1160 (2012).
89. Lehnertz, B. *et al.* Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. *Curr. Biol.* **13**, 1192–1200 (2003).
90. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes to Cells* **11**, 805–814 (2006).
91. Saksouk, N. *et al.* Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation. *Mol. Cell* **56**, 580–594 (2014).
92. Xin, H., Yoon, H. G., Singh, P. B., Wong, J. & Qin, J. Components of a Pathway Maintaining Histone Modification and Heterochromatin Protein 1 Binding at the Pericentric Heterochromatin in Mammalian Cells. *J. Biol. Chem.* **279**, 9539–9546 (2004).
93. Hoyt, S. J. *et al.* From telomere to telomere: the transcriptional and epigenetic state of human repeat elements. *bioRxiv* 2021.07.12.451456 (2021) doi:10.1101/2021.07.12.451456.
94. Bakhtiari, M. *et al.* Variable number tandem repeats mediate the expression of proximal genes. *Nat. Commun. 2021 121* **12**, 1–12 (2021).
95. Garg, P. *et al.* Pervasive cis effects of variation in copy number of large tandem repeats on local DNA methylation and gene expression. *Am. J. Hum. Genet.* **108**, 809–824 (2021).
96. Khristich, A. N. & Mirkin, S. M. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.* **295**, 4134–4170 (2020).
97. Depienne, C. & Mandel, J. L. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? *Am. J. Hum. Genet.* **108**, 764–785 (2021).

98. Trizzino, M. *et al.* Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* **27**, 1623–1633 (2017).
99. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nat. 2001 4096822* **409**, 860–921 (2001).
100. Venter, J. C. *et al.* The Sequence of the Human Genome. *Science (80-.).* **291**, 1304–1351 (2001).
101. Altemose, N. *et al.* Complete genomic and epigenetic maps of human centromeres. *bioRxiv* **16**, 2021.07.12.452052 (2021).
102. Gershman, A. *et al.* Epigenetic Patterns in a Complete Human Genome. *bioRxiv* 2021.05.26.443420 (2021) doi:10.1101/2021.05.26.443420.
103. (IWGSC), T. I. W. G. S. C. *et al.* Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science (80-.).* **361**, (2018).
104. Platt, R. N., II, Vandewege, M. W. & Ray, D. A. Mammalian transposable elements and their impacts on genome evolution. *Chromosom. Res.* **26**, 25 (2018).
105. Torres-Padilla, M.-E. On transposons and totipotency. *Philos. Trans. R. Soc. B* **375**, (2020).
106. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012).
107. De Iaco, A. *et al.* DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
108. Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
109. McKinley, K. L. & Cheeseman, I. M. The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol. 2015 171* **17**, 16–29 (2015).
110. Darrow, E. M. *et al.* Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl. Acad. Sci.* **113**, E4504–E4512 (2016).
111. Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol. 2015 161* **16**, 1–21 (2015).
112. Garrick, D., Fiering, S., Martin, D. I. K. & Whitelaw, E. Repeat-induced gene silencing in mammals. *Nat. Genet. 1998 181* **18**, 56–59 (1998).
113. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number variation. *Nat. Genet. 2007 3910* **39**, 1256–1260 (2007).
114. Rennison, D. J., Owens, G. L. & Taylor, J. S. Opsin gene duplication and divergence in ray-finned fish. *Mol. Phylogenet. Evol.* **62**, 986–1008 (2012).
115. Fondon, J. W. & Garner, H. R. Molecular origins of rapid and continuous morphological evolution. *Proc. Natl. Acad. Sci.* **101**, 18058–18063 (2004).
116. Deutekom, J. C. T. V. *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).
117. Hobbs, S. L. A., Kpodar, P. & DeLong, C. M. O. The effect of T-DNA copy number, position and methylation on reporter gene expression in tobacco transformants. *Plant Mol. Biol. 1990 156* **15**, 851–864 (1990).
118. Assaad, F. F., Tucker, K. L. & Signer, E. R. Epigenetic repeat-induced gene silencing (RIGS) inArabidopsis. *Plant Mol. Biol. 1993 226* **22**, 1067–1085 (1993).
119. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Genome Res.* **14**, 9–25 (1975).
120. Holliday, R. & Pugh, J. DNA modification mechanisms and gene activity during development. *Science (80-.).* **187**, 226–232 (1975).
121. Henikoff, S. Conspiracy of silence among repeated transgenes. *BioEs-says* **20**, 532–535 (1998).
122. Weiler, K. S. & Wakimoto, B. T. HETEROCHROMATIN AND GENE EXPRESSION IN DROSOPHILA. *https://doi.org/10.1146/annurev.ge.29.120195.003045* **29**, 577–605 (2003).
123. Muller, H. J. Types of visible variations induced by X-rays in Drosophila. *J. Genet.* **22**, 299–334 (1930).
124. Girton, J. R. & Johansen, K. M. Chapter 1 Chromatin Structure and the Regulation of Gene Expression: The Lessons of PEV in Drosophila. *Adv. Genet.* **61**, 1–43 (2008).

125. Blewitt, M. & Whitelaw, E. The Use of Mouse Models to Study Epigenetics. *Cold Spring Harb. Perspect. Biol.* **5**, a017939 (2013).
126. Blewitt, M. E. *et al.* An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc. Natl. Acad. Sci.* **102**, 7629–7634 (2005).
127. Daxinger, L. *et al.* An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome Biol. 2013 149* **14**, 1–17 (2013).
128. Preis, J. I., Downes, M., Oates, N. A., Rasko, J. E. J. & Whitelaw, E. Sensitive Flow Cytometric Analysis Reveals a Novel Type of Parent-of-Origin Effect in the Mouse Genome. *Curr. Biol.* **13**, 955–959 (2003).
129. Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I. & Whitelaw, E. Metastable epialleles in mammals. *Trends Genet.* **18**, 348–351 (2002).
130. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet. 2007 84* **8**, 272–285 (2007).
131. Duhl, D. M. J., Vrieling, H., Miller, K. A., Wolff, G. L. & Barsh, G. S. Neomorphic agouti mutations in obese yellow mice. *Nat. Genet. 1994 81* **8**, 59–65 (1994).
132. Miller, M. W. *et al.* Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation. *Genes Dev.* **7**, 454–467 (1993).
133. Isbel, L. *et al.* Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse. *Elife* **5**, (2016).
134. Michaud, E. J. *et al.* Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. *Genes Dev.* **8**, 1463–1472 (1994).
135. Youngson, N. A. *et al.* No evidence for cumulative effects in a Dnmt3b hypomorph across multiple generations. *Mamm. Genome 2013 245* **24**, 206–217 (2013).
136. Ashe, A. *et al.* A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. *Genome Biol. 2008 912* **9**, 1–16 (2008).
137. Sorolla, M. A., Marqués, M., Parisi, E. & Sorolla, A. An N-ethyl-N-Nitrosourea Mutagenesis Screen in Mice Reveals a Mutation in Nuclear Respiratory Factor 1 (Nrf1) Altering the DNA Methylation State and Correct Embryonic Development. *Anim. 2021, Vol. 11, Page 2103* **11**, 2103 (2021).
138. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* **99**, 247–257 (1999).
139. Yehezkel, S., Segev, Y., Viegas-Péquignot, 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).
140. 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).
141. Rice, J. C. *et al.* Histone Methyltransferases Direct Different Degrees of Methylation to Define Distinct Chromatin Domains. *Mol. Cell* **12**, 1591–1598 (2003).
142. García-Cao, M., O'Sullivan, R., Peters, A. H. F. M., Jenuwein, T. & Blasco, M. A. Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nat. Genet. 2004 361* **36**, 94–99 (2003).
143. Groh, S. *et al.* Morc3 silences endogenous retroviruses by enabling Daxx-mediated histone H3.3 incorporation. *Nat. Commun. 2021 121* **12**, 1–18 (2021).
144. Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nat. 2010 4647290* **464**, 927–931 (2010).
145. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nat. 2010 4637278* **463**, 237–240 (2010).
146. Lemmers, R. J. L. F. *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).
147. van den Boogaard, M. L. *et al.* Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. *Am. J. Hum. Genet.* **98**, 1020–1029 (2016).

148. Xu, G.-L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191 (1999).

149. Gordon, C. T. *et al.* De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nat. Genet.* **49**, 249–255 (2017).

150. 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).

151. Wijmenga, C. *et al.* Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet. 1992 21* **2**, 26–30 (1992).

152. Weiffenbach, B. *et al.* Mapping the facioscapulohumeral muscular dystrophy gene is complicated by chromosome 4q35 recombination events. *Nat. Genet. 1993 42* **4**, 165–169 (1993).

153. Zampatti, S. *et al.* Facioscapulohumeral muscular dystrophy (FSHD) molecular diagnosis: from traditional technology to the NGS era. *neurogenetics 2019 202* **20**, 57–64 (2019).

154. Hewitt, J. E. *et al.* Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystropothhy. *Hum. Mol. Genet.* **3**, 1287–1295 (1994).

155. 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).

156. Bakker, E. *et al.* The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve. Suppl.* S39-44 (1995).

157. 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).

158. van Overveld, P. G. M. *et al.* Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* **35**, 315–317 (2003).

159. Zeng, W. *et al.* Specific Loss of Histone H3 Lysine 9 Trimethylation and HP1γ/Cohesin Binding at D4Z4 Repeats Is Associated with Facioscapulohumeral Dystrophy (FSHD). *PLoS Genet.* **5**, e1000559 (2009).

160. Gabellini, D., Green, M. R. & Tupler, R. Inappropriate Gene Activation in FSHD: A Repressor Complex Binds a Chromosomal Repeat Deleted in Dystrophic Muscle. *Cell* **110**, 339–348 (2002).

161. 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).

162. 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).

163. Geel, M. van *et al.* Identification of a novel β-tubulin subfamily with one member (TUBB4Q) located near the telomere of chromosome region 4q35. *Cytogenet. Genome Res.* **88**, 316–321 (2000).

164. 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).

165. Dixit, M. *et al.* DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci.* **104**, 18157–18162 (2007).

166. Masny, P. S. *et al.* Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei. *Eur. J. Hum. Genet. 2010 184* **18**, 448–456 (2009).

167. Klooster, R. *et al.* Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. *Eur. J. Hum. Genet. 2009 1712* **17**, 1615–1624 (2009).

168. Tupler, R. *et al.* Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. *J. Med. Genet.* **33**, 366–370 (1996).

169. Hewitt, J. E. *et al. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy*. *Human Molecular Genetics* vol. 3 https://academic.oup.com/hmg/article-abstract/3/8/1287/554670 (1994).

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

171. Gabriëls, J. *et al.* Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* **236**, 25–32 (1999).
172. Lemmers, R. J. L. F. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329**, 1650–3 (2010).
173. Deak, K. L. *et al.* Genotype-phenotype study in an FSHD family with a proximal deletion encompassing p13E-11 and D4Z4. *Neurology* **68**, 578–582 (2007).
174. Lemmers, R. J. L. F. *et al.* High-resolution breakpoint junction mapping of proximally extended D4Z4 deletions in FSHD1 reveals evidence for a founder effect. *Hum. Mol. Genet.* **00**, (2021).
175. Lemmers, R. J. L. F. *et al.* Chromosome 10q-linked FSHD identifies DUX4 as principal disease gene. *J. Med. Genet.* **0**, 1–9 (2021).
176. Hamanaka, K. *et al.* Homozygous nonsense variant in LRIF1 associated with facioscapulohumeral muscular dystrophy. *Neurology* **94**, e2441–e2447 (2020).
177. de Greef, J. C. *et al.* Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mutat.* **30**, 1449–1459 (2009).
178. Theadom, A. *et al.* Prevalence of Muscular Dystrophies: A Systematic Literature Review. *Neuroepidemiology* **43**, 259–268 (2014).
179. Greef, J. C. de *et al.* Clinical features of facioscapulohumeral muscular dystrophy 2(CME). *Neurology* **75**, 1548 (2010).
180. Sacconi, S. *et al.* FSHD1 and FSHD2 form a disease continuum. *Neurology* **92**, E2273–E2285 (2019).
181. Wohlgemuth, M. *et al.* A family-based study into penetrance in facioscapulohumeral muscular dystrophy type 1. *Neurology* **91**, e444–e454 (2018).
182. Tonini, M. M. O. *et al.* Asymptomatic carriers and gender differences in facioscapulohumeral muscular dystrophy (FSHD). *Neuromuscul. Disord.* **14**, 33–38 (2004).
183. 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).
184. Zatz, M. *et al.* The Facioscapulohumeral Muscular Dystrophy (FSHD1) Gene Affects Males More Severely and More Frequently Than Females. *J. Med. Genet* **77**, 155–161 (1998).
185. Deenen, J. C. W. *et al.* Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* **83**, 1056–9 (2014).
186. Padberg, G. W. A. M. *Facioscapulohumeral disease*. (1982).
187. Statland, J. M. & Tawil, R. Risk of functional impairment in Facioscapulohumeral muscular dystrophy. *Muscle Nerve* **49**, 520–527 (2014).
188. Klinge, L. *et al.* Severe phenotype in infantile facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* **16**, 553–558 (2006).
189. Goselink, R. J. M. *et al.* Early onset facioscapulohumeral dystrophy – a systematic review using individual patient data. *Neuromuscul. Disord.* **27**, 1077–1083 (2017).
190. Padberg, G. W. *et al.* On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. *Muscle Nerve* **18**, S73–S80 (1995).
191. Goselink, R. J. M. *et al.* Ophthalmological findings in facioscapulohumeral dystrophy. *Brain Commun.* **1**, (2019).
192. Lutz, K. L., Holte, L., Kliethermes, S. A., Stephan, C. & Mathews, K. D. Clinical and genetic features of hearing loss in facioscapulohumeral muscular dystrophy. *Neurology* **81**, 1374–1377 (2013).
193. Trevisan, C. P. *et al.* Facioscapulohumeral muscular dystrophy: a multicenter study on hearing function. *Audiol. Neurootol.* **13**, 1–6 (2008).
194. Brouwer, O. F., Padberg, G. W., Wijmenga, C. & Frants, R. R. Facioscapulohumeral Muscular Dystrophy in Early Childhood. *Arch. Neurol.* **51**, 387–394 (1994).
195. Funakoshi, M., Goto, K. & Arahata, K. Epilepsy and mental retardation in a subset of early onset 4q35-facioscapulohumeral muscular dystrophy. *Neurology* **50**, 1791–1794 (1998).

196. Goselink, R. J. M. *et al.* Early onset as a marker for disease severity in facioscapulohumeral muscular dystrophy. *Neurology* **92**, E378–E385 (2019).
197. Lemmers, R. J. F. L. *et al.* Contractions of D4Z4 on 4qB Subtelomeres Do Not Cause Facioscapulohumeral Muscular Dystrophy. *Am. J. Hum. Genet.* **75**, 1124–1130 (2004).
198. Lemmers, R. J. L. F. *et al.* Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **81**, 884–94 (2007).
199. van Geel, M. *et al.* Genomic Analysis of Human Chromosome 10q and 4q Telomeres Suggests a Common Origin. *Genomics* **79**, 210–217 (2002).
200. Lemmers, R. J. L. F. *et al.* Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat. Genet.* **32**, 235–236 (2002).
201. Lemmers, R. J. L. F. *et al.* Worldwide Population Analysis of the 4q and 10q Subtelomeres Identifies Only Four Discrete Interchromosomal Sequence Transfers in Human Evolution. *Am. J. Hum. Genet.* **86**, 364–377 (2010).
202. Jones, T. I., Himeda, C. L., Perez, D. P. & Jones, P. L. Large family cohorts of lymphoblastoid cells provide a new cellular model for investigating facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* **27**, 221–238 (2017).
203. Scionti, I. *et al.* Large-Scale Population Analysis Challenges the Current Criteria for the Molecular Diagnosis of Fascioscapulohumeral Muscular Dystrophy. *Am. J. Hum. Genet.* **90**, 628–635 (2012).
204. Lemmers, R. J. *et al.* Deep characterization of a common D4Z4 variant identifies biallelic DUX4 expression as a modifier for disease penetrance in FSHD2. *Eur. J. Hum. Genet.* **26**, 94–106 (2018).
205. 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).
206. Ricci, E. *et al.* Progress in the Molecular Diagnosis of Facioscapulohumeral Muscular Dystrophy and Correlation between the Number of KpnI Repeats at the 4q35 Locus and Clinical Phenotype. (1999) doi:10.1002/1531-8249.
207. 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).
208. Mul, K. *et al.* Phenotype-genotype relations in facioscapulohumeral muscular dystrophy type 1. *Clin. Genet.* **94**, 521–527 (2018).
209. Wang, C.-H. *et al.* Correlation between muscle involvement, phenotype and D4Z4 fragment size in facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* **22**, 331–8 (2012).
210. Tawil, R. *et al.* Evidence for anticipation and association of deletion size with severity in facioscapulohumerd muscular dystrophy. *Ann. Neurol.* **39**, 744–748 (1996).
211. Scionti, I. *et al.* Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling. *J. Med. Genet.* **49**, 171–178 (2012).
212. 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 2013 141* **14**, 1–12 (2013).
213. Park, H. J. *et al.* Low D4Z4 copy number and gender difference in Korean patients with facioscapulohumeral muscular dystrophy type 1. *Neuromuscul. Disord.* **25**, 859–64 (2015).
214. Goto, K., Nishino, I. & Hayashi, Y. K. Rapid and accurate diagnosis of facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* **16**, 256–261 (2006).
215. Lemmers, R. J. L. F., van der Wielen, M. J. R., Bakker, E., Frants, R. R. & van der Maarel, S. M. Rapid and accurate diagnosis of facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* **16**, 615–7; author reply 617-8 (2006).
216. Himeda, C. L. *et al.* Myogenic Enhancers Regulate Expression of the Facioscapulohumeral Muscular Dystrophy-Associated DUX4 Gene. *Mol. Cell. Biol.* **34**, 1942–1955 (2014).
217. Cabianca, D. S. *et al.* A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**, 819–31 (2012).

218. Deak, K. L. *et al.* Genotype-phenotype study in an FSHD family with a proximal deletion encompassing p13E-11 and D4Z4. *Neurology* **68**, 578–582 (2007).

219. Lemmers, R. J. L. F. *et al.* D4F104S1 deletion in facioscapulohumeral muscular dystrophy. *Neurology* **61**, 178–183 (2003).

220. Lemmers, R. J. L. F. *et al.* Inter-and Intrachromosomal Sub-Telomeric Rearrangements on 4q35: Implications for Facioscapulohumeral Muscular Dystrophy (FSHD) Aetiology and Diagnosis. *Hum. Mol. Genet.* **7**, 1207–1214 (1998).

221. De Greef, J. C. *et al.* Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. *Neurology* **69**, 1018–1026 (2007).

222. Lemmers, R. J. L. F. *et al.* SMCHD1 mutation spectrum for facioscapulohumeral muscular dystrophy type 2 (FSHD2) and Bosma arhinia microphthalmia syndrome (BAMS) reveals disease-specific localisation of variants in the ATPase domain. *J. Med. Genet.* **56**, 693–700 (2019).

223. Lemmers, R. J. L. F. *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).

224. Brideau, N. J. *et al.* Independent Mechanisms Target SMCHD1 to Trimethylated Histone H3 Lysine 9-Modified Chromatin and the Inactive X Chromosome. *Mol. Cell. Biol.* **35**, 4053–68 (2015).

225. Gurzau, A. D. *et al.* SMCHD1's ubiquitin-like domain is required for N-terminal dimerization and chromatin localization. *Biochem. J.* **478**, 2555–2569 (2021).

226. 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).

227. Lemmers, R. J. L. F. *et al.* Cis D4Z4 repeat duplications associated with facioscapulohumeral muscular dystrophy type 2. *Hum. Mol. Genet.* **27**, 3488–3497 (2018).

228. 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).

229. Cascella, R. *et al.* Digenic inheritance of shortened repeat units of the D4Z4 region and a loss-of-function variant in SMCHD1 in a Family with FSHD. *Front. Neurol.* **9**, 1027 (2018).

230. Goossens, R. *et al.* Intronic SMCHD1 variants in FSHD: Testing the potential for CRISPR-Cas9 genome editing. *J. Med. Genet.* **56**, 828–837 (2019).

231. Calandra, P. *et al.* Allele-specific DNA hypomethylation characterises FSHD1 and FSHD2. *J. Med. Genet.* **53**, 348–55 (2016).

232. Jones, T. I. *et al.* Identifying diagnostic DNA methylation profiles for facioscapulohumeral muscular dystrophy in blood and saliva using bisulfite sequencing. *Clin. Epigenetics* **6**, 1–16 (2014).

233. Balog, J. *et al.* Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4. *Epigenetics* **10**, 1133–1142 (2015).

234. 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).

235. Blewitt, M. E. *et al.* SmcHD1, containing a structural-maintenance-of- chromosomes hinge domain, has a critical role in X inactivation. (2008) doi:10.1038/ng.142.

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

237. Pedersen, L. C., Inoue, K., Kim, S., Perera, L. & Shaw, N. D. A ubiquitin-like domain is required for stabilizing the N-terminal ATPase module of human SMCHD1. *Commun. Biol.* **2**, 255 (2019).

238. Chen, K. *et al.* Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3535-44 (2015).

239. Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16**, 1–21 (2015).

240. Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nat. 2016 5357613* **535**, 575–579 (2016).

241. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

242. Jansz, N. *et al.* Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters. *Nat. Struct. Mol. Biol.* **25**, 766–777 (2018).
243. Wang, C.-Y., Jégu, T., Chu, H.-P., Oh, H. J. & Lee, J. T. SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X. *Cell* **174**, 406-421.e25 (2018).
244. Gdula, M. R. *et al.* The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome. *Nat. Commun.* **10**, 30 (2019).
245. Gendrel, A.-V. *et al.* Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. *Mol. Cell. Biol.* **33**, 3150–65 (2013).
246. 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 2013 61* **6**, 1–16 (2013).
247. Wanigasuriya, I. *et al.* Smchd1 is a maternal effect gene required for genomic imprinting. *Elife* **9**, 1–27 (2020).
248. 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).
249. 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).
250. 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).
251. Das, S. & Chadwick, B. P. Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. *PLoS One* **11**, e0160022 (2016).
252. Van Overveld, P. G. M. *et al.* Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet. 2003 354* **35**, 315–317 (2003).
253. Haynes, P., Bomsztyk, K. & Miller, D. G. Sporadic DUX4 expression in FSHD myocytes is associated with incomplete repression by the PRC2 complex and gain of H3K9 acetylation on the contracted D4Z4 allele. *Epigenetics Chromatin* **11**, 47 (2018).
254. Balog, J. *et al.* Monosomy 18p is a risk factor for facioscapulohumeral dystrophy. *J. Med. Genet.* **55**, 469–478 (2018).
255. Campbell, A. E. *et al.* NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *Elife* **7**, (2018).
256. Signorelli, M. *et al.* Evaluation of blood gene expression levels in facioscapulohumeral muscular dystrophy patients. *Sci. Reports 2020 101* **10**, 1–11 (2020).
257. Banerji, C. R. S., Panamarova, M. & Zammit, P. S. DUX4 expressing immortalized FSHD lymphoblastoid cells express genes elevated in FSHD muscle biopsies, correlating with the early stages of inflammation. *Hum. Mol. Genet.* **29**, 2285–2299 (2020).
258. Ferreboeuf, M. *et al.* DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. *Hum. Mol. Genet.* **23**, 171–181 (2014).
259. Williams, K. *et al.* Muscle group specific transcriptomic and DNA methylation differences related to developmental patterning in FSHD. *bioRxiv* 2021.09.28.462147 (2021) doi:10.1101/2021.09.28.462147.
260. Leidenroth, A. *et al.* Evolution of DUX gene macrosatellites in placental mammals. *Chromosoma* **121**, 489–497 (2012).
261. 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).
262. Eidahl, J. O. *et al.* Mouse Dux is myotoxic and shares partial functional homology with its human paralog DUX4. *Hum. Mol. Genet.* **25**, 4577–4589 (2016).
263. Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **49**, 935–940 (2017).
264. Ishiuchi, T. *et al.* Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–671 (2015).
265. Bošković, A. *et al.* Higher chromatin mobility supports totipotency and precedes pluripotency in vivo. *Genes Dev.* **28**, 1042–1047 (2014).

266. Huang, Z. *et al.* The chromosomal protein SMCHD1 regulates DNA methylation and the 2c-like state of embryonic stem cells by antagonizing TET proteins. *Sci. Adv.* **7**, eabb9149 (2021).

267. Grow, E. J. *et al.* p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. *Nat. Genet. 2021 538* **53**, 1207–1220 (2021).

268. Wu, K. *et al.* SETDB1-Mediated Cell Fate Transition between 2C-Like and Pluripotent States. *Cell Rep.* **30**, 25-36.e6 (2020).

269. Akiyama, T. *et al.* Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. *DNA Res.* **22**, 307–318 (2015).

270. Maksakova, I. A. *et al.* Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics and Chromatin* **6**, 15 (2013).

271. Genet, M. & Torres-Padilla, M.-E. The molecular and cellular features of 2-cell-like cells: a reference guide. *Development* **147**, (2020).

272. Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**, 391-405.e19 (2018).

273. Cossec, J. C. *et al.* SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* **23**, 742-757.e8 (2018).

274. Eckersley-Maslin, M. *et al.* Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. *Genes Dev.* **33**, 194–208 (2019).

275. Fu, X., Wu, X., Djekidel, M. N. & Zhang, Y. Myc and Dnmt1 impede the pluripotent to totipotent state transition in embryonic stem cells. *Nat. Cell Biol.* **21**, 835–844 (2019).

276. Chen, Z. & Zhang, Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat. Genet. 2019 516* **51**, 947–951 (2019).

277. De Iaco, A., Verp, S., Offner, S., Grun, D. & Trono, D. DUX is a non-essential synchronizer of zygotic genome activation. *Dev.* **147**, (2020).

278. Bosnakovski, D., Gearhart, M. D., Ho Choi, S. & Kyba, M. Dux facilitates post-implantation development, but is not essential for zygotic genome activation. *Biol. Reprod.* **104**, 83–93 (2021).

279. Guo, M. *et al.* Precise temporal regulation of Dux is important for embryo development. *Cell Res. 2019 2911* **29**, 956–959 (2019).

280. Yu, H. *et al.* rRNA biogenesis regulates mouse 2C-like state by 3D structure reorganization of peri-nucleolar heterochromatin. *Nat. Commun. 2021 121* **12**, 1–21 (2021).

281. Xie, S. Q. *et al.* Nucleolar-based Dux repression is essential for 2-cell stage exit. *bioRxiv* 2021.11.11.468235 (2021) doi:10.1101/2021.11.11.468235.

282. Lafontaine, D. L. J., Riback, J. A., Bascetin, R. & Brangwynne, C. P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol. 2020 223* **22**, 165–182 (2020).

283. McStay, B. Nucleolar organizer regions: genomic 'dark matter' requiring illumination. *Genes Dev.* **30**, 1598–1610 (2016).

284. 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).

285. Tam, R., Smith, K. P. & Lawrence, J. B. The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres. *J. Cell Biol.* **167**, 269–279 (2004).

286. Winokur, S. T., Bengtsson, U., Vargas, J. C., Wasmuth, J. J. & Altherr, M. R. The Evolutionary Distribution and Structural Organization of the Homeobox-Containing Repeat D4Z4 Indicates a Functional Role for the Ancestral Copy in the FSHD Region. *Hum. Mol. Genet.* **5**, 1567–1575 (1996).

287. Yao, Z. *et al.* DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum. Mol. Genet.* **23**, 5342–5352 (2014).

288. Jagannathan, S. *et al.* Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum. Mol. Genet.* **25**, 4419–4431 (2016).

289. Geng, L. N. *et al.* DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* **22**, 38–51 (2012).

290. Young, J. M. *et al.* DUX4 Binding to Retroelements Creates Promoters That Are Active in FSHD Muscle and Testis. *PLoS Genet.* **9**, e1003947 (2013).
291. Block, G. J. *et al.* Asymmetric Bidirectional Transcription from the FSHD-Causing D4Z4 Array Modulates DUX4 Production. *PLoS One* **7**, e35532 (2012).
292. 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**, (2015).
293. Van Den Heuvel, A. *et al.* Single-cell RNA sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. *Hum. Mol. Genet.* **28**, 1064–1075 (2019).
294. Snider, L. *et al.* Facioscapulohumeral Dystrophy: Incomplete Suppression of a Retrotransposed Gene. *PLoS Genet.* **6**, e1001181 (2010).
295. Jiang, S. *et al.* Single-nucleus RNA-seq identifies divergent populations of FSHD2 myotube nuclei. *PLOS Genet.* **16**, e1008754 (2020).
296. Wang, L. H. *et al.* MRI-informed muscle biopsies correlate MRI with pathology and DUX4 target gene expression in FSHD. *Hum. Mol. Genet.* **28**, 476–486 (2019).
297. Chau, J. *et al.* Relationship of DUX4 and target gene expression in FSHD myocytes. *Hum. Mutat.* **42**, 421–433 (2021).
298. Choi, S. H. *et al.* DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **44**, 5161–73 (2016).
299. Bosnakovski, D. *et al.* A novel P300 inhibitor reverses DUX4-mediated global histone H3 hyperacetylation, target gene expression, and cell death. *Sci. Adv.* **5**, 7781–7792 (2019).
300. Resnick, R. *et al.* DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. *Cell Rep.* **29**, 1812-1820.e5 (2019).
301. Kowaljow, V. *et al.* The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul. Disord.* **17**, 611–23 (2007).
302. 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).
303. Lek, A. *et al.* Applying genome-wide CRISPR-Cas9 screens for therapeutic discovery in facioscapulohumeral muscular dystrophy. *Sci. Transl. Med.* **12**, 271 (2020).
304. Bosnakovski, D. *et al.* High-throughput screening identifies inhibitors of DUX4-induced myoblast toxicity. *Skelet. Muscle* **4**, 1–11 (2014).
305. Dmitriev, P. *et al.* DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. *Free Radic. Biol. Med.* **99**, 244–258 (2016).
306. Shadle, S. C. *et al.* DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. *PLOS Genet.* **13**, e1006658 (2017).
307. Shadle, S. C. *et al.* DUX4-induced bidirectional HSATII satellite repeat transcripts form intranuclear double-stranded RNA foci in human cell models of FSHD. *Hum. Mol. Genet.* **28**, 3997–4011 (2019).
308. Bosnakovski, D. *et al.* Low level DUX4 expression disrupts myogenesis through deregulation of myogenic gene expression. *Sci. Rep.* **8**, 1–12 (2018).
309. Bosnakovski, D. *et al.* The DUX4 homeodomains mediate inhibition of myogenesis and are functionally exchangeable with the Pax7 homeodomain. *J. Cell Sci.* **130**, 3685–3697 (2017).
310. Von Maltzahn, J., Jones, A. E., Parks, R. J. & Rudnicki, M. A. Pax7 is critical for the normal function of satellite cells in adult skeletal muscle. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16474–16479 (2013).
311. Seale, P. *et al.* Pax7 Is Required for the Specification of Myogenic Satellite Cells. *Cell* **102**, 777–786 (2000).
312. Bosnakovski, D. *et al.* An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J.* **27**, 2766–2779 (2008).
313. Haynes, P., Kernan, K., Zhou, S. L. & Miller, D. G. Expression patterns of FSHD-causing DUX4 and myogenic transcription factors PAX3 and PAX7 are spatially distinct in differentiating human stem cell cultures. *Skelet. Muscle* **7**, 1–13 (2017).

314. Banerji, C. R. S. *et al.* PAX7 target genes are globally repressed in facioscapulohumeral muscular dystrophy skeletal muscle. *Nat. Commun. 2017 81* **8**, 1–13 (2017).

315. Banerji, C. R. S. & Zammit, P. S. PAX7 target gene repression is a superior FSHD biomarker than DUX4 target gene activation, associating with pathological severity and identifying FSHD at the single-cell level. *Hum. Mol. Genet.* **28**, 2224–2236 (2019).

316. Banerji, C. R. S. PAX7 target gene repression associates with FSHD progression and pathology over 1 year. *Hum. Mol. Genet.* **29**, 2124–2133 (2020).

317. Van Den Heuvel, A. *et al.* Facioscapulohumeral dystrophy transcriptome signatures correlate with different stages of disease and are marked by different MRI biomarkers. *Sci. Reports 2022 121* **12**, 1–18 (2022).

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

319. Goll, M. G. *et al.* Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science (80-.).* **311**, 395–398 (2006).

320. Vogan, A. A. *et al.* The Enterprise, a massive transposon carrying Spok meiotic drive genes. *Genome Res.* **31**, 789–798 (2021).

321. Pech, M., Igo-Kemenes, T. & Zachau, H. G. Nucleotide sequence of a highly repetitive component of rat DNA. *Nucleic Acids Res.* **7**, 417–432 (1979).

322. Kazakov, V. Why did the heated discussion arise between Erb and Landouzy–Dejerine concerning the priority in describing the facio-scapulo-humeral muscular dystrophy and what is the main reason for this famous discussion? *Neuromuscul. Disord.* **11**, 421 (2001).