UNIVERSITY OF CALIFORNIA,
IRVINE

Understanding the role of DUX4 in FSHD pathology and its connection to early human development

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Sciences

by

Nam Viết Nguyễn

Dissertation Committee:
Professor Kyoko Yokomori, Chair
Professor Ali Mortazavi
Professor Xing Dai
Assistant Professor Rémi Buisson

2022

DEDICATION

To

the Nguyen and the Truong family

my partner, Kevin Cabrera and my friends

who have given me an enormous amount of growth

and support for me to pursue my childhood dream

TABLE OF CONTENTS

LIST OF FIGURES

Appendix I

LIST OF TABLES

ACKNOWLEDGEMENTS

During my time at the Yokomori lab, I was trained and mentored by many inspiring individuals. First of all, I would like to thank my advisor Dr. Kyoko Yokomori for giving me the opportunity to grow and mature as a scientist. She has been always proactive in providing me with resources and support to work on multidisciplinary research projects when I asked for help. She also encouraged critical thinking, questioning, and careful interpretation of data. For the past five years in her lab, I have learned on many important aspects of being a researcher and a scholar. I also would like to thank my co-advisor Dr. Ali Mortazavi for giving me guidance and advise on sequencing technologies and bioinformatic analysis. His assistance allowed my projects to move forward quickly. The rest of my thesis committee members, Dr. Xing Dai and Dr. Rémi Buisson also gave me support and suggestions on my projects.

I am also thankful for everyone who was involved with my projects. The past and current members of the Yokomori lab: Jonathan Chau, Yumeng Li, Dr. Xiangduo Kong, Altay Bayrakci, Preston Carroll and Marrisa Fukunaga who assisted me on various aspects of my research and made my time at the Yokomori lab memorable. Particularly, I truly appreciate Dr. Xiangduo Kong's training and his hard work in the Yokomori lab. I have learned a lot from working with him at the bench. I also appreciate the help and friendship from the members of the Mortazavi lab: Kate Williams, Shan Jiang (Mandy), Weihua Zeng (Benny), Cassandra McGill (Cassie), Heidi Yang, Jasmine Sakr and the other members of the lab.

Finally, I would like to thank my classmates including my partner Kevin Cabrera for all the good memories that we shared. My time as a graduate student is definitely an adventure of a lifetime.

Vita

NAM VIẾT NGUYỄN

EDUCATION

2013 **B.S.** in Biochemistry, University of California, Santa Barbara

2016 **M.S.** in Pharmaceutical Sciences, Western University of Pomona

2022 **Ph.D.** in Biomedical Science, University of California, Irvine

FIELD OF STUDY

Computational Biology in Human Genetics and Diseases

PUBLICATIONS

***N. V. Nguyen**, *X. Kong, Y. Li, J. S. Sakr, K. Williams, S. Sharifi, J. Chau, A. Bayrakci, S. Mizuno, S. Takahashi, T. Kiyono, R. Tawil, A. Mortazavi, K. Yokomori, Engineered SMCHD1 and D4Z4 mutations reveal roles of D4Z4 heterochromatin disruption and feedforward DUX4 network activation in FSHD, *bioRxiv*, https://doi.org/10.1101/2022.10.14.512332.

Williams, K., Kong, X., **Nguyen, N.**, McGill, C., Tawil, R., Yokomori, K., Mortazavi, A. Muscle group specific transcriptomic and DNA methylation differences related to developmental patterning influence FSHD, bioRxiv, https://doi.org/10.1101/2021.09.28.462147.

J. Chau, X. Kong, **N. Viet Nguyen**, K. Williams, M. Ball, R. Tawil, T. Kiyono, A. Mortazavi, K. Yokomori, Relationship of DUX4 and target gene expression in FSHD myocytes. *Hum Mutat*. **42**, 421–433 (2021).

S. Jiang, K. Williams, X. Kong, W. Zeng, **N. V. Nguyen**, X. Ma, R. Tawil, K. Yokomori, A. Mortazavi, Single-nucleus RNA-seq identifies divergent populations of FSHD2 myotube nuclei. *PLoS Genet*. **16**, 1–26 (2020).

Nguyen, N. V., Tran, J. T., & Sanchez, D. J. HIV blocks Type I IFN signaling through disruption of STAT1 phosphorylation. Innate immunity, 24(8), 490–500. https://doi.org/10.1177/1753425918803674 (2018).

SELECTED PRESENTATIONS

Asilomar Chromatin, Chromosomes, and Epigenetics Conference 2019
Presentation: The Role of SMCHD1 in FSHD

FSHD International Research Congress 2021
Oral Presentation: Genetic engineering and characterization of isogenic FSHD mutant myocytes

FSHD International Research Congress 2022
Poster Presentation: Genetic engineering and characterization of isogenic FSHD mutant myocytes

TEACHING EXPERIENCE

Teaching Associate, COSMOS for high school students 2021
University of California, Irvine

CERTIFICATES

Business Concepts for STEM Scientists, 2020
GPS-STEM and Beall Applied Innovation, UC Irvine

EXTRACURRICULAR ACTIVITIES

Bioinformatic Support Group Officer 2018-2019
University of California, Irvine

Biological Chemistry Department, Social Chair 2019-2020
University of California, Irvine

ABSTRACT OF THE DISSERTATION

Understanding the role of DUX4 in FSHD pathology and its connection to early human development

by

Nam Viết Nguyễn

Doctor of Philosophy in Biomedical Science

University of California, Irvine, 2022

Professor Kyoko Yokomori, Chair

Facioscapulohumeral muscular dystrophy (FSHD) is a rare genetic disease and is considered one of the most prominent muscular dystrophies in human. The disease typically begins with weakening of muscles starting from the upper body (face, shoulders, arms) that eventually spread to the lower body (legs). FSHD patients' symptoms generally appear at around adolescent and continue to worsen overtime. Currently there is no cure for FSHD; however, clinical management is available to slow down the progression of muscle loss. With the recent advancement in genetic research technology, major mutations that cause the disease were discovered. Most FSHD patients (95%) have a contraction of the D4Z4 repeat macrosatellite at the subtelomeric region of chromosome 4q with a specific haplotype (4qA). These patients are designated as FSHD1. Around 5% of FSHD cases acquire other mutations that disrupt the heterochromatic establishment of D4Z4 repeats such as DNA methylation (termed FSHD2). Their mutations so far were found in

the epigenetic modifier genes (SMCHD1, LRIF1, DNMT3B) which might be involved in D4Z4 heterochromatin.

DUX4, a transcription factor, is encoded in the D4Z4 repeats. It was found to be re-activated from the last D4Z4 repeat in FSHD patients and has been linked to the development of the disease. During my time in the Yokomori lab, I have characterized FSHD mutant myocytes which were generated by Dr. Xiangduo Kong by CRISPR-Cas9 from a healthy permissive myoblast line. The mutants carry either deletion of D4Z4 repeats or SMCHD1 homozygous mutations or both. I discovered that the mutant cells shared significant characteristics with FSHD patient cells by having a de-repression/reactivation of DUX4. DUX4 was undetected in my RNA-seq analysis but still sufficient to activate FSHD signature genes in FSHD mutants while the parental wildtype cell line had minimal to zero expression of such genes.

DUX4 is a double homeobox protein and is involved in embryonic genome activation (EGA). Homeobox family transcription factors are considered to be central in governing early mammalian development. I further characterized DUX4 major downstream targets such as H3.X/Y, LEUTX, and DUXA and demonstrated that they enhance the DUX4 network through positive feedback loop. I also identified a set of embryonic genes that were induced by LEUTX overexpression in FSHD mutant myocytes. Particularly, some of these genes (DPRX, DPPA3) were previously found to be upregulated at 8-cell stage of embryonic development corresponding to LEUTX timing of expression from 4-cell to 8-cell stage. Altogether, these findings suggested that misexpression of LEUTX in FSHD muscles could disrupt muscle differentiation by activation of embryonic genes.

Chapter 1: INTRODUCTION

The complex epigenetic regulation of *DUX4* and its role in human development and diseases

1.1 Abstract:

DUX4 is a retro-transcription factor that is involved in embryonic genome activation (EGA) in early embryos, whose misexpression in adult skeletal muscle is linked to Facioscapulohumeral Dystrophy (FSHD). FSHD is a third common muscular dystrophy. Studies have shown that *DUX4* misexpression occurs in adult muscle due to the disruption of epigenetic regulations that control the heterochromatin structure at D4Z4 macrosatellite repeat array, in which the DUX4 gene is embedded. Consequently, *DUX4* misexpression leads to activation of downstream target genes and pathways that gradually results in progressive loss of muscles though the detailed mechanism has not been defined. Many DUX4 target genes have been identified yet many of their functions are not well-defined. The function of DUX4 outside of FSHD context also is not completely determined. We discuss here the discovery of *DUX4* as well as its evolution from the *DUX* gene family and how the recent findings have led to a broader understanding of early human development. We also discuss the multifaceted epigenetic regulation of DUX4 and its target genes along with their known functions in maintenance and amplification of DUX4 gene network. This review serves to speculate the interesting case of lowly expressed (often undetectable in bulk RNA-seq) *DUX4* and its incredible long-term impact on adult skeletal muscle. The understanding of DUX4 and its associated genes in human diseases and development might provide a deeper perspective on the complexity of human gene network.

1.2 Introduction

Overview of FSHD and DUX4

FSHD (Facioscapulohumeral dystrophy) is present in around 1 out of 20000 people globally. The cause of FSHD is not clearly understood but it is generally accepted that misexpression of *DUX4* in skeletal muscle myocytes is associated with the disease. Many studies have collectively revealed that the heterochromatin of D4Z4 repeat array is disrupted and the last repeat has a complete ORF for *DUX4* (*1–4*). Heterochromatic disruption and relaxation of D4Z4 repeats allow for de-repression of *DUX4*, resulting in progressive muscle wasting in adult overtime. FSHD patients are divided into two classes: the D4Z4 contraction (FSHD1) in which the repeat array is shortened to less than 10 copies; and the non-contraction (FSHD2) which is mostly linked to the heterozygous mutation of SMCHD1 (structural maintenance of chromosomes hinge domain 1). FSHD2 patients also obtain a low-moderate number of D4Z4 repeats (8-20 copies). The D4Z4 contraction (FSHD1) makes up 95% of the total cases.

DUX4 is located at the sub-telomeric region of the chromosome 4 and only the 4qA but not 4qB allele has the ability to encode for translatable *DUX4* because of the polyA signal downstream of the last D4Z4 repeat. Expression of DUX4 is time-specific in early human development where DUX4 is activated at the 2-cell stage to instruct embryonic genome activation (EGA) (*5*, *6*). In adults, DUX4 is found to be expressed in testis and thymus; and it appears to be subsequently silenced in adult tissues (*7–9*). Heterochromatic factors must be recruited to silent *DUX4* in later development while mutations related to

FSHD appear to inhibit this process. The disruption of heterochromatic D4Z4 repeats in FSHD patients can include H3K9me3, H3K27me3, DNA methylation and other complexes(*2*, *3*, *10*, *11*). The re-activation of *DUX4* arises from ineffective silencing of *DUX4* in FSHD adult muscle. Though *DUX4* is lowly expressed, its influence is substantial, going beyond immediate downstream gene network. Many DUX4 targets have been identified and characterized to understand the disease pathology (ZSCAN4, LEUTX, DUXA, DUXB, H3.X/Y, MBD3L family). The discovery of DUX4 has opened up another area of study as DUX4 was initially considered a pseudogene without any function. However, the current understanding of DUX4 targets is limited and there are still more mysteries of DUX4 yet to unfold.

Physiological role of DUX4 in embryonic development:

Following fertilization, embryo undergoes cleavages and cell divisions without significant overall growth. Due to the lack of gene transcription, the initial cleavage stage of an embryo is assisted by maternal RNAs. As the embryo develops, maternal materials are being degraded and embryonic genome activation (EGA) starts to take place, turning on the first wave of embryonic transcription program. DUX4 is activated at 2-cell stage and peaks at 4-cell stage. At 8-cell stage, DUX4 is being strongly downregulated at RNA and protein level(*6*, *12*, *13*). Cleavage-specific genes such as *ZSCAN4*, *KDM4E*, and *PRAMEF* family were revealed to be activated by DUX4 (Figure 1.1) (*13*). Zscan4 genes are expressed in 2-cell stage in mice and act as transcriptional repressors at heterochromatic regions(*14*). KDM4E (H3K9me demethylase) belongs to KDM4 family and has been characterized so far to be an

epigenetic regulator for embryonic development in bovine(*15*, *16*). This evidence indicates that DUX4 induces chromatin regulators which might be critical for instruction of broader program of gene activation/inhibition during embryonic development. However, PRAMEF family is currently uncharacterized and unknown of biological mechanism. DUX4 also was found to activate endogenous retroelements such as *HERVL* during cleavage. Furthermore, Dux, a homolog of DUX4 in mice was discovered to serve a similar role (*12*). The findings reveal the key conserved functions of the double homeobox genes (*Dux* and *DUX4*) in regulating early embryonic program. During this critical period, embryonic genome becomes activated, allowing the embryo to achieve totipotency. Because of their conserved functions, family of genes regulated by Dux and DUX4 are likely to represent the core ancestral network, possibly setting the momentum for subsequent development processes(*13*).

Expression of DUX4 in other tissues

In human adult, *DUX4* is expressed in testis but suppressed in most somatic tissues(*8*). RT-qPCR analysis of FSHD and unaffected skeletal muscles and unaffected testes revealed that DUX4 targets (PRAMEF1, TRIM43, ZSCAN4, MBD3L2) were activated in FSHD skeletal muscle and testis(*17*). These genes whose expression might be toxic in skeletal muscle are not well-understood in testis. It is also not understood whether DUX4 targets have similar biological functions in both FSHD skeletal muscle and testis. In addition, DUX4-bound repetitive elements such as MaLR and ERV were shown to be active promoters for retrotransposon transcription that could have some crucial role in testis(*8*).

Fascinatingly, DUX4 is also involved in cancer cells and immune evasion(*18*). Aberrant expression of DUX4 suppressed MHC class I and prompted cancer cells to fail immune checkpoint blockade. This finding suggests that DUX4 can also instruct certain immune activity. Intriguingly, DUX4 is also expressed in thymus though whether DUX4 targets are upregulated in this organ is largely unknown(*7*). It was implied that DUX4 might be responsible for some immunological activity in both testis and thymus. Testis is immunologically privileged while thymus is a site for the production and maturation of immune cells. DUX4 might partially share a similar role in both thymus and testis. Further studies needed to address the biological functions of DUX4-activated genes in the somatic tissues.

Functional conservation of DUX4 and DUX family

Since *DUX4* is specific to primates, its functions probably diverge from other mammalian species. In addition, both *Dux* and *DUX4* evolved separately from parental *DUXC* more than 100 million years ago through retrotranposition(*19*). However, *DUXC* was lost in both mouse and primate lineages during evolution (Figure 1.2). Interestingly, *DUXC* was found to be retained in Laurasiatheria (dog, cow, dolphin, and bat) and contains an intron while *DUX4* is intronless(*20*). In canines, DUXC was also revealed to activate cleavage program and retroelements similar to DUX4(*21*). Expression of DUXC was also detected in testis and thymus. *DUX4* split from other lineages might have provided profound evidence for primate evolution. Two other human *DUX* genes, *DUXA* and *DUXB*, are DUX4 targets but contain introns(*22*). DUXA and DUXB were recently found to be

expressed in early embryo development(*23*). It is possible that both DUXA and DUXB share some functions with DUX4 to orchestrate the embryonic cleavage program. DUX4 does not seem to act alone but could partner with its target (DUXA and DUXB) to regulate gene network. However, detailed gene network of DUX family has not been studied extensively. Mechanism of how DUX4 induces a diversity of genes is not well-characterized but past studies altogether have suggested that DUX4 is associated with numerous biological pathways(*8*, *13*, *17*, *24*–*26*). Many more biological roles of DUX4 are still currently under investigation.

1.3 Epigenetic regulation of DUX4

DUX4 transcription is located at D4Z4 repeats

4q D4Z4 repeats are polymorphic and a genomic site for *DUX4* expression. Because *DUX4* is a strong transcription factor that can activate many genes and biological pathways (EGA, retroelements, immune responses, germline, etc.), it is undoubted that transcription of *DUX4* should be tightly controlled. Moreover, *DUX4*-fl transcript (a full-length isoform that causes FSHD) is extremely difficult to detect as DUX4 protein is estimated to be present in < 0.1% of patient muscle cells(*27*). Only one study was able to capture *DUX4*-fl transcript in patient muscle using single-nucleus RNA-seq but at a very low level(*28*). Therefore, *DUX4*-fl transcript abundancy is not required to achieve its biological functions. The organization of D4Z4 repeat array is unusual but very crucial for *DUX4* transcription since contraction/disturbance of heterochromatic D4Z4 repeats is likely to cause FSHD. Additionally, for most FSHD patients, disease-causing mutation is not located on the gene

sequence. This evidence indicates that disruption of *DUX4* regulatory region is sufficient to make a physiological impact. In fact, artificial contraction of D4Z4 repeats using CRISPR-Cas9 is sufficient to cause FSHD phenotype (*29*). The epigenetic regulation of DUX4 by D4Z4 repeats is unique and very complex; and little is known how D4Z4 evolved to control *DUX4* expression. However, it is suggested that D4Z4 array originated from amplification of a retrotransposed copy of a *DUXC* (*19*). Interestingly, besides the well-studied subtelomeric regions of D4Z4 at 4q and 10q, there are other D4Z4 homologs located elsewhere on human genome, mostly at centromere (chromosomes 3, 13, 14, 15, 21, 22, and Y)(*4*). It is not clear why other D4Z4 homologs exist as they are largely uncharacterized and do not have a complete *DUX4* ORF. Fascinatingly, D4Z4 homologs of 4q and 10q were indicated to undergo mitotic rearrangements(*30*, *31*). D4Z4 repeats duplication was also demonstrated in some FSHD patients with an extra shortened D4Z4 array in 4q which contributes to FSHD pathology(*32*). Indeed, sub-telomeric region is prone to chromosomal rearrangement more than any other genomic regions. It is not surprising that there could be some form of connection between the D4Z4 homologs. Furthermore, 4qA was implied to be the original form in human before duplication events happened that gave rise to other D4Z4 variants/haplotypes(*33*). Perhaps, more unique D4Z4 arrangements and 4q/10q haplotypes in human population and whether they have any effect on *DUX4* expression remain to be found.

Epigenetic regulation of DUX4

As discussed earlier, *DUX4* is immediately silenced following EGA and only was found to be expressed in testis and thymus of somatic tissues and also cancer cells so far. Silencing of *DUX4* is achieved by the heterochromatic D4Z4 repeat array. In general, DNA methylation and histone modification are essential for genome reprogramming during embryogenesis(*34*). D4Z4 heterochromatic mechanism to silence *DUX4* mainly consists of DNA methylation, H3K9 trimethylation (H3K9me3) and sometimes H3K27 trimethylation (H3K27me3) (*1*, *2*, *11*). Contraction of D4Z4 repeats to less than 10 units (FSHD1) enables de-repression/relaxation of D4Z4, allowing for aberrant active *DUX4* expression in adult muscle. Heterochromatic D4Z4 can also be disrupted by having a mutation for a specific gene modifier without a contraction (FSHD2). Such mutations were discovered so far to be SMCHD1 (structural maintenance of chromosomes hinge domain 1), LRIF1 (Ligand Dependent Nuclear Receptor Interacting Factor 1), and DNMT3B (DNA methyltransferase 3 beta)(*35*–*37*). Each of these mutations is involved in either DNA methylation or H3K9me trimethylation of D4Z4 repeats. Each D4Z4 repeat has high frequency of CpGs, allowing D4Z4 array to be hypermethylated in healthy individuals (10-100 units). Shortening of D4Z4 repeats (1-8 units) reduces instances of CpGs and results in hypomethylation of D4Z4 in FSHD1 patients. In fact, size of D4Z4 repeats array inversely correlates with the severity of FSHD as a low number of D4Z4 (1-3 units) could trigger an earlier disease onset(*38*–*40*). However, some evidence suggested that level of DNA methylation varies in patients and insufficient to predict the disease severity(*41*). Therefore, DNA methylation alone cannot be the determinant feature for disease diagnosis.

Indeed, together with DNA methylation, H3K9me3 significantly solidifies the heterochromatin of D4Z4. Moreover, H3K9me3 which is established by SUV39H1 was lost

in both FSHD1 and FSHD2 patients at D4Z4(*2*). Surprisingly, DNA methylation loss does not lead to H3K9me3 loss at D4Z4, implying that these two processes are independent of each other. H3K9me3 reduction not only occurs on the contracted disease allele but also on other non-contracted 4q and 10q D4Z4 alleles. As discussed previously above, there is some form of communication among the D4Z4 homologs. It is possible that 4q and 10q chromatin interact with each other in 3D space, disrupting heterochromatin of one could affect the other D4Z4. Furthermore, spreading of H3K9me3 is likely to occur as more H3K9me3 being recruited to nearby genomic region. It is not surprising that higher number of D4Z4 repeats (> 20 units) enhances this process to silence *DUX4* effectively as longer D4Z4 repeat array is able to recruit more H3K9me3. H3K9me3 expansion allows more recruitment of downstream silencing factors to stabilize heterochromatin domains on the body of the silenced gene in a cell-type specific manner. Such silencing factors include HP1γ and cohesin which are also recruited to D4Z4. In fact, the binding of HP1γ and cohesin to D4Z4 was found to be reduced in FSHD patients(*2*). Importantly, the binding of HP1γ and cohesin to D4Z4 is cell type-specific, suggesting that their binding is involved in cell type-specific chromatin organization. The other repressive histone mark H3K27me3 was also found in D4Z4 heterochromatin but not observed to be reduced in FSHD(*2*). However, H3K27me3 was found to be reduced specifically in DUX4-expressing cells, suggesting the stochastic nature of DUX4 epigenetic regulation by histone marks(*11*). Nonetheless, H3K9me3 seems to play a more critical role in maintenance of the D4Z4 heterochromatin as chromatin interaction of other D4Z4 homologs altogether spread H3K9me3, recruiting more silencing factors to repress *DUX4*.

Other gene modifiers of DUX4

SMCHD1 is classically known to be involved in X-chromosome inactivation specifically through DNA methylation(*42*, *43*). SMCHD1 belongs to a non-SMC family protein and consists of SMC hinge domain and an N-terminal ATPase domain. Generally, SMCHD1 homozygous mutation is lethal but a heterozygous mutation is haploinsufficient to cause D4Z4 hypomethylation in FSHD2 patients(*10*). Specifically, SMCHD1 mutation causes hypomethylation in both 4q and 10q; and SMCHD1 binding was also found to be reduced at D4Z4 in FSHD2 patients(*36*) .The likelihood of developing disease phenotype with SMCHD1 mutation is associated with low-moderate number of D4Z4 copies (typically 8-20 copies)(*40*). SMCHD1 might be involved in de-novo DNA methylation in early development as knocking down the gene in adult cells did not seem to have any effect on D4Z4 DNA methylation(*44*). It is unclear which mechanism of SMCHD1 plays a role in silencing *DUX4* during early development but reduced activity of SMCHD1 due to heterozygous mutation triggers an incomplete repression of *DUX4* in adult. It was also discovered that reduced H3K9me3 also reduces SMCHD1 binding at D4Z4 in FSHD1 patients(*4*). Another study further confirmed that SMCHD1 homozygous deletion in adult myoblast also reduces H3K9me3 at D4Z4 but not DNA methylation (*29*). Perhaps in adult tissues, SMCHD1 is not required for establishment nor maintenance of DNA methylation. In contrast, there is an interconnection between SMCHD1 and H3K9me3 in adult tissues as binding of SMCHD1 potentially recruits binding of H3K9me3 and vice versa. It is not clear whether SMCHD1 or H3K9me3 is the upstream effector. Interestingly, SMCHD1 is also known to be involved in another unrelated developmental disease called BAM (Bosma arhinia and micropthalmia)(*45*, *46*). SMCHD1 mutations that cause each disease tend to be

mutually exclusive though not entirely confirmed. Further studies still required to understand mechanisms of both disease etiology relating to SMCHD1 mutations. Homozygous mutation for LRIF1, another gene modifier, is also present in FSHD2 population. Similar to SCMHD1, LRIF1 is involved in X-inactivation and is indicated to employ H3K9me3 to D4Z4 repeat array. When completely lost, LRIF1 could cause de-represssion of *DUX4* in skeletal muscle(*37*). Particularly, the de-repression of *DUX4* by LRIF1 homozygous mutation is H3K9me3-depedent. This finding further emphasized the critical role of H3K9me3 in the silencing of *DUX4*. Additionally, DNMT3B (DNA methyltransferase) heterozygous mutation was found in a rare population of FSHD2 patients(*35*). Unlike LRIF1, DNMT3B is involved in DNA methylation, particularly in the de novo methylation of embryo and germline development(*47*). Interestingly, D4Z4 hypomethylation caused by DNMT3B mutation was first discovered in ICF1 (immunodeficiency, centromeric instability and facial abnormalities type 1), another unrelated disease(*35*). However, DNMT3B mutations are dominant mutations (heterozygous) in FSHD while recessive mutations (homozygous) in ICF1. FSHD phenotype is typically not observed in ICF1 patients due to their shorter life span before FSHD could develop at adolescence. At the current knowledge, there is no report of ICF1 carriers (DNM3TB heterozygous mutation) that develop FSHD because neither did the studied carriers inherit the permissive 4qA nor low-to-moderate D4Z4 units. It is also not fully confirmed if FSHD patients could have multiple inherited diseases. Nonetheless, all these mutations seem to be involved at the critical period of early development during which de novo epigenetic silencing of *DUX4* was disrupted, causing catastrophic disturbance in the

developing muscle tissues downstream in later adult life. Major epigenetic regulation of D4Z4 repeats is summarized in Figure 1.3.

There are still more players of D4Z4 gene modifiers being discovered. For instance, a study which developed a sophisticated technique termed enChIP-MS identified and characterized NuRD and CAF-1 complex as D4Z4-associated proteins. The NuRD and CAF-1 complex represses *DUX4* expression which can be reversed by MBD3L protein family(*48*–*50*). MBD3L family inhibits binding of MBD3 in the NuRD complex, releasing repression of *DUX4*. Meanwhile, a long non-coding RNA, DBE-T, which is located upstream of D4Z4 repeat was revealed to recruit ASH1L to de-repress *DUX4* in FSHD patients(*51*). Taken together, we propose that *DUX4* expression is very complex and highly regulated at many different levels, having one or more mutations for *DUX4* gene modifiers is likely to trigger de-repression of *DUX4*. This implication might explain why there are many distinct FSHD mutations that are associated with FSHD, highlighting the importance of each of the players that silences *DUX4*.

1.4 DUX4 gene network and the Homeobox family

DUX4 target genes

With the recent advance in DNA sequencing technologies in the past decades, huge effort was made to characterize FSHD transcriptome by multiple researchers. Several studies have uncovered DUX4-induced genes (DUX4 targets) using FSHD samples or ectopic expression of *DUX4* in normal tissues(*24*, *25*, *52*, *53*). Because DUX4 protein is present at a very low level in patients (< 0.1% myocytes), induced expression of *DUX4* is

necessary to identify and confirm DUX4 targets(*27*, *52*). Furthermore, DUX4 targets are more reliable and informative as biomarkers of DUX4 due to the difficulty to detect DUX4 in patient myocyte population(*17*, *24*, *52*). However, the majority of DUX4 gene signature is poorly characterized and some are unknown of functions. Some targets were confirmed to be involved in EGA and/or germline development (ZSCAN4, KDM4E, PRAMEF family, MBD3L, retroelements, etc.). Single-cell and single-nuclei RNA-seq studies in patient samples also identified and confirmed expression of DUX4 targets that overlapped across multiple FSHD studies(*28*, *54*). The single-cell and single-nuclei RNA-seq studies also identified distinct population of FSHD cells expressing higher DUX4 gene signature, implying the stochasticity and heterogeneity of DUX4 activation. Particularly, DUX4 target expression increases significantly as myoblasts differentiate into fused myotubes. RNA FISH study also suggested that *DUX4*-fl transcript can spread to other *DUX4*-negative nuclei of the same myotube to encode DUX4 protein that could be far away from the transcript source(*55*). *DUX4-fl* transcript also tent to localize inside nuclei while DUX4 target genes were much more wide-spread. Together, these findings further validated that DUX4 activation is stronger in differentiated myocytes as the effect of *DUX4-fl* transcript is global in fused myotube. It is unclear if DUX4 was designed to conserve this behavior in any other multinucleated cells but the nature of myotube fusion seems to boost DUX4 activation that possibly affect muscle more than other tissues. It is still possible that a small population of patient muscle cells might be free of DUX4 effect initially. However, FSHD clinical phenotype requires years to advance, indicating that DUX4 impact eventually affects the majority of myocyte population.

Due to genetic variation and difficulty to extract muscle biopsies in FSHD patients, and also because exogenous overexpression of DUX4 tends to be cytotoxic to cells, FSHD mutants were generated by using CRISPR-Cas9 to identify FSHD-specific gene signature (*29*). The mutant study was able to confirm DUX4 target expression from the artificial contraction of D4Z4 in the permissive myoblast of a healthy individual. Besides the DUX4 direct targets which were also confirmed by ChIP-seq, the study also found indirect DUX4 target genes and their related pathways (extracellular matrix, apoptosis, immune responses, embryonic development) upregulated in D4Z4-contracted mutants (*17*). Some of these pathways were also observed in FSHD patient studies(*17*, *53*). Furthermore, these genes were decreased in lesser extent than the DUX4 direct target genes by DUX4 shRNA depletion. It is unclear how indirect target genes are regulated by DUX4 and whether their misexpression have any effect on myogenesis. Indeed, myogenesis pathway was revealed to be inhibited by DUX4 activation(*17*, *26*, *56*). Taken altogether, these findings collectively suggested that low level of sporadic DUX4 is sufficient to cause a significant impact on further downstream genes/pathways in muscle cell population. This potentially explains why DUX4 could effectively disturb muscle differentiation program in FSHD patients as incompatible programs are being activated in muscle cells. It is not well-defined on how these DUX4-induced programs can affect myocyte differentiation though they might be partially associated with DUX4 biological functions in early embryonic development (EGA) or in germline. Future studies needed to examine the effect of DUX4-induced program/pathways in muscle development and to pinpoint the mechanism leading to DUX4 pathology.

DUX4 belongs to PRD-like class of homeobox family

The homeobox family is a very diverse group of genes that are commonly characterized by a homeodomain. The protein homeodomains are variable but most of them have 60 amino acids in length(*57*). Past studies have proposed to divide homeobox genes into 11 classes with over 100 gene families in animal(*57*, *58*). Homeobox genes are highly conserved throughout evolution and essential for early human development. A major example of homeobox gene class is the *HOX* gene family(*59*). The *HOX* gene family is well-known for its critical role in pattern formation along the anterior-posterior body axis. The double homeobox family including *DUX4* was proposed to belong to the PAIRED (PRD)-like class of the homeobox family(*58*). Interestingly, some of DUX4 targets such as DUXA, DUXB and LEUTX are also double homeobox and belong to PRD-like class. Furthermore, a recent single-cell study of early human development also found that DUXA, DUXB and LEUTX are indeed involved in EGA (*60*). Their expression is peak at 8-cell stage coincidentally after DUX4 peak at 4-cell stage. This evidence indicates another link between DUX4 and its immediate downstream effectors which are also PRD-like proteins. Additionally, DUXA, DUXB and LEUTX are also restricted to early human embryonic development(*23*). However, their function is not known completely because of their low expression which might challenge future studies to identify their binding sites. Interestingly, the double homeobox genes including *DUX4, DUXA and DUXB* is closely related to *sDUX* (a single-homeobox gene) from a common mammal ancestor(*20*). It was hypothesized that the evolutionary precursor of the DUX4 family was a single homeodomain protein(*21*). Primate-specific *DUX* family has probably evolved many times through duplication and retrotransposition as copies of DUX homologs have dispersed as

pseudogenes and have been uncovered in human genome by recent findings. It is not clear if these double homeobox pseudogenes are just remnants of evolution or they still have any biological functions in human. For example, ten *DUXA* pseudogenes (DUXAP1-10) were identified and particularly DUXAP8 as long-noncoding RNA was found to be involved in multiple cancers and also proposed as a cancer biomarker(*22*, *61*). Interestingly, DUXAP8 was also found to be upregulated in FSHD mutants (*29*). LEUTX was also proposed to evolve in human through tandem duplication and divergence from Otx family(*58*). An analysis of three separates single-cell studies of LEUTX also revealed that LEUTX target genes are involved in 8-cell embryo stage(*62*). Interestingly, NANOG is among the list of LEUTX upregulated targets. This implication might suggest for a connection between DUX4 network and the NANOG homeobox family(*62*). Nonetheless, with our current knowledge, we propose that the double homeobox DUX4, DUXA, DUXB and LEUTX cooperate by some undefined mechanism and play a critical role in early development. Future FSHD studies also should explore how active double homeobox genes in myocytes affects muscle development in adult.

Remarkably, PAX3/PAX7 also belongs to PRD-like class. PAX3/PAX7 is involved in embryonic myogenesis and muscle regeneration in satellite cells(*63*, *64*). PAX3/PAX7 was found to compete with DUX4 for genomic binding sites and could reduce DUX4 toxicity(*65*). DUX4 additionally was found to be at MYOD binding sites which might interfere with the function of PAX7, hence inhibiting MYOD targets and impairing muscle program. There is still a debate whether PAX7 or DUX4 plays a major role in FSHD pathology and whether PAX7 or DUX4-related pathway is a more accurate model for FSHD signature(*66*). The homeodomains of PAX7 and DUX4 contain similar sequences, this evidence potentially

explains their shared binding sites in human genome and the repression of PAX7 targets in FSHD due to DUX4 activity(*56*). Despite their antagonistic relationship, it is still possible that both DUX4 and PAX7 orchestrate as PRD-like factors to regulate early muscle development. Nevertheless, our current knowledge of the homeobox genes is still limited. We do not fully understand their interconnectedness and how they regulate each other in network. More studies needed to map their relationships particularly in early development where their function is essential.

DUX4 gene network

Many recent FSHD studies have characterized the DUX4 and its relationship with the DUX4 targets to understand DUX4 gene network and FSHD pathology(*8*, *17*, *24*, *25*, *28*, *52*, *54*, *55*). It is facilitating that low expression of DUX4 is capable to turn on multiple diverse pathways which have complicated and challenged many attempts to understand DUX4. Role of DUX4 in early embryonic development could be more characterized if examined together with its closely related homeobox family (DUXA, DUXB, LEUTX, NANOG, HOX, etc.). Their function could be interconnected as they work in orchestra to activate/inhibit developmental genes at an appropriate time event. The magnitude of gene activation or inhibition is also essential to control the level of gene expression to signal the rest of the network to follow. In addition, DUX4 is expressed in a brief burst during cleavage, hence its targets must carry out DUX4 program when DUX4 is no longer present(*13*). For example, DUXA and LEUTX which are novel transcription factors for early embryogenesis have been shown to act in positive feedback loop to enhance DUX4 activity

in FSHD mutants (*28*, *55*). Motif analysis also revealed that DUXA and LEUTX also share common binding sites with DUX4(*28*). In addition, DUXA is a later target than LEUTX because LEUTX expression is turned on earlier in differentiation and in greater magnitude than DUXA in terms of normalized TPM (transcript per millions). Therefore, the effect of LEUTX must be more significant than DUXA to enhance DUX4 activity initially during myoblast differentiation. But in later differentiation, DUXA potentially drives the DUX4 program by activating/inhibiting DUX4 targets or unknown genes. DUXA and LEUTX are still undercharacterized and their target genes are still under investigation. In addition, DUXB which is PRD-like class is also unknown of function besides its novel expression in early embryogenesis (8-cell stage). DUXB is also turned on in later differentiation in FSHD mutants, thus it might continue to control DUX4 program after DUXA peak. However, DUXB gene targets have not been identified. It is not clear if DUX4 network in FSHD muscle is similar to DUX4 gene network during early embryo development as distinct cell types might behave differently to DUX4 expression. Unfortunately, studying embryo during early development can be challenging due to ethical policies, hence DUX4 and its gene targets have only been investigated comprehensively in the context of FSHD.

Intriguingly, other non PRD-like class of DUX4 targets could also activate DUX4 program. H3.X/Y was discovered recently to enhance DUX4 targets by incorporation onto DUX4 target genes(*67*). H3.X/Y histone variants enable relaxation of chromatin, hence increasing DUX4 target gene expression. H3.X/Y specially can increase the perdurance and re-expression of DUX4 targets after a brief burst of DUX4 induction. H3.X/Y is also expressed in early embryo, indicating its important role in cleavage program(*67*). H3.X/Y expression was also clearly detected in FSHD double mutants as early as the myoblast stage

despite the fact that DUX4 target genes typically appear detectable during myoblast differentiation in FSHD (*29*). Moreover, depletion of H3.X/Y has almost similar effect as DUX4 depletion in FSHD mutants. Therefore, H3.X and H3.Y as DUX4 gene targets emerge as one of the earliest regulators of DUX4 program to enhance the network activation, possibly controlling downstream expression such as LEUTX and DUXA. MBD3L family (MBD3L2, MB3DL2B, MBD3L3) is also one of the earliest DUX4 targets which directly increases DUX4 expression by interfering the binding of MBD3 proteins in the NuRD complex and releasing DUX4 repression (*49*). Therefore, the MBD3L family is also involved in the enhancement of DUX4 initially. Taken together, we hypothesized that DUX4 relies on these early target genes in order to initiate DUX4 gene network and set up optimal parameters for gene expression of later targets such as LEUTX and DUXA. In turn, LEUTX and DUXA readjust the network by activation/inhibition of some sets of genes including DUX4 target genes (Figure 1.4). This hypothesis might explain the discordancy between DUX4 and its targets in RNA FISH study as the target genes still continue their expression without any presence of *DUX4* transcript in later differentiated myocytes(*55*). The hypothesis also explains why a short burst of DUX4 expression is sufficient to cause activation of a vast number of genes as some DUX4 target genes are chronologically turned on to amplify the network, taking turns to insert their task into DUX4 program on an unknown time scale that has not been yet characterized. We hope that more future studies will investigate the connection and relationship of the possible genes in the DUX4 network. This will not only serve to understand FSHD etiology but also to better solidify an interesting and complex gene network that is so critical in early embryonic cleavage program.

1.5 Conclusion:

As discussed above, DUX4 is a master regulator and also highly regulated at many different aspects (epigenetic regulation, D4Z4 repeat array length, DUX4 feedback network/perdurance, etc.). Perhaps, the biological design of DUX4 has accumulated throughout primate evolution by retrotransposition and duplication events as evidence from D4Z4 homologs and *DUX* pseudogenes which have been uncovered in many studies within the past decade. The advantages of *DUX4* divergence from the ancient *DUXC* have not been studied or proposed. Other mammals in the animal kingdom also carry some version of *DUX,* suggesting the conservational function of *DUX* in early development particularly during embryo cleavage across species. So far, DUX4 is known to be expressed normally in early embryo, testis and thymus. Low and burst expression of DUX4 is able to activate a pantheon of gene classes including retroelements, developmental and immune genes, germline genes and so on. Nonetheless, the functional mechanism of DUX4 program is not well-defined in embryogenesis but misexpression of DUX4 in muscle cell lineage is associated with progressive muscle loss in adult. Certainly, DUX4 gene network alone is very complex with many unknown relationships among DUX4 target genes. So far DUX4 is known to activate some of its target genes which also belong to PRD-like family (DUXA, DUXB, LEUTX) to govern embryogenesis and cell fate. Furthermore, there is a potential connection of DUX4 to other homeobox gene family such as NANOG to orchestrate embryonic development. Altogether, recent discoveries on DUX4 and its related genes have opened up to more interesting inquiry to understand human development.

1.6 Future aspect:

Because of the significant number of involved genes in the DUX4 network, it might require a lot of effort to understand the DUX4 program which has evolved since the earliest existence of mammal ancestors. Examination of all the dynamic relationships of DUX4 gene network is truly challenging with our raw mental capacity in order to comprehend the network. Therefore, I propose the use of artificial intelligence (AI) to solve for DUX4 gene regulatory network (GRN) model. Particularly provided with all the possible known genes in the network, I propose to solve for a simplified GRN model that can explain these parameters: (1) the type of connection between any pair of genes (nodes) at any given time internal: activation, inhibition or no relationship (none) and (2) magnitude of a connection (arbitrary value) (Figure 1.5). Such model must approximately fit TPM values provided from single-cell or RNA-seq time course data when gene expression changes over a discrete time scale (or pseudo-time in single-cell data) (Figure 1.5). The most optimal model will accurately define every connection between any pair of genes (parameters) to fit all the TPM values. DUX4 program could be a very good example to solve for a GRN provided with numerous published studies on DUX4 and its target gene expression. Some known relationships already have been defined in the DUX4 network from perturbance studies (H3.X/Y, DUXA, LEUTX) but other unknown relationships (activation/inhibition/none) and their magnitude for the rest of the genes in the network could be deduced from machine learning and AI. Recently, a concept from information theory called transfer entropy (TE) has been applied to single cell transcriptomic data to deduce GRN (*68*). TE values can reflect the degree of dependency of genes to its regulator. However, TE values cannot

distinguish whether a gene is inhibited or activated. Many more machine learning methods are currently still under investigation.

We are living at the time where advancement in technologies including artificial intelligence (AI) is accelerating. Though we are still in the very early phase of applying AI to research, I hope that future machine learning approaches will be able to solve for the model that I proposed. Further advanced methods will be required, and challenges will be met before any GNR can be solved effectively. If a model for a gene network like DUX4 network is achieved, it will open more doors to understand possibly many more different gene networks and systems in human biology.

1.7 Figure

Figure 1.1: DUX4 is activated at 2-cell stage following by target genes at 4-cell to 8-cell stage

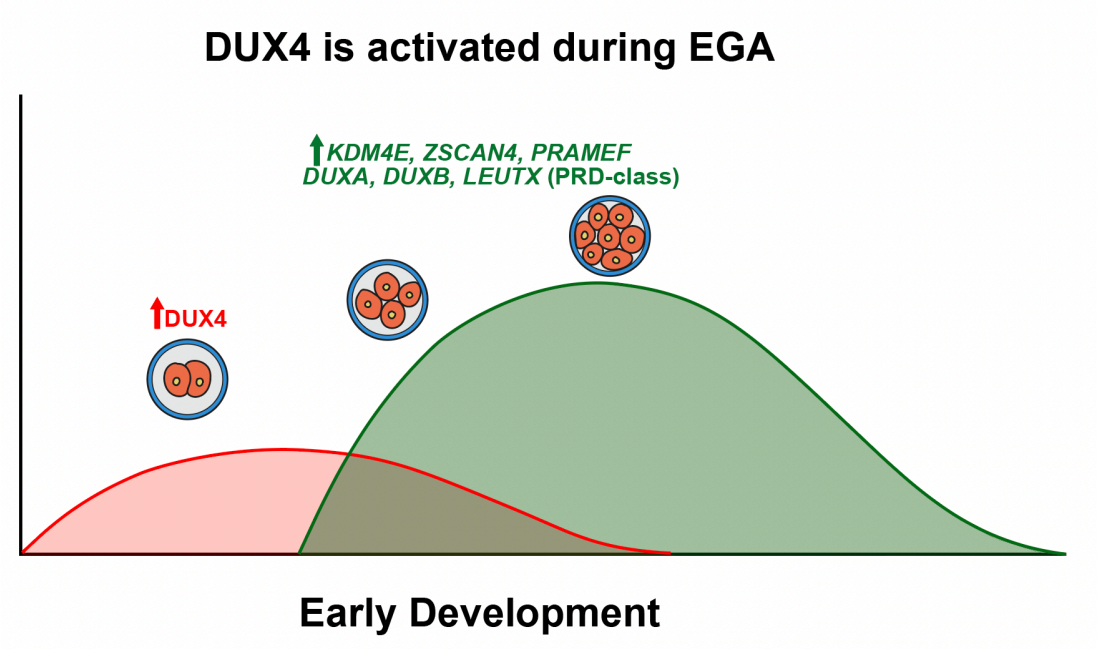


Figure 1.2: Evolution of DUX family in mammalians

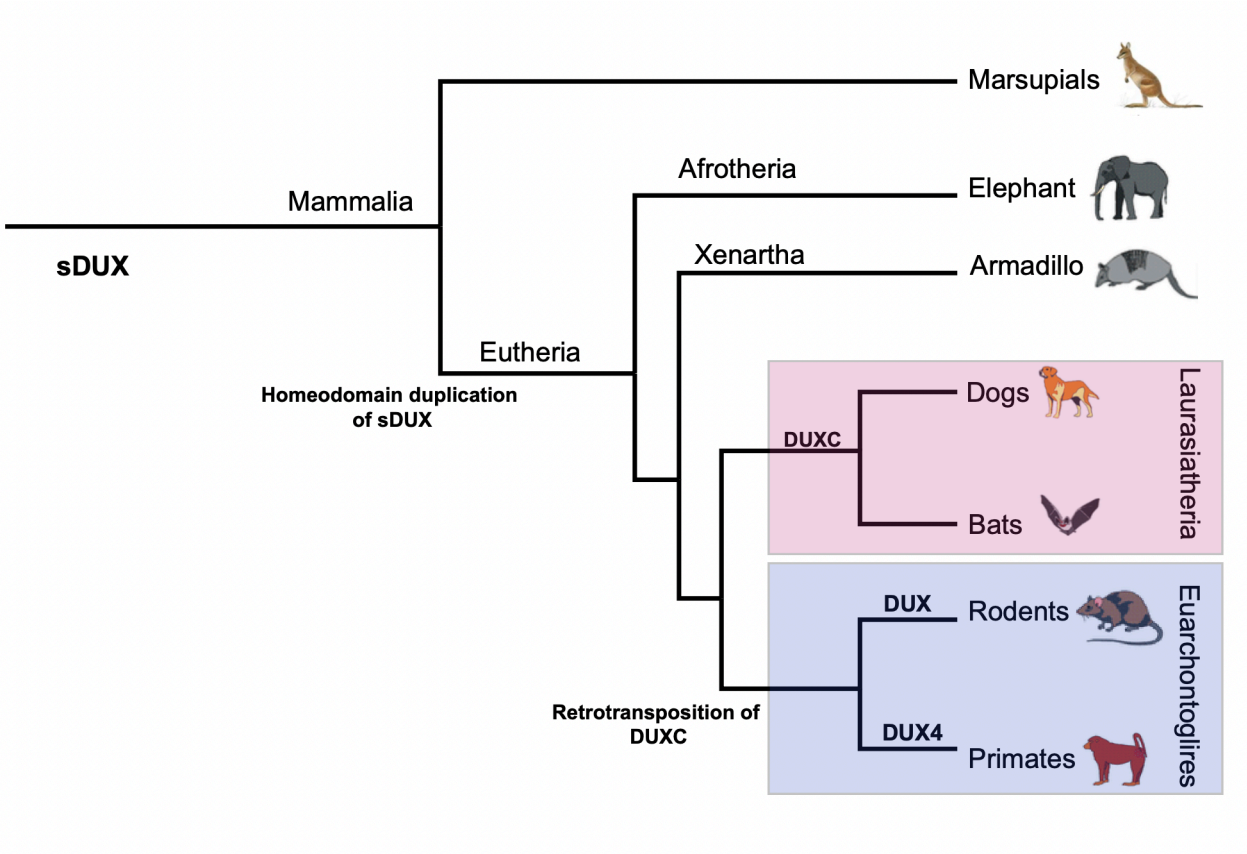


Figure 1.3: Epigenetic regulation of D4Z4 repeats and DUX4

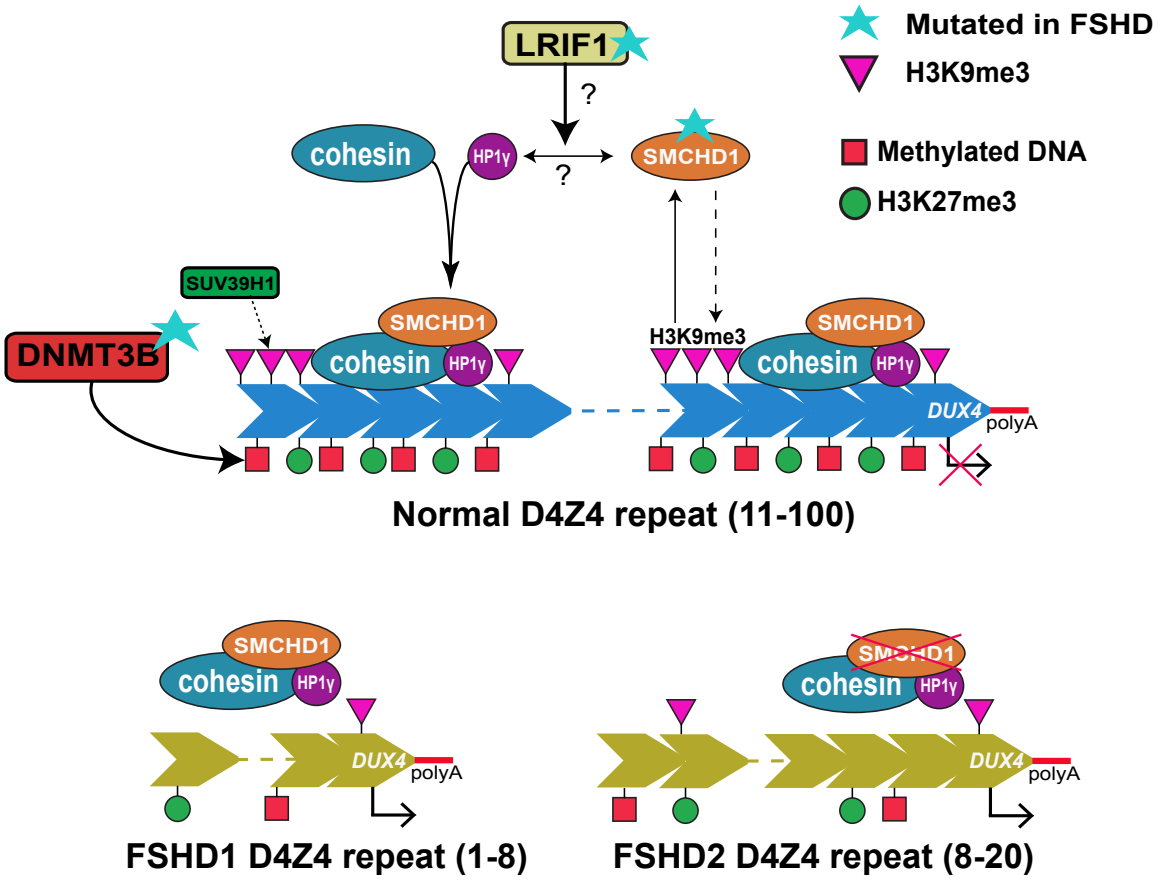


Figure 1.4: DUX4 target genes relationship

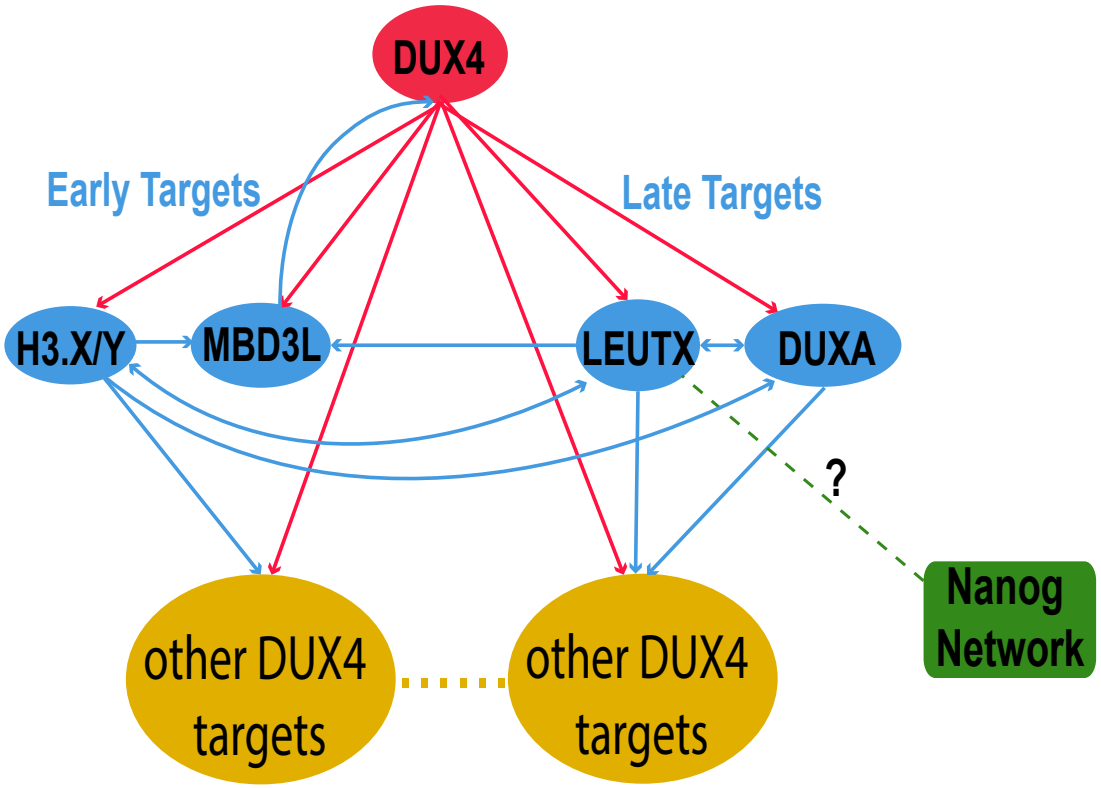
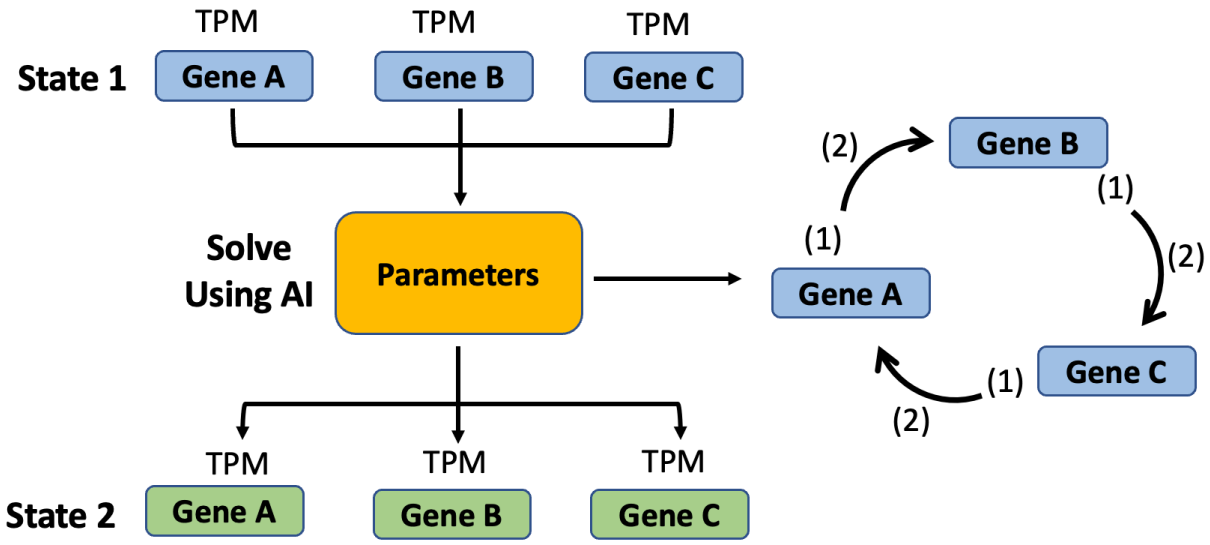


Figure 1.5: Gene Regulatory Network (GRN) method proposal



References:

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

2. W. Zeng, J. C. de Greef, Y.-Y. Chen, R. Chien, X. Kong, H. C. Gregson, S. T. Winokur, A. Pyle, K. D. Robertson, J. A. Schmiesing, V. E. Kimonis, J. Balog, R. R. Frants, A. R. Ball Jr., L. F. Lock, P. J. Donovan, S. M. van der Maarel, K. Yokomori, 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).

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

4. W. Zeng, Y. Y. Chen, D. A. Newkirk, B. Wu, J. Balog, X. Kong, A. R. Ball Jr., S. Zanotti, R. Tawil, N. Hashimoto, A. Mortazavi, S. M. van der Maarel, K. Yokomori, 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).

5. N. L. Vastenhouw, W. X. Cao, H. D. Lipshitz, The maternal-to-zygotic transition revisited. *Development*. **146**, dev161471 (2019).

6. J. L. Whiddon, A. T. Langford, C.-J. Wong, J. W. Zhong, S. J. Tapscott, Conservation and innovation in the DUX4-family gene network. *Nat Genet*. **49**, 935–940 (2017).

7. S. Das, B. P. Chadwick, Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. *PLoS One*. **11**, e0160022–e0160022 (2016).

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

9. L. Snider, L. N. Geng, R. J. L. F. Lemmers, M. Kyba, C. B. Ware, A. M. Nelson, R. Tawil, G. N. Filippova, S. M. van der Maarel, S. J. Tapscott, D. G. Miller, Facioscapulohumeral Dystrophy: Incomplete Suppression of a Retrotransposed Gene. *PLoS Genet*. **6**, e1001181- (2010).

10. S. Sacconi, R. J. L. F. Lemmers, J. Balog, P. J. van der Vliet, P. Lahaut, M. P. van Nieuwenhuizen, K. R. Straasheijm, R. D. Debipersad, M. Vos-Versteeg, L. Salviati, A. Casarin, E. Pegoraro, R. Tawil, E. Bakker, S. J. Tapscott, C. Desnuelle, S. M. van der Maarel, The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. *Am J Hum Genet*. **93**, 744–751 (2013).

11. P. Haynes, K. Bomsztyk, D. G. Miller, 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).

12. A. de Iaco, E. Planet, A. Coluccio, S. Verp, J. Duc, D. Trono, DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet*. **49**, 941–945 (2017).

13. P. G. Hendrickson, J. A. Doráis, E. J. Grow, J. L. Whiddon, J.-W. Lim, C. L. Wike, B. D. Weaver, C. Pflueger, B. R. Emery, A. L. Wilcox, D. A. Nix, C. M. Peterson, S. J. Tapscott, D. T. Carrell, B. R. Cairns, Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet*. **49**, 925–934 (2017).

14. K. Ishiguro, Y. Nakatake, N. Chikazawa-Nohtomi, H. Kimura, T. Akiyama, M. Oda, S. B. H. Ko, M. S. H. Ko, Expression analysis of the endogenous Zscan4 locus and its coding proteins in mouse ES cells and preimplantation embryos. *In Vitro Cell Dev Biol Anim*. **53**, 179–190 (2017).

15. L. Hillringhaus, W. W. Yue, N. R. Rose, S. S. Ng, C. Gileadi, C. Loenarz, S. H. Bello, J. E. Bray, C. J. Schofield, U. Oppermann, Structural and Evolutionary Basis for the Dual Substrate Selectivity of Human KDM4 Histone Demethylase Family *. *Journal of Biological Chemistry*. **286**, 41616–41625 (2011).

16. X. Liu, Y. Wang, Y. Gao, J. Su, J. Zhang, X. Xing, C. Zhou, K. Yao, Q. An, Y. Zhang, H3K9 demethylase KDM4E is an epigenetic regulator for bovine embryonic development and a defective factor for nuclear reprogramming. *Development*. **145**, dev158261 (2018).

17. L. N. Geng, Z. Yao, L. Snider, A. P. Fong, J. N. Cech, J. M. Young, S. M. van der Maarel, W. L. Ruzzo, R. C. Gentleman, R. Tawil, S. J. Tapscott, DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev Cell*. **22**, 38–51 (2012).

18. G. L. Chew, A. E. Campbell, E. de Neef, N. A. Sutliff, S. C. Shadle, S. J. Tapscott, R. K. Bradley, DUX4 Suppresses MHC Class I to Promote Cancer Immune Evasion and Resistance to Checkpoint Blockade. *Dev Cell*. **50**, 658-671.e7 (2019).

19. J. Clapp, L. M. Mitchell, D. J. Bolland, J. Fantes, A. E. Corcoran, P. J. Scotting, J. A. L. Armour, J. E. Hewitt, Evolutionary Conservation of a Coding Function for D4Z4, the Tandem DNA Repeat Mutated in Facioscapulohumeral Muscular Dystrophy. *The American Journal of Human Genetics*. **81**, 264–279 (2007).

20. A. Leidenroth, J. E. Hewitt, A family history of DUX4: Phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evol Biol*. **10** (2010), doi:10.1186/1471-2148-10-364.

21. C.-J. Wong, J. L. Whiddon, A. T. Langford, A. E. Belleville, S. J. Tapscott, Canine DUXC: implications for DUX4 retrotransposition and preclinical models of FSHD. *Hum Mol Genet*. **31**, 1694–1704 (2022).

22. H. A. F. Booth, P. W. H. Holland, Annotation, nomenclature and evolution of four novel homeobox genes expressed in the human germ line. *Gene*. **387**, 7–14 (2007).

23. E. Madissoon, E.-M. Jouhilahti, L. Vesterlund, V. Töhönen, K. Krjutškov, S. Petropoulos, E. Einarsdottir, S. Linnarsson, F. Lanner, R. Månsson, O. Hovatta, T. R. Bürglin, S. Katayama, J. Kere, Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. *Sci Rep*. **6**, 28995 (2016).

24. Z. Yao, L. Snider, J. Balog, R. J. L. F. Lemmers, S. M. van der Maarel, R. Tawil, S. J. Tapscott, DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum Mol Genet*. **23**, 5342–5352 (2014).

25. M. Ferreboeuf, V. Mariot, B. Bessières, A. Vasiljevic, T. Attié-Bitach, S. Collardeau, J. Morere, S. Roche, F. Magdinier, J. Robin-Ducellier, DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles, doi:10.1093/hmg/ddt409ï.

26. D. Bosnakovski, M. D. Gearhart, E. A. Toso, E. T. Ener, S. H. Choi, M. Kyba, Low level DUX4 expression disrupts myogenesis through deregulation of myogenic gene expression. *Sci Rep*. **8**, 16957 (2018).

27. K. Tsumagari, S.-C. Chang, M. Lacey, C. Baribault, S. v Chittur, J. Sowden, R. Tawil, G. E. Crawford, M. Ehrlich, Gene expression during normal and FSHD myogenesis. *BMC Med Genomics*. **4**, 67 (2011).

28. S. Jiang, K. Williams, X. Kong, W. Zeng, N. V. Nguyen, X. Ma, R. Tawil, K. Yokomori, A. Mortazavi, Single-nucleus RNA-seq identifies divergent populations of FSHD2 myotube nuclei. *PLoS Genet*. **16**, 1–26 (2020).

29. X. Kong, N. V. Nguyen, Y. Li, J. S. Sakr, K. Williams, S. Sharifi, J. Chau, A. Bayrakci, S. Mizuno, S. Takahashi, T. Kiyono, R. Tawil, A. Mortazavi, K. Yokomori, *bioRxiv*, doi:10.1101/2022.10.14.512332.

30. J. C. T. van Deutekom, E. Bakker, R. J. L. F. Lemmers, M. J. R. van der Wielen, E. Bik, M. H. Hofker, G. W. Padberg, R. R. Frants, Evidence for Subtelomeric Exchange of 3.3 kb Tandemly Repeated Units between Chromosomes 4q35 and 10q26: Implications for Genetic Counselling and Etiology of FSHD1. *Hum Mol Genet*. **5**, 1997–2003 (1996).

31. P. G. M. van Overveld, R. J. F. L. Lemmers, G. Deidda, L. Sandkuijl, G. W. Padberg, R. R. Frants, S. M. van der Maarel, Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. *Hum Mol Genet*. **9**, 2879–2884 (2000).

32. R. J. L. F. Lemmers, P. J. van der Vliet, J. P. Vreijling, D. Henderson, N. van der Stoep, N. Voermans, B. van Engelen, F. Baas, S. Sacconi, R. Tawil, S. M. van der Maarel, Cis D4Z4 repeat duplications associated with facioscapulohumeral muscular dystrophy type 2. *Hum Mol Genet*. **27**, 3488–3497 (2018).

33. R. J. L. F. Lemmers, P. J. van der Vliet, K. J. van der Gaag, S. Zuniga, R. R. Frants, P. de Knijff, S. M. van der Maarel, Worldwide Population Analysis of the 4q and 10q

Subtelomeres Identifies Only Four Discrete Interchromosomal Sequence Transfers in Human Evolution. *The American Journal of Human Genetics*. **86**, 364–377 (2010).

34. E. Li, Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet*. **3**, 662–673 (2002).

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

36. R. J. L. F. Lemmers, R. Tawil, L. M. Petek, J. Balog, G. J. Block, G. W. E. Santen, A. M. Amell, P. J. van der Vliet, R. Almomani, K. R. Straasheijm, Y. D. Krom, R. Klooster, Y. Sun, J. T. den Dunnen, Q. Helmer, C. M. Donlin-Smith, G. W. Padberg, B. G. M. van Engelen, J. C. de Greef, A. M. Aartsma-Rus, R. R. Frants, M. de Visser, C. Desnuelle, S. Sacconi, G. N. Filippova, B. Bakker, M. J. Bamshad, S. J. Tapscott, D. G. Miller, S. M. van der Maarel, Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet*. **44**, 1370–1374 (2012).

37. K. Hamanaka, D. Šikrová, S. Mitsuhashi, H. Masuda, Y. Sekiguchi, A. Sugiyama, K. Shibuya, R. J. L. F. Lemmers, R. Goossens, M. Ogawa, K. Nagao, C. Obuse, S. Noguchi, Y. K. Hayashi, S. Kuwabara, J. Balog, I. Nishino, S. M. van der Maarel, Homozygous nonsense variant in LRIF1 associated with facioscapulohumeral muscular dystrophy. *Neurology*. **94**, e2441–e2447 (2020).

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

39. R. Tawil, J. Forrester, R. C. Griggs, J. Mendell, J. Kissel, M. McDermott, W. King, B. Weiffenbach, D. Figlewicz, T. F.-D. Group, Evidence for anticipation and association of deletion size with severity in facioscapulohumerd muscular dystrophy. *Ann Neurol*. **39**, 744–748 (1996).

40. R. J. L. F. Lemmers, J. J. Goeman, P. J. van der Vliet, M. P. van Nieuwenhuizen, J. Balog, M. Vos-Versteeg, P. Camano, M. A. Ramos Arroyo, I. Jerico, M. T. Rogers, D. G. Miller, M. Upadhyaya, J. J. G. M. Verschuuren, A. Lopez de Munain Arregui, B. G. M. van Engelen, G. W. Padberg, S. Sacconi, R. Tawil, S. J. Tapscott, B. Bakker, S. M. van der Maarel, Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. *Hum Mol Genet*. **24**, 659–669 (2015).

41. A. Nikolic, T. I. Jones, M. Govi, F. Mele, L. Maranda, F. Sera, G. Ricci, L. Ruggiero, L. Vercelli, S. Portaro, L. Villa, C. Fiorillo, L. Maggi, L. Santoro, G. Antonini, M. Filosto, M. Moggio, C. Angelini, E. Pegoraro, A. Berardinelli, M. A. Maioli, G. D'Angelo, A. di Muzio,

G. Siciliano, G. Tomelleri, M. D'Esposito, F. della Ragione, A. Brancaccio, R. Piras, C. Rodolico, T. Mongini, F. Magdinier, V. Salsi, P. L. Jones, R. Tupler, Interpretation of the Epigenetic Signature of Facioscapulohumeral Muscular Dystrophy in Light of Genotype-Phenotype Studies. *Int J Mol Sci*. **21**, 2635 (2020).

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

43. A. V. Gendrel, A. Apedaile, H. Coker, A. Termanis, I. Zvetkova, J. Godwin, Y. A. Tang, D. Huntley, G. Montana, S. Taylor, E. Giannoulatou, E. Heard, I. Stancheva, N. Brockdorff, Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome. *Dev Cell*. **23**, 265–279 (2012).

44. J. Amiel, J. D. Robin, C. Laberthonnière, R. Bernard, A. D. Gurzau, C. T. Gordon, M.-C. Gaillard, J. Déjardin, V. Malan, B. Reversade, J. M. Murphy, A. Sefiani, L. Ratbi, A. Nowak, E. Salort Campana, C. Missirian, S. Xue, J. Dumonceaux, M. Thomas, A. Schlupp-Robaglia, B. Wollnik, C. El-Yazidi, V. Mariot, S. Attarian, K. Nguyen, N. Broucqsault, B. Binetruy, C. Dion, M. E. Blewitt, S. Roche, F. Magdinier, SMCHD1 is involved in de novo methylation of the DUX4 -encoding D4Z4 macrosatellite . *Nucleic Acids Res*, 1–18 (2019).

45. N. D. Shaw, H. Brand, Z. A. Kupchinsky, H. Bengani, L. Plummer, T. I. Jones, S. Erdin, K. A. Williamson, J. Rainger, A. Stortchevoi, K. Samocha, B. B. Currall, D. S. Dunican, R. L. Collins, J. R. Willer, A. Lek, M. Lek, M. Nassan, S. Pereira, T. Kammin, D. Lucente, A. Silva, C. M. Seabra, C. Chiang, Y. An, M. Ansari, J. K. Rainger, S. Joss, J. C. Smith, M. F. Lippincott, S. S. Singh, N. Patel, J. W. Jing, J. R. Law, N. Ferraro, A. Verloes, A. Rauch, K. Steindl, M. Zweier, I. Scheer, D. Sato, N. Okamoto, C. Jacobsen, J. Tryggestad, S. Chernausek, L. A. Schimmenti, B. Brasseur, C. Cesaretti, J. E. García-Ortiz, T. P. Buitrago, O. P. Silva, J. D. Hoffman, W. Mühlbauer, K. W. Ruprecht, B. L. Loeys, M. Shino, A. M. Kaindl, C.-H. Cho, C. C. Morton, R. R. Meehan, V. van Heyningen, E. C. Liao, R. Balasubramanian, J. E. Hall, S. B. Seminara, D. Macarthur, S. A. Moore, K. Yoshiura, J. F. Gusella, J. A. Marsh, J. M. Graham, A. E. Lin, N. Katsanis, P. L. Jones, W. F. Crowley, E. E. Davis, D. R. FitzPatrick, M. E. Talkowski, SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nat Genet*. **49**, 238–248 (2017).

46. C. T. Gordon, S. Xue, G. Yigit, H. Filali, K. Chen, N. Rosin, K. Yoshiura, M. Oufadem, T. J. Beck, R. McGowan, A. C. Magee, J. Altmüller, C. Dion, H. Thiele, A. D. Gurzau, P. Nürnberg, D. Meschede, W. Mühlbauer, N. Okamoto, V. Varghese, R. Irving, S. Sigaudy, D. Williams, S. F. Ahmed, C. Bonnard, M. K. Kong, I. Ratbi, N. Fejjal, M. Fikri, S. C. Elalaoui, H. Reigstad, C. Bole-Feysot, P. Nitschké, N. Ragge, N. Lévy, G. Tunçbilek, A. S. M. Teo, M. L. Cunningham, A. Sefiani, H. Kayserili, J. M. Murphy, C. Chatdokmaiprai, A. M. Hillmer, D. Wattanasirichaigoon, S. Lyonnet, F. Magdinier, A. Javed, M. E. Blewitt, J. Amiel, B. Wollnik, B. Reversade, De novo mutations in SMCHD1 cause Bosma arhinia

microphthalmia syndrome and abrogate nasal development. *Nat Genet*. **49**, 249–255 (2017).

47. F. Chédin, The DNMT3 Family of Mammalian De Novo DNA Methyltransferases. *Prog Mol Biol Transl Sci*. **101**, 255–285 (2011).

48. A. E. Campbell, S. C. Shadle, S. Jagannathan, J.-W. Lim, R. Resnick, R. Tawil, S. M. van der Maarel, S. J. Tapscott, NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *Elife*. **7**, e31023 (2018).

49. S. G. Jin, C. L. Jiang, T. Rauch, H. Li, G. P. Pfeifer, MBD3L2 Interacts with MBD3 and Components of the NuRD Complex and Can Oppose MBD2-MeCP1-mediated Methylation Silencing. *Journal of Biological Chemistry*. **280**, 12700–12709 (2005).

50. C. L. Jiang, S. G. Jin, D. H. Lee, Z. J. Lan, X. Xu, T. R. O'Connor, P. E. Szabó, J. R. Mann, A. J. Cooney, G. P. Pfeifer, MBD3L1 and MBD3L2, Two New Proteins Homologous to the Methyl-CpG-Binding Proteins MBD2 and MBD3: Characterization of MBD3L1 as a Testis-Specific Transcriptional Repressor. *Genomics*. **80**, 621–629 (2002).

51. D. S. Cabianca, V. Casa, B. Bodega, A. Xynos, E. Ginelli, Y. Tanaka, D. Gabellini, A Long ncRNA Links Copy Number Variation to a Polycomb/Trithorax Epigenetic Switch in FSHD Muscular Dystrophy. *Cell*. **149**, 819–831 (2012).

52. S. Jagannathan, S. C. Shadle, R. Resnick, L. Snider, R. N. Tawil, S. M. van der Maarel, R. K. Bradley, S. J. Tapscott, Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum Mol Genet*. **25**, 4419–4431 (2016).

53. A. M. Rickard, L. M. Petek, D. G. Miller, Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum Mol Genet*. **24**, 5901–5914 (2015).

54. A. van den Heuvel, A. Mahfouz, S. L. Kloet, J. Balog, B. G. M. van Engelen, R. Tawil, S. J. Tapscott, S. M. van der Maarel, Single-cell RNA sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. *Hum Mol Genet*. **28**, 1064–1075 (2019).

55. J. Chau, X. Kong, N. Viet Nguyen, K. Williams, M. Ball, R. Tawil, T. Kiyono, A. Mortazavi, K. Yokomori, Relationship of DUX4 and target gene expression in FSHD myocytes. *Hum Mutat*. **42**, 421–433 (2021).

56. D. Bosnakovski, E. A. Toso, L. M. Hartweck, A. Magli, H. A. Lee, E. R. Thompson, A. Dandapat, R. C. R. Perlingeiro, M. Kyba, The DUX4 homeodomains mediate inhibition of myogenesis and are functionally exchangeable with the Pax7 homeodomain. *J Cell Sci*. **130**, 3685–3697 (2017).

57. P. W. H. Holland, H. A. F. Booth, E. A. Bruford, Classification and nomenclature of all human homeobox genes. *BMC Biol*. **5**, 47 (2007).

58. P. W. H. Holland, Evolution of homeobox genes. *WIREs Developmental Biology*. **2**, 31–45 (2013).

59. T. R. Bürglin, "Homeodomain Subtypes and Functional Diversity" in *A Handbook of Transcription Factors*, T. R. Hughes, Ed. (Springer Netherlands, Dordrecht, 2011; https://doi.org/10.1007/978-90-481-9069-0_5), pp. 95–122.

60. V. Töhönen, S. Katayama, L. Vesterlund, E.-M. Jouhilahti, M. Sheikhi, E. Madissoon, G. Filippini-Cattaneo, M. Jaconi, A. Johnsson, T. R. Bürglin, S. Linnarsson, O. Hovatta, J. Kere, Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat Commun*. **6**, 8207 (2015).

61. C. Xue, X. Cai, J. Jia, Long Non-coding RNA Double Homeobox A Pseudogene 8: A Novel Oncogenic Propellant in Human Cancer. *Front Cell Dev Biol*. **9** (2021) (available at https://www.frontiersin.org/article/10.3389/fcell.2021.709069).

62. E.-M. Jouhilahti, E. Madissoon, L. Vesterlund, V. Töhönen, K. Krjutškov, A. Plaza Reyes, S. Petropoulos, R. Månsson, S. Linnarsson, T. Bürglin, F. Lanner, O. Hovatta, S. Katayama, J. Kere, The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. *Development*. **143**, 3459–3469 (2016).

63. F. Relaix, D. Rocancourt, A. Mansouri, M. Buckingham, A Pax3/Pax7-dependent population of skeletal muscle progenitor cells. *Nature*. **435**, 948–953 (2005).

64. M. Buckingham, F. Relaix, The Role of Pax Genes in the Development of Tissues and Organs: Pax3 and Pax7 Regulate Muscle Progenitor Cell Functions. *Annu Rev Cell Dev Biol*. **23**, 645–673 (2007).

65. D. Bosnakovski, Z. Xu, E. Ji Gang, C. L. Galindo, M. Liu, T. Simsek, H. R. Garner, S. Agha-Mohammadi, A. Tassin, F. Coppée, A. Belayew, R. R. Perlingeiro, M. Kyba, An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J*. **27**, 2766–2779 (2008).

66. C. R. S. Banerji, P. S. Zammit, 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).

67. R. Resnick, C. J. Wong, D. C. Hamm, S. R. Bennett, P. J. Skene, S. B. Hake, S. Henikoff, S. M. van der Maarel, S. J. Tapscott, DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. *Cell Rep*. **29**, 1812-1820.e5 (2019).

68. J. Kim, S. T. Jakobsen, K. N. Natarajan, K.-J. Won, TENET: gene network reconstruction using transfer entropy reveals key regulatory factors from single cell transcriptomic data. *Nucleic Acids Res*. **49**, e1–e1 (2021).

Chapter 2

Engineered SMCHD1 and D4Z4 mutations reveal roles of D4Z4 heterochromatin disruption and feedforward DUX4 network activation in FSHD

2.1 Abstract

Facioscapulohumeral dystrophy (FSHD) is commonly associated with contraction of D4Z4 repeats on chromosome 4q (FSHD1). Mutations in the *SMCHD1* gene are linked to both minor cases with no prominent repeat loss (FSHD2) and severe cases of FSHD1. Abnormal upregulation of the transcription factor DUX4, encoded in the *D4Z4* repeat, is believed to play a central role in FSHD. However, defining the disease mechanism has been hampered by the heterogeneity of patient-derived cells, difficulty to detect DUX4 in patient myocytes, and limited animal models because D4Z4 repeats are primate-specific. To overcome these limitations, we engineered isogenic human skeletal myoblast lines with D4Z4 and/or *SMCHD1* mutations. We found a highly synergistic effect of double mutations on triggering two key disease processes, D4Z4 heterochromatin disruption and cross-stimulation of DUX4 targets, such as histone H3.X/Y and LEUTX transcription factor. Thus, engineered human myocyte models provide unique insights into the molecular mechanisms underpinning FSHD.

2.2 Introduction

Facioscapulohumeral dystrophy (FSHD) is one of the most common muscular dystrophies with a prevalence of ~1 in 8,333. FSHD causes progressive wasting of facial, shoulder, and upper arm as well as lower leg musculature (1, 2). The majority of FSHD

cases (>95%) are caused by monoallelic contraction of 3.3 kb D4Z4 macrosatellite repeat sequences located at the subtelomeric region of chromosome 4q (4qter D4Z4) (FSHD1 (MIM 158900)) (3, 4). FSHD1 is associated with 1~10 copies of D4Z4 repeats in the contracted allele in contrast to 11~150 copies in the intact allele. Lower copy numbers (1~3 copies) are tied to earlier onset and more severe phenotypes compared to higher copy numbers (8~10 copies). However, clinical manifestations are variable, suggesting a potential contribution of additional modifier gene(s) (2, 5). FSHD2, which is the rare form of FSHD (<5% of cases), is mainly linked to mutations of *SMCHD1* (MIM 158901) (2, 6). Although FSHD2 was thought previously to involve no D4Z4 repeat contraction, accumulating evidence indicates its association with relatively short D4Z4 repeats (8~20 repeats) (2). Moreover, mutations of the *SMCHD1* gene have been found in severe cases of FSHD1, suggesting that it may act as a modifier gene to increase the disease severity (7, 8).

D4Z4 contains an open reading frame for the double-homeobox transcription factor *DUX4* gene (9-11). *DUX4* gene is normally expressed early in embryogenesis and participates in EGA (12, 13). Expression of short and long isoforms have been detected but only the full-length *DUX4* transcript (*DUX4fl*) that includes a transactivation domain can activate target genes, and its abnormal upregulation is associated with FSHD (11, 14, 15). The *DUX4* gene, embedded in the D4Z4 repeat, lacks a poly-adenylation (poly(A)) signal sequence, and only those individuals with "permissive" 4qA haplotypes carrying the canonical poly(A) signal downstream of the last D4Z4 repeat develop FSHD, strongly suggesting that the disease is tightly linked to functional *DUX4* mRNA production from the last D4Z4 copy (14). Indeed, upregulation of DUX4 target genes is readily detectable in patient cells, supporting the significance of DUX4-mediated gene activation in FSHD.

Curiously, however, *DUX4fl* RNA and protein expressions are extremely infrequent and occur at low levels in patient myocytes, albeit higher than in control cells (11, 14-16). Consequently, expression of DUX4 target genes, rather than *DUX4* itself, is used as markers for the FSHD phenotype (5, 10, 17-20). However, dynamics and regulation of the DUX4 target gene network activation is not well understood.

Hypomethylation of D4Z4 DNA is a signature change in FSHD patient cells (21-23). We also found that D4Z4 repeats contain heterochromatic regions marked by histone H3 lysine 9 trimethylation (H3K9me3), heterochromatin binding protein HP1γ and the higher-order chromatin organizer cohesin (24). This heterochromatin structure is compromised in both FSHD1 and FSHD2 (24). Indeed, reduction of H3K9me3 at D4Z4 by inhibition or depletion of SUV39H1 causes *DUX4fl* expression (24, 25). Thus, perturbation of heterochromatinization of D4Z4 appears to be directly linked to FSHD pathogenesis (15, 26). SMCHD1 binds to D4Z4, and its haploinsufficiency results in derepression of *DUX4fl* expression, indicating a direct role of SMCHD1 in *DUX4fl* regulation (6). We found that SMCHD1 binding to D4Z4 is H3K9me3-dependent (25). This raised the possibility that even in FSHD1 with no mutation in *SMCHD1*, SMCHD1 binding to D4Z4 may be compromised (due to the loss of H3K9me3), which may contribute to *DUX4* upregulation (25). In severe cases of FSHD1, this effect may be further exacerbated by the actual mutations in *SMCHD1* itself (7, 8). SMCHD1 has been implicated in regulation of DNA methylation at certain CpG islands and at the inactive X chromosome (27-29), and thus, it is assumed to also regulate DNA methylation at D4Z4. However, SMCHD1 represses gene expression in both DNA methylation-dependent and independent ways (30), and whether SMCHD1 modulates DUX4fl expression through DNA methylation has not been determined.

Overexpression of recombinant DUX4 in *in vitro* myoblasts and in *in vivo* model organisms is highly toxic (31, 32), and there is an ongoing effort to characterize the mechanism of DUX4-induced cell toxicity (33-36). However, the frequency of DUX4 expression in patient cells is often less than 1%, which is substantially lower than that in the recombinant DUX4-inducible systems. In fact, recent single cell/nucleus-sequencing and in situ RNA detection analyses revealed no significant evidence for cell death in FSHD patient myocytes that endogenously express DUX4 (37, 38). Moreover, the subcellular localization of the recombinant *DUX4fl* mRNA differs from that of the endogenous *DUX4fl* mRNA (37, 39). Collectively, these observations raise the question whether DUX4-induced acute cytotoxicity is a physiologically relevant mechanism of disease pathogenesis. To circumvent these issues, efforts are being made to express DUX4 at low level in an inducible fashion in mice (40, 41). D4Z4 repeats (and *DUX4* within), however, are primate-specific. Likewise, DUXA and LEUTX, two main transcription factors activated by DUX4, are absent in mice (42). Consequently, the genes and molecular network activated by human *DUX4* introduced in non-primate model organisms are different from those in patient muscle. Thus, a physiologically relevant model for FSHD is still lacking.

Crucially, whether genetic changes (D4Z4 contraction and/or mutations in *SMCHD1*) are sufficient to recapitulate patient phenotypes remains undetermined. To experimentally address this and to further investigate the disease mechanism, here, we report the development of CRISPR-engineered isogenic mutant myoblast cell lines carrying either D4Z4 mutations, SMCHD1 mutation or both. Analyses of these cell lines provided evidence for a feedback loop between SMCHD1 and H3K9me3 at D4Z4 and demonstrated a synergistic effect of double mutations on DUX4 target gene induction. Additional

heterochromatin disruption enhanced the effect of individual mutations. Furthermore, our results uncovered differentiation-insensitive early and -sensitive late DUX4 target genes and their positive cross-regulation. Taken together, our analyses of isogenic FSHD mutant myocytes reveal heterochromatin disruption and feedforward gene expression network downstream of DUX4 as two key processes underlying FSHD pathogenesis.

2.3 Results

Elimination of SMCHD1 has a minor effect on *DUX4* expression

FSHD occurs only in individuals with "permissive" 4qA haplotypes, in which the presence of a poly(A) signal downstream of the last D4Z4 copy would allow the expression of functional DUX4fl mRNA (14). To compare the effects of haplotypes, SMCHD1 shRNA depletion was initially performed in healthy control cell lines with non-permissive and permissive haplotypes (Supplemental Fig. S2.1). Interestingly, in both cell types, this caused upregulation of DUX4 target genes at the myoblast stage in a statistically significant and comparable fashion compared to control shRNA-treated cells, suggesting that a low level functional *DUX4* mRNA can be expressed in non-permissive cells (Supplemental Fig. S2.1A). Notably, DUX4 target gene expression was further stimulated upon myotube differentiation in permissive haplotype cells. In contrast, despite comparable SMCHD1 depletion efficiency, no further stimulation was observed in non-permissive haplotype myotubes (Supplemental Fig. S2.1B). Thus, the permissive haplotype is critical for differentiation-induced stimulation of DUX4 and target gene expression.

Even in cells with a permissive haplotype, DUX4 target gene expression in SMCHD1-depleted myotubes is much lower than that in FSHD2 myotubes (Supplemental Fig. S2.1C).

Thus, transient SMCHD1 depletion is not sufficient for a robust FSHD phenotype. To generate stable mutant cell lines, gRNAs specific for *SMCHD1* were designed for CRISPR knockout (KO) (Supplemental Fig. S2.2A). Despite screening 300 clones, we failed to obtain heterozygous mutant cells due to high efficiency of CRISPR mutation. Thus, unlike FSHD2 cells, in which *SMCHD1* mutation is heterozygous, our *SMCHD1* (SM) mutant cells are *SMCHD1* null (Fig. 2.1A). As in HCT116 colorectal cancer cells (43), *SMCHD1* KO in adult myoblasts is not lethal. *SMCHD1* mutation in permissive haplotype cells upregulated DUX4 target genes in a statistically significant fashion compared to the parental control cells (Fig. 2.1D). However, even after a complete loss of SMCHD1, the amount of DUX4 target gene transcripts was much lower than that in FSHD2 patient cells (which harbor a heterozygous *SMCHD1* mutation). Thus, by itself, a somatic KO mutation of *SMCHD1* is not sufficient to recapitulate FSHD2 (Fig. 2.1E).

D4Z4 contraction induces variegated DUX4 target gene expression, which is enhanced by concurrent loss of SMCHD1

A set of gRNAs was designed at the single-nucleotide polymorphism (SNP) regions to enhance mutations of 4q over 10q D4Z4 repeats (DEL mutants) (Fig. 2.1A-C and Supplemental Fig. S2.2B). Parental control cells have 24 and 18 copies of D4Z4 on chromosome 4q and 24 and 12 copies on chromosome 10q based on PFGE (Fig. 2.1B). We found that CRISPR-CAS9-mediated D4Z4 disruption was sufficient to upregulate DUX4 target genes (e.g., TRIM43, LEUTX and MBD3L2) in early myotubes, which was therefore used as a screening phenotype (Fig. 1D and Supplemental Fig. S2.2B). We generated DEL mutant clones using DNA plasmid- and protein/RNA-based CRISPR-Cas9 systems by

screening ~300 and 180 clones, respectively (see Methods and Supplemental Figure S2.2B). We observed a tendency of inverse correlation between high DUX4 target gene expression and efficient myotube differentiation (Supplemental Fig. S2.2C). For our experiments, we selected clones with a relatively high DUX4 target gene expression as well as efficient proliferation and differentiation capabilities (Supplemental Fig. S2.2C). We found that D4Z4 mutant (DEL) clones showed higher DUX4 target gene expression than SM clones, but still much lower than FSHD patient cells (DEL3 as an example in Fig. 2.1D and E). Importantly, target gene upregulation is DUX4-dependent as DUX4 shRNA depletion abolished the expression of target genes (Fig. 2.1F). We found that target gene expression is highly variable even in the same clone with different passage numbers and/or in different experiments (Fig. 2.1G).

Although PFGE analyses indicated repeat contraction (Fig. 2.1B), we analyzed the regions and resulting transcripts using genomic and RNA nanopore long-read sequencing, respectively. For DEL mutants generated by the plasmid-based CRISPR-CAS9 treatment, we found that gRNA-mediated cutting of D4Z4 repeats in the 4qA allele resulted not only in deletion of repeats upstream of the last copy, but also in insertion of inverted 2.5 copies of D4Z4 repeats as well as 2 copies of a gRNA plasmid sequence separated by a small fragment of a CAS9 plasmid sequence (Fig. 2.1C; Supplemental Fig. S2.2D). These inverted repeats could give rise to the ~2 repeat signal in PFGE when in fact only the last copy was left downstream of this insertion, thus creating the 4qA allele with one D4Z4 repeat (Fig. 2.1B). For 4qB, we found the repeat shortening and inversion, the length consistent with the PFGE band, which would not yield any significant DUX4 gene expression due to the absence of canonical poly(A) signal sequence (14) (Fig. 2.1C). No rearrangements were

detected in 10q D4Z4 alleles, consistent with the expectation with 4q-tailored gRNAs (Fig. 2.1C). The insertion of plasmid sequences resulted in expression of EGFP from the PKG promoter as well as small spliced RNA fragments containing *DUX4* exons 2 and 3 (corresponding to the 3'-UTR of *DUX4* mRNA) (termed a chimeric DUX4 3'-UTR) (Supplemental Fig. S2.2D). We confirmed that overexpression of the chimeric DUX4 3'UTR fragment has no effect on DUX4 target gene expression (Supplemental Fig. S2.2E). In contrast, protein/RNA-based CRISPR-Cas9 mutagenesis resulted in repeat contraction down to one copy of D4Z4 at 4qA and 2 copies at 4qB, and also reduced the copy number of 10qD4Z4 alleles to one copy as confirmed by nanopore sequencing (Fig. 2.1C, DEL9). DEL9 cells exhibit comparable upregulation of DUX4 target genes as the plasmid-based DEL mutants (Fig. 2.1H, *MBD3L2* as an example). Consistent with the one intact *DUX4* gene in the 4qA allele, RNA nanopore sequencing of both types of DEL mutant clones confirmed an intact wild type *DUX4fl* transcript, indicating that the last D4Z4 copy at 4qA retained the ability to express *DUX4fl* (Supplemental Fig. S2.2D). Thus, we used both of these mutant clones for further analyses.

The above results demonstrate that D4Z4 mutations cause upregulation of *DUX4* and target genes, but their expression is highly variegated and tends to be much lower than in FSHD patient cells (Fig. 2.1E). In FSHD1 patients, the lower D4Z4 repeat numbers associate with more severe clinical phenotypes (2, 44). Our results suggest, however, that even when only one repeat copy left, additional mechanism(s) are required to recapitulate the full FSHD-associated gene expression program. As *SMCHD1* mutations are associated with severe cases of FSHD1 (combined with D4Z4 contraction) (7, 8), we generated double mutant cells (DEL_SM), by introducing *SMCHD1* mutation in DEL clones (Supplemental Fig.

S2.2B). In comparison to single DEL mutants, DEL_SM mutants consistently upregulated DUX4 target genes at a higher level (Fig. 2.1E and I). These results demonstrate that *SMCHD1* acts as a modifier gene whose loss acts synergistically with D4Z4 contraction to enhance and stabilize the DUX4 target gene expression.

Synergistic effect of double mutations recapitulates patient cell phenotype

To determine the effects of these engineered mutations, we analyzed the genome-wide gene expression changes during myoblast differentiation by RNA-seq of 3-5 clones of each mutant type (SM, DEL and DEL_SM) in comparison to the isogenic control as well as FSHD1 and FSHD2 patient cells. Although we chose clones with relatively efficient differentiation, we observed prominent delays in differentiation of DEL and DEL_SM clones even in those with comparable doubling time (Supplemental Fig. S2.3A). A mild delay was also observed in FSHD patient cells whereas the delay was minimal in *SMCHD1* mutant (SM) cells. Time course principal component analyses (PCA) of gene expression during differentiation also confirmed delays in DEL and DEL_SM mutant cells compared to the isogenic control and SM cells (Supplemental Fig. S2.3B). To compensate for the differentiation delay, we compared DEL and DEL_SM samples at days 4∼5 and 13∼14 of with control and SM samples on days 3 and 12 of differentiation as “early” and “late” myotubes, respectively, in the current study.

Using edgeR, we identified up- and downregulated differential expressed genes (DEGs) (*P*-values <0.01 and Log_2 fold change > 1.5) in each mutant type (Supplemental Table S2.1). Interestingly, at both myoblast and early myotube stages, the number of genes that are upregulated in the DEL_SM mutant cells is significantly more (673 and 1228,

respectively) than the sum of de-repressed genes in the DEL- or SM-only mutants combined (203 and 228, respectively) (Fig. 2.2C and Supplemental Fig. S2.4A). PCA of myoblast and early myotube stages revealed that DEL_SM clones and FSHD patient myocytes cluster together (Fig. 2.2A). While PC1 (49.2% variance) largely separates myoblast and differentiated myotube stages, there is a significant clustering of patient and DEL_SM myotubes comparing to the control and single mutants, and two clusters are further separated in PC2 (14.0% variance) (Fig. 2.2A). The top DEGs in PC1 include myogenesis genes that are downregulated (Supplemental Table. S2). Box plots show that albeit at varying degrees, these genes are commonly downregulated in all mutants at early myotube stage (Fig. 2.2B, top; Supplemental Table S2.1). Thus, all three types of mutations (DEL, SM and DEL_SM) impede expression of muscle genes. The top mis-expressed genes in PC2 include a subset of DUX4 target genes that are significantly upregulated in patient and DEL_SM, but not in single mutant, early myotubes (Fig. 2.2B, bottom; Supplemental Fig. S2.4B).

Expression of 64 representative DUX4 target genes (38, 42) increased in DEL_SM during differentiation and became predominant in late myotubes (Fig. 2.2C, orange). Hierarchically clustered heatmaps in three types of mutant clones compared to isogenic control and FSHD patient myocytes indicated that DUX4 target gene expression is strongest at late myotube stage in DEL, DEL_SM and patient cells (Fig. 2.2D). In SM mutants, however, DUX4 target expression was over all very low throughout differentiation (Fig. 2.2D). Consistent with the RT-qPCR analyses (Fig. 2.1E and I), target gene expression is highly prominent in DEL_SM mutants, which is clear even in myoblast and early myotube stages (Fig. 2.2D; Supplemental Figs. S2.4B and S2.5A). This is in a stark contrast to DEL

mutants, in which target gene expression is relatively weak in early stages (Fig. 2.2D; Supplemental Fig. S2.4B and S2.5A). These results further highlight the synergy between D4Z4 contraction and SMCHD1 KO.

Characteristics of non-DUX4 target DEGs are recapitulated in mutant cells

We performed Gene Ontology (GO) enrichment analysis on the DEGs from DEL_SM for three stages of differentiation (myoblasts, early and late myotubes) (Fig. 2.3A and B; Supplemental Table S2.3). While similar GO terms were enriched for downregulated genes at all three stages, we found largely distinct GO term enrichment for upregulated genes at each stage, suggesting that upregulation of DEGs are closely linked to differentiation stages of myocytes (Fig. 2.3A and B).

Notably, we observed prominent upregulation of genes related to extracellular matrix (ECM), immune response, ER stress, apoptosis, and embryonic genes as well as down regulation of muscle-related genes across patient and mutant cells (Fig. 2.3C). Similar gene expression changes have been reported in FSHD patient myocytes (16, 35, 45) and some in the recombinant DUX4 overexpression study (10). Our results indicate that D4Z4 and *SMCHD1* mutations are sufficient to recapitulate this patient gene expression phenotype. Importantly, DUX4 depletion reversed these changes of representative genes, strongly suggesting that most of these changes are triggered by the mutation-induced DUX4 expression (Fig. 2.3D). However, DUX4 may regulate these genes indirectly as the depletion effect was not as robust as that on defined DUX4 target genes (Fig. 2.1F).

Both positive and negative regulators of apoptosis are upregulated especially in patient, DEL and DEL_SM myoblasts and early myotubes (e.g., *BAX, BAK1, FAS* and *FADD* for

pro-apoptosis, and *CITED2, ANKLE2*, and *ARAF* for cell survival) (Supplemental Table S2.3). Unexpectedly, these gene expression changes appear to taper off in late myotubes when DUX4 target genes are most highly expressed (Figs. 2.2C and D, 2.3C). This is consistent with our previous observations that no significant evidence for apoptosis/necrosis was detected in FSHD patient cells (data not shown) (37, 38). These results argue against the cytotoxic fate of DUX4-activated cells.

DEL mutations diminish D4Z4 heterochromatin, which is exacerbated by H3K9me3 reduction induced by SMCHD1 KO

Previously it was shown that disruption of heterochromatin integrity at the *DUX4* promoter region in the D4Z4 repeat is linked to DUX4 de-repression in both FSHD1 and FSHD2 myoblasts (24, 25). To interrogate the status of H3K9me3 and DNA methylation at the DUX4 promoter region, we performed ChIP-qPCR and MeDIP, respectively, in several clones each of mutant myoblasts (Fig. 2.4A and B). The reduction in H3K9me3 was observed in both DEL and SM, and was more substantial and consistent in DEL_SM (Fig. 2.4A). Thus, the loss of *SMCHD1* affects H3K9me3 at the *DUX4* promoter, and further enhances H3K9me3 reduction in D4Z4 mutants, correlating with their synergy on DUX4 and target gene expression (Figs. 2.2 and 2.4A). We previously showed that SMCHD1 binding to D4Z4 is H3K9me3-dependent (25). Thus, combined, our results uncover a positive feedback relationship between SMCHD1 and H3K9me3 at D4Z4.

In contrast to FSHD2 patient cells carrying *SMCHD1* mutations, loss of *SMCHD1* in our cell lines did not affect DNA methylation (Fig. 2.4B). Thus, a somatic mutation of *SMCHD1* by itself does not recapitulate the DNA methylation changes observed in FSHD2

myoblasts. The relatively subtle changes of MeDIP signals in our D4Z4 mutant cells (and no apparent change in FSHD1 patient cells) may be due to the fact that PCR primers do not distinguish 4q and 10q D4Z4. While DNA hypomethylation was shown to occur at both 4q and 10q alleles in FSHD2 cells, the changes occurred only in the contracted 4qA D4Z4 in FSHD1 cells (21-23). To distinguish different alleles, we separated reads specific to 4qA, 4qB and 10q D4Z4 based on their SNPs (24) and analyzed the read count ratio of 4q/10q in MeDIP compared to input genomic DNA at the *DUX4* promoter region (Fig. 2.4C). This allowed us to observe a significant decrease of MeDIP signals in FSHD1, DEL and DEL_SM mutants at 4qA D4Z4 over 10q, but not at 4qB over 10q (Fig. 2.4C). Importantly, we observed no significant difference between control and SM1 mutant cells as well as between DEL and DEL_SM cells (Fig. 2.4C). These results suggest a negligible contribution of *SMCHD1* KO to the maintenance of D4Z4 DNA methylation in adult myocytes.

An earlier study suggested that SMCHD1 is important for de novo methylation of D4Z4 repeats (43). Therefore, we compared recovery of DNA methylation in control- or SM1 mutant myoblasts after treatment with the DNA methylation inhibitor, 5AzaC. Following treatment with 5AzaC for 48 hours, cells were allowed to recover for 48 hours. MeDIP revealed no differences in recovery between the control and *SMCHD1* mutant cells (Fig. 2.4D). Taken together, our results indicate that SMCHD1 is not required for either the maintenance or the re-establishment of DNA methylation at the DUX4 promoter in adult myoblasts. It remains possible that *SMCHD1* must be mutated earlier in development to have an effect on DNA methylation at D4Z4, or there may be an additional mechanism/modifier gene involved in FSHD2.

At late myotube stages in DEL mutants, we observed a further reduction in DNA methylation and/or H3K9me3 at the *DUX4* promoter, accompanied by more efficient expression of *DUX4* and target genes (Figs. 2.2D and 2.4E). In conclusion, our results indicate that the degree of loss of heterochromatin caused by FSHD mutations correlates with the degree of activation of the DUX4 network and severity of the disease.

Inhibition of DNA methylation boosts expression of the DUX4fl network in mutant cells

One major difference between our somatic mutant cells and FSHD2 cells that were tested was the level of DNA methylation (Fig. 2.4B). To examine the role of DNA methylation, we treated control and mutant cells with 5AzaC. We observed robust induction of DUX4 target genes in DEL and DEL_SM mutant, but not in control cells (Fig. 2.5A; Supplemental Fig. S2.5B). Upregulation is more prominent in early myotubes than at myoblast stages (Fig. 2.5A). Although DUX4 target genes were significantly upregulated in SM myotubes by 5AzaC, they were induced over 100-fold higher in DEL myotubes (Fig. 2.5B and C). Importantly, in both cases, upregulation was DUX4-dependent (Fig. 2.5D; Supplemental Fig. S2.6A). Because 5AzaC also inhibits RNA methylation, we tested 5AzadC, which only gets incorporated into DNA, and obtained the same results (Supplemental Fig. S2.6B). Moreover, MeDIP indicates that both 5AzaC and 5AzadC caused a comparable reduction in DNA methylation at D4Z4 (Supplemental Fig. S2.6C). We conclude that inhibition of DNA methylation leads to an induction of the DUX4 gene network.

To specifically detect the *DUX4fl* transcript that can be translated into the protein, we split our RNAScope probes to 4ZZ specific to the middle region and 2ZZ specific to the

3'end of the DUX4fl transcript (Fig. 2.5E). Therefore, colocalization of the two probes should reflect the presence of DUX4fl. Using this strategy, we were able to confirm that the prominent nuclear foci of *DUX4* RNA previously detected (37) represents *DUX4fl* transcripts (Fig. 2.5E). 5AzaC treatment indeed increased the colocalized signals of *DUX4fl* in DEL mutant myotubes (Fig. 2.5E, right). Consistently, DUX4 protein expression at the early myotube stage is significantly induced by 5AzaC treatment to a level that is comparable to those at late myotube stages of DEL_SM and patient cells (Fig. 2.5F-H). Collectively, these findings indicate that heterochromatin disruption stimulates target gene induction through upregulation of DUX4fl expression in mutant cells. Moreover, we found that contraction of D4Z4 makes this locus highly sensitive to additional heterochromatin destabilization, emphasizing a potential role for epigenetic modifiers in the disease penetrance and severity in FSHD1.

Coherent feedforward loop of early and late DUX4 target genes

LEUTX is a DUX4 target gene that encodes for a transcription factor critical for zygotic genome activation in early development (46). We observed significant increase of *LEUTX* transcript and protein signals in early mutant myotubes after 5AzaC treatment (Figs. 2.5C and 2.6A, respectively). Interestingly, however, LEUTX induction in myoblasts is very weak (Fig. 2.6A). This is in contrast to the significant induction of another DUX4 target H3.X/Y in myoblasts by 5AzaC (Fig. 2.6A). We found that genes, such as *H3Y1*, *MBD3L2*, *KHDC1L*, are among the top DUX4 target genes activated, and are efficiently expressed and can be stimulated by 5AzaC in the myoblast stage in DEL_SM cells (Figs. 2.2D and 2.6C; Supplemental Fig. S2.7). In contrast, *LEUTX*, *DUXA*, *RFPL1* and *KLF17*, are

expressed only at a low level and fail to get stimulated by 5AzaC in the myoblast stage and are induced much more efficiently after cells were differentiated into myotubes (Fig. 2.6C; Supplemental Fig. S2.7). These results reveal that there are two classes of DUX4 targets: early response genes and differentiation-dependent late target genes.

Histone H3Y1 (H3.Y) is one of the highest induced DUX4 targets in mutant cells (Fig. 2.2D). H3.X/Y was found to be incorporated in the target gene regions and increased their expression (47). Indeed, H3.X/Y binding to *MBD3L2*, *ZSCAN4* and *LEUTX* gene regions was all induced in 5AzaC-treated DEL_SM, but not control myotubes, providing an additional mechanism for increased target gene expression by 5AzaC (Fig. 2.6B). Consistently, shRNA depletion of H3.X/Y effectively blocks *LEUTX* activation in DEL_SM myotubes (Fig. 2.6D). The effect is nearly comparable to DUX4 depletion, emphasizing the strong reliance of DUX4 target gene expression on H3.X/Y (Fig. 2.6D). Conversely, overexpression of H3.X significantly stimulated *LEUTX* expression in DEL_SM cells (Fig. 2.6E). However, consistent with the notion that *LEUTX* is a differentiation-dependent late target gene, induction of LEUTX mRNA and protein is much more robust in myotubes than in myoblasts, despite comparable H3.X/Y overexpression in both cells (Fig. 2.6E and G).

Overexpression of LEUTX also stimulated *H3.X/Y* (as well as *MDB3L2* and *TRIM43*), indicating a positive feedback loop (Fig. 2.6F and G). Indeed, H3.X and H3.Y gene promoters both contain LEUTX (OTX2) binding motifs, raising the possibility that LEUTX directly upregulates *H3.X/Y* genes (Fig. 2.6H). Intriguingly, genome-wide analyses of overexpression and depletion of H3.X/Y and LEUTX by RNA-seq reveal that both of them globally affect DUX4 target gene expression (Figs. 2.6I and J). Taken together, the results indicate that the differentiation-independent early DUX4 target H3.X/Y gets incorporated

and sets the epigenetic stage for upregulation of other DUX4 target genes, including the differentiation-dependent LEUTX. Once expressed, LEUTX further stimulates expression of H3.X/Y as well as other DUX4 target genes to fortify the DUX4-triggered gene network. Taken together, our results suggest that the DUX4 gene network is composed of a coherent feedforward portion with DUX4 working together with H3.X/Y to drive other target genes as well as a positive feedback loop between H3.X/Y and LEUTX (and possibly other TFs) that is subsequently activated and locks in target gene expression (Fig. 2.7B).

2.4 Discussion

In the present study, we investigated the effects of genetic FSHD mutations in healthy isogenic human skeletal myoblast cells. Particularly, we compared effects of D4Z4 contraction or loss of SMCHD1, either alone or combined. We found that, although loss of SMCHD1 by itself has only minor consequences, this mutation strongly synergizes with D4Z4 contraction, resulting in upregulation of the DUX4fl target gene network. These results support the notion of SMCHD1 mutations as modifiers that enhance FSHD1 severity (7, 8). Interestingly, SMCHD1 mutation affected H3K9me3 but not DNA methylation (Fig. 2.7A). Our analyses uncovered two key processes in the establishment of the robust FSHD disease phenotype: (1) disruption of heterochromatin at D4Z4 repeats and (2) coherent feedforward loop of the DUX4 target gene expression (Fig. 2.7B and C). Our mutant cell models closely recapitulate the patient cell phenotype and display sensitivity to disease modifiers. Importantly, the analysis of late myotubes allowed us to separate robust DUX4 target gene activation from cytotoxicity/apoptosis. Our results suggest that transcriptional

activation of the DUX4 target gene network, rather than short-term cytotoxic effects, is relevant for FSHD disease progression.

Our CRISPR mutation using gRNA specific to 4q D4Z4 repeat upstream of the DUX4 transcription start site has led to repeat contraction as initially assessed by PFGE. A similar CRISPR approach was used to remove telomere repeats in human neuroblastoma cells (48). We found that despite some repeat inversion or, CRISPR plasmid insertion to D4Z4 in one set of DEL mutants, the last copy of 4qA D4Z4 remained intact enabling the expression of DUX4fl. Perhaps it is not surprising that repetitive sequences can be recombinogenic, which makes it important to assess repeat alteration by direct sequencing of genomic DNA and transcript RNA using nanopore long-read sequencing.

We previously demonstrated that D4Z4 heterochromatin is marked by H3K9me3, which is required for the recruitment of HP1γ, cohesin and SMCHD1 (24, 25). Reduction of H3K9me3 leads to decrease of SMCHD1 binding, accompanied by increased expression of DUX4 (25). In the currently study, albeit more variable than in DEL mutants, we found that SMCHD1 mutation reduced H3K9me3 and increased DUX4 target gene expression in a modest, but statistically significant manner, suggesting the positive feedback loop between SMCHD1 and H3K9me3 contributing to DUX4 suppression. Interestingly, reduction of H3K9me3 was also observed in D4Z4 transgenic mice crossed with smchd1 mutant mice (49). Furthermore, a recent study showed that restoration of SMCHD1 expression in FSHD2 iPSCs increased the level of H3K9me3 and HP1γ at D4Z4 chromatin (50). SMCHD1 is recruited to and compacts human inactive X chromosome in part through interaction with HP1-bound H3K9me3 chromatin via HP1-binding protein HBiX1 (LRIF1) (51, 52). Suggestively, a homozygous mutation of LRIF1 has also been linked to FSHD2 (53). Thus,

disruption of the observed positive feedback loop between SMCHD1 and H3K9me3 at D4Z4 may be critical for FSHD pathogenesis.

Smchd1 was originally identified to play a role in the maintenance of DNA methylation at CpG islands and inactive X chromosome (27-29). Since FSHD2 cells (with SMCHD1 mutations) tend to exhibit a strong DNA hypomethylation phenotype (21) (and this study), it was speculated that SMCHD1 mutation results in loss of DNA methylation leading to DUX4 upregulation. A recent study showed that Smchd1 binds and antagonizes Tet enzymes (54). Consequently, loss of Smchd1 in mouse ES cells leads to reduced DNA methylation and upregulation of Dux, a functional homolog of human DUX4 (54). Our results, however, suggest that such a DNA methylation-dependent mechanism does not appear to operate in human myocytes. We cannot formally exclude the possibility that there is a narrow developmental window in which haploinsufficiency of SMCHD1 effectively blocks the initial establishment of DNA methylation at D4Z4. Indeed, it was shown that D4Z4 remethylation during reprogramming cannot take place in SMCHD1-mutated FSHD2 iPSCs, suggesting that SMCHD1 is involved in the initial establishment of D4Z4 DNA methylation earlier in development (43). However, since that study were done using FSHD2 iPSCs rather than SMCHD1 knockout iPSCs, possible contributions of additional modifier gene(s) in FSHD2 iPSCs cannot be excluded. Interestingly, SMCHD1 gene correction in FSHD2 iPSCs failed to increase DNA methylation (50). The fact that SMCHD1 mutation did not affect D4Z4 DNA methylation, yet cooperated strongly with D4Z4 contraction in activating DUX4 target gene expression in our mutant cells, suggests that SMCHD1 suppresses DUX4 in a DNA methylation-independent manner. This might involve the SMCHD1-H3K9me3 positive feedback loop discussed above. This is consistent

with a previous report that SMCHD1 can suppress gene expression in a DNA methylation-independent manner (30). Interestingly, we observed that 5AzaC-induced DNA hypomethylation in SMCHD1 mutant cells is not sufficient to recapitulate the robust phenotype of FSHD2 patient cells (with SMCHD1 mutation) used in our study. These observations hint at the existence of additional factors that synergize with SMCHD1 mutations to generate full FSHD2 pathogenesis (Fig. 2.7C).

In addition to DUX4 target genes, D4Z4 disruption combined with loss of SMCHD1 in our mutant cells yielded a gene-expression pattern that recapitulated key features of patient myocytes. These include upregulation of genes related to ECM, immune and stress responses as well as embryonic genes, and downregulation of muscle genes (10, 16, 35, 45, 55). Crucially, DUX4 depletion reversed these changes in gene expression, strongly suggesting that they are DUX4-dependent. In contrast to the almost complete suppression of the DUX4 target gene network, this reversal is only partial, suggesting that expression changes of these other pathways are largely indirect downstream effects. Notably, changes in the expression of genes involved in the regulation of apoptosis (both anti- and pro) are more prominent in myoblasts, and appear to taper off later in myotube differentiation when DUX4 target genes are most highly induced. Our results strongly suggest that activation of the DUX4 target gene network is separate from DUX4-induced cell toxicity. Therefore, further analyses of dynamics and consequences of DUX4 and target gene expression during muscle differentiation will be important.

Our results indicate that DUX4 target genes can be sub-divided into early and late genes based on their differential dependency on myotube differentiation. The highly induced early DUX4 targets, histone variants H3.X and/or H3.Y, were previously shown to

be incorporated into the DUX4 target gene regions and promote their expression (47). We found that H3.X/Y depletion almost completely inhibits the expression of LEUTX, a late target gene. Interestingly, another differentiation-insensitive early DUX4 target, MBD3L, can disrupt the repressive functions of MBD2 or MBD3, and upregulate DUX4 expression as a positive feedback regulator (56, 57). Thus, these early DUX4 targets are epigenetic initiators that set the stage to promote activation of the DUX4 gene network. In contrast, late target genes such as LEUTX, are not induced until cells differentiate into myotubes, suggesting an additional differentiation-coupled mechanism of gene regulation. Nevertheless, overexpression or depletion of H3.X/Y as well as LEUTX globally affects DUX4 target expression. LEUTX may directly bind and control these genes and/or indirectly promote their expression through feedback activation of H3.X/Y (and possibly other TFs), as indicated by our study. Taken together, our results strongly suggest that DUX4 triggers the sequential induction of epigenetic regulators and downstream transcription factors to ensure a coherent feed forward effect on the DUX4 gene network (58).

Using CRISPR engineering, we have created SMCHD1 and/or D4Z4 mutant human skeletal myoblast lines and examined the mutation effects during myocyte differentiation. Our results highlight the fact that FSHD is a heterochromatin abnormality disorder, in which stabilization of otherwise variegated DUX4fl expression from contracted D4Z4 allele by chromatin modifier(s) is a critical driver of FSHD pathogenesis. Furthermore, our observations revealed a hierarchy within the DUX4 target genes, highlighting a coherent feed forward mechanism of the DUX4 target gene network activation involving early and late genes, triggered by the expression of the DUX4fl from the last D4Z4 repeat. Our results

also suggest that in adult myocytes SMCHD1 functions in the context of H3K9me3 rather than regulation of DNA methylation, raising the possibility of additional modifiers in FSHD2. The mutant cells described in the present study will be a valuable resource for further investigation of the molecular mechanism, and serve as a possible platform for therapy development.

2.5 Materials and Methods

Generation of SMCHD1 knockout mutants in immortalized permissive control myoblast with CRISPR-Cas9

The CRISPR-SMCHD1-1 and 2 constructs used in this study for induction of double-strand breaks (DSB) at SMCHD1 gene were designed by CRISPRdirect (https://crispr.dbcls.jp/) (Supplemental Figure S2.2A). ~3x105 Immortalized human control myoblasts were seeded in a 35mm cell culture dish. One day later, the cells were transfected with both 1 μg of CRISPR-SMCHD1-1 and 1 μg of CRISPR-SMCHD1-2 or CRISPR-SMCHD1-1 alone together with 0.5 μg of a puromycin-resistance plasmid using Lipofectamine 3000. The media was changed after 4 hours. The next day the media was replaced with fresh media containing 2 μg/ml puromycin (Sigma-Aldrich) for 3 days. Single cell clones were isolated by FACS sorting into 96 well plates. Ten to 14 days later, genomic DNA of the single cell clones with good proliferation was extracted using a QuickExtract DNA extraction kit (Epicentre). The exon 23-24 region was amplified by a pair of PCR primer (SMCHD1_PCR) to check for genomic deletions on a 1.5% agarose gel. The deletion mutants were confirmed by Sanger sequencing and western blot. SMCHD1 knock out mutants were identified by western blot and pooled amplicon sequencing. Using

custom barcoded primers (SMCHD1_seqPCR), pooled amplicons from multiple individuals were sequenced at Genewiz to determine the genomic sequence at the gRNA target site of each cell line. The off-target loci with the highest prediction scores were amplified by a PCR primer pair (SMCHD1_off_target_PCR) and sequenced by Sanger sequencing. For the SMCHD1 mutant cell lines used in this paper, the corresponding off-target sequences (5'-TTTTCAATTTCAGTCAACGA-3', chr9:+72135246) were not changed.

Generation of D4Z4 contraction mutants with CRISPR-Cas9

The CRISPR-Cas9 system was used to delete D4Z4 repeat units on 4qA in permissive control myoblasts. Guide RNAs (gRNAs) target a D4Z4 "1-kb" subregion sequence, which excludes the regions that are repeated elsewhere in the genome (59) as well as the DUX4 gene/promoter region to avoid to induce DUX4 mutation (Supplemental Figure S2B). Based on this sequence, gRNAs were designed for 4q D4Z4 using the CRISPR Design Tool (https://zlab.bio/guide-design-resources) with low predicted off-target effects. Two rounds of CRISPR-Cas9 induced D4Z4 repeat array contraction was performed to obtain D4Z4 contraction mutants with satisfactory repeat number. Cas9 (41815, Addgene plasmid) gRNA-D4Z4-1/ 2 (in gRNA expression vector pH082 pU6-gRNA2.0-GFP) and a puromycin-resistance plasmid were cotransfected into parental myoblast as indicated in Supplemental Figure S2.2B. One mutant with 10 units of 4qA D4Z4 repeat was used as parental cells for the 2nd round of D4Z4 deletion mutant generation. Alternatively, the Alt-R CRISPR-Cas9 genome editing system (IDT) was used for the recombinant Cas9 protein and gRNA delivery for single and double mutations of D4Z4 and SMCHD1 (DEL5, DEL7,

DEL8, DEL9 and DEL9_SM, respectively). CRISPR/Cas9 (1081060, Integrated DNA technologies (IDT))/tracrRNA (1073190, IDT)/crRNA(IDT) RNP were delivered to the myoblasts using CRISPRMAX Cas9 Transfection Reagent (CMAX00003, Invitrogen). Transfections were performed as described in the IDT protocol for "Alt-R CRISPR/Cas9 System". The differentiation efficiency and DUX4 target gene MBD3L2 expression of the single colony cell lines were tested. Based on the results, several cell lines were subjected to PFGE and blot hybridization to confirm the size of D4Z4 regions at 4q and 10q as well as nanopore genomic sequencing (see below).

Genomic and RNA nanopore long-read sequencing

Genomic Nanopore libraries constructed from genomic DNA using Cas9 Sequencing Kit (SQK-CS9109), Nuclease-free duplex buffer (IDT Cat # 11-01- 03-01), and the following Alt-R CRISPR reagent from IDT: tracrRNA (1073190, IDT) resuspended at 100 μM in TE pH 7.5, Cas9 nuclease V3 (1081060, IDT), 3 different S. pyogenes Cas9 Alt-RTM crRNAs (two upstream target sites 5'-CCTATTAAACGTCACGGACA-3' and 5'-GATACCGACAGCAATAGTCC-3' and one downstream target site '5-AAATCTTCTATAGGATCCAC-3') resuspended at 100 μM in TE pH 7.5. The Long Fragment Buffer (LFB) from the Cas9 Sequencing Kit (SQK-CS9109) was used during the wash steps. Libraries were loaded on R9.4.1 Flow Cells (FLO-MIN106D) and sequenced on MinION Mk1B instrument using the MinKNOW software. Oxford Nanopore's base calling software, Guppy version 6.0.1+652ffd1, was run in super accurate (sup) mode with the dna_r9.4.1_450bps_sup.cfg configuration file.

RNA-seq Nanopore libraries were constructed from 200 fmol Illumina libraries using Ligation Sequencing Kit (SQK-LSK110) and NEBNext® Companion Module for Oxford Nanopore Technologies® Ligation Sequencing (E7180S). The Short Fragment Buffer (SFB) from the Ligation Sequencing Kit (SQK-LSK110) was used during the wash steps. 50 fmol sample libraries were loaded on R9.4.1 Flow Cells (FLO-MIN106D) and sequenced on MinION Mk1B instrument using the MinKNOW software. Oxford Nanopore's base calling software, Guppy version 6.0.1+652ffd1, was run in super accurate (sup) mode with the dna_r9.4.1_450bps_sup.cfg configuration file. Adapters were trimmed from reads with the Porechop software package by adding a custom adapter sequence that includes both the illumina primer and nanopore adapter (top: 5`-AATGTACTTCGTTCAGTTACGTATTGCTAAGCAGTGGTATCAACGCAGAGTAC-3` and bottom: 5`-GTACTCTGCGTTGATACCACTGCTTAGCAATACGT-3`). Reversed reads were flipped using a custom script.

Cell culture and differentiation

Immortalized control, FSHD1, FSHD2 and control-derived mutant skeletal myoblast cells were grown in high glucose DMEM (Gibco) supplemented with 20% FBS (Omega Scientific, Inc.), 1% Pen-Strep (Gibco), and 2% Ultrasor G (Crescent Chemical Co.). Immortalization and single cell clone isolation of primary FSHD1 myoblasts were performed as previously described for Control and FSHD2 myoblasts (60). Upon reaching 80% confluence, myoblast differentiation was induced by using high glucose DMEM medium supplemented with 2% FBS and ITS supplement (insulin 0.1%, 0.000067%

sodium selenite, 0.055% transferrin; Invitrogen). Fresh differentiation media was changed every day.

Antibodies

Immunofluorescence (IF) or Western blot (WB) is performed using antibodies specific for SMCHD1 (NBP1-49968, Novus Bio.), DUX4 (NBP2-12886, Novus Bio.), LEUTX (PA5-59595, Thermofisher), Actin (A4700, Sigma), Histone H3 (ab18521, Abcam), and Histone H3.X/Y (MABE243I, Sigma).

Immunofluorescent staining

Staining was performed as previously described (61). Briefly, cells grown on coverslips were fixed in 4% paraformaldehyde for 10 min at room temperature, permeabilized with 0.5% Triton X-100 in PBS, and blocked in blocking buffer (0.02% saponin, 0.05% NaN3, 1% BSA, 4% horse serum and 0.1% gelatin in PBS) for 15 min at 37°C. The coverslips were incubated overnight with primary antibodies at 4 °C followed by three PBS washes, then incubated with fluorescent secondary antibody for 30 min at 37°C, washed with PBS 3 times, counter-stained with DAPI, and mounted with Prolong Diamond Antifade Mountant. Images were acquired with a Zeiss LSM510 confocal laser microscope.

Western blotting

Cells were lysed in 2X Laemlli Buffer with 4% beta-mercaptoethanol, sonicated, boiled, and separated by 4%–20% TEO-Tricine gel (Abcam). Then the samples were transferred to nitrocellulose membranes, blocked with Pierce Protein-Free T20 (PBS) Blocking Buffer (Thermo Fisher Scientific), and blotted with the desired antibodies. Horseradish peroxidase-conjugated anti-mouse-IgG (Promega), anti-rabbit-IgG (Promega), or anti-rat-IgG (Abcam) were used as secondary antibodies. Immunoblots were developed with SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific). Images were acquired using the Image Analyzer (LAS-4000, Fujifilm).

RNA isolation and quantitative real-time RT-PCR (RT-qPCR)

RNA was extracted using RNeasy Plus Mini kit (Qiagen, Cat No. 74134), and complementary DNA (cDNA) was made using 500ng of total RNA with SuperScript IV VILO Master (Thermo Fisher Scientific, Cat No. 11756050) following the manufacturer's instructions. qPCR was performed by using AzuraView GreenFast qPCR Blue Mix LR (Azura Genomics Inc., Cat No. AZ-2320). The genes and their corresponding PCR primers are listed in Supplemental Table S2.4. For some genes, the expression was detected by probe qPCR, which was performed by using TaqMan Fast Advanced Master Mix (Thermo Fisher Cat No. 4444557). The commercially available TaqMan Gene Expression Assay probes (Thermo Fisher Scientific) were listed in Supplemental Table S2.5.

RNAScope in situ hybridization

RNAscope was performed using RNAscope Multiplex Fluorescent Reagent Kit v2 (Advanced Cell Diagnostics, Inc., Cat. No. 323100) according to the manufacturer's protocol as previously described (37). The following RNAScope probes (Advanced Cell Diagnostics, Inc.) were used: LEUTX probe set (Hs-LEUTX-C2, Cat. No. 547251-C2), 6ZZ DUX4fl probe set (HS-DUX4-O6-C1, Cat. No. 546151), 4ZZ DUX4fl probe set (Hs-DUX4-O7-C2, Cat. No. 1089191-C2), and 2ZZ DUX4fl probe set (Hs-DUX4-O8-C3, Cat. No. 1089201-C3). The 4ZZ and 2ZZ DUX4fl probes target 701-1388 and1481-1697 of DUX4fl mRNA (NM_001306068.2) respectively (Figure 2.5E).

5-Azacytidine (5AzaC) and 5-Aza-2'-deoxycytidine (5AzadC) treatment

Six μM of 5AzaC (Sigma-Aldrich, A2385) was added to myoblasts at ~70% confluency for 48h. With or without differentiation, the cells were harvested or fixed at indicated day after the drug treatment for subsequent RNA-seq, ChIP-qPCR, RT-qPCR, MeDIP and IFA analyses. Alternatively, myoblasts were incubated with 6 μM of 5AzadC (Sigma, A3656) for 24h, followed by 2 days release. To assess the SMCHD1 depletion effect, cells were treated with 5 μM of 5AzaC for 48 hours. For RNA-seq experiment, myoblasts were allowed to grow for additional 48 hours before collection or before changing to differentiating media. For SMCHD1 mutant maintenance experiment, cells were allowed to recover for 48 hours before collection.

ShRNA depletion or overexpression of proteins using lentiviral systems

The shRNA plasmids for SMCHD1 (5'-TTATTCGAGTGCAACTAATTT-3', TRCN0000253777), DUX4 (5'-AGATTTGGTTTCAGAATGAGA-3', TRCN0000421072) and a control shRNA (shCTRL, 5'-CAACAAGATGAAGAGCACCAA-3', SHC002), were obtained from the Sigma Mission library. The shRNA pLVshH3.X/Y (5'- GCGGGAAATCAGAAAGTAC-3', the siH3.X/Y targeting sequence in previous paper (47)), pLVshDUX4 (5'-GGCAAACCTGGATTAGAGTT-3' (18)), and corresponding pLVshControl (5'-CCTAAGGTTAAGTCGCCCTCG-3') were synthesized and cloned in pLV(shRNA)-Puro-U6 vector by VectorBuilder. To construct the H3.X and LEUTX overexpression lentiviral plasmids pLVX_H3.X and pLVX_LEUTX, human H3.X and LEUTX ORF sequences were amplified from FSHD1 myotube cDNA using primer pairs H3.X_PCR and LEUTX_PCR respectively, and cloned into pLVX Lentiviral vector (Addgene, plasmid #135182). An empty pLVX vector was used as a negative control. Lentivirus packaging, transduction and puromycin selection were performed as previously described with slight modifications (37). Briefly myoblasts were infected twice at 48 hour and 24 hours prior to differentiation. The differentiated cells were harvested at days 3-7 when 60-80% cells fused to form myotubes.

Detection the ratio of 4qA- and 10q-specific nucleotide polymorphisms (SNPs) using Amplicon sequencing

Input and MeDIP DNA were amplified by PCR using Phusion DNA Polymerase in two steps: The first PCR primers (D4Z4_seqPCR) are derived from the 4q/10q specific Q-PCR primers (24) with adapters for the 2nd PCR primers attachment. The 2nd PCR and

Amplicon sequencing were performed following the previous paper (62) with modifications. The second PCR primer pair (Illumina_seqPCR) was used to attach Illumina adaptors and to barcode samples. Amplification was carried out with 18 cycles for the first PCR and 24 cycles for the second PCR. Resulting amplicons from the second PCR were gel extracted, quantified, mixed and sequenced using NovaSeq6000 (Illumina). The ratio of 4qA/B - and 10q-derived D4Z4 sequences was calculated based on the reads number of their specific SNPs.

Pulsed-field gel electrophoresis (PFGE) and Southern blotting for 4q/10q D4Z4 repeat array length analysis

PFGE and southern blotting were performed as described (59) with slight modifications. Briefly, suspended myoblasts were mixed with melted 1% UltraPure Low Melting Point Agarose (Bio-Rad) at 37°C to form plugs, each containing ~1.5 x10^6 cells. The plugs were digested with pronase (Sigma), rinsed, and treated with enzymes (EcoRI plus HindIII or EcoRI plus BlnI, Roche). The digested DNA in plugs was subjected to PFGE. Electrophoresis was done in CHEF-DR III system (Bio-Rad) at 6 V/cm and 15°C for 13 h, with the switch time increasing linearly from 1 to 6 sec. The gel was washed in 0.25 M HCl for 30 minutes to depurinate DNA fragments, rinsed with H_2O, washed in denaturation buffer (0.6 M NaOH, 0.4 M NaCl) for 30 minutes, and transferred in that solution to a Biodyne B Membrane (KPL). The membrane was neutralized, cross-linked by UV irradiation, and was probed with the 1-kb 4q/10q specific probe described in the paper (59). Southern blots were visualized using Typhoon scanner (GE Healthcare).

ChIP-qPCR Analysis

Around 3 x 10^6 cells were crosslinked with 1% formaldehyde for 10 minutes and quenched with 0.125 M glycine for 5 minutes. Cells were washed with ice-cold PBS and harvested in cell lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP40, protease inhibitors). Nuclear extract was collected in RIPA buffer (1 % NP-40, 0.5% Sodium Deoxycholate, 0.1% SDS, protease inhibitors). Nuclei were sonicated for 20 cycles (30 seconds on and off) using the Bioruptor (Diagnode) to obtain fragment length around 150-500 bp. Chromatin extract was quickly spun down to remove cell debris. Around 20 µg or 10 µg of chromatin were incubated overnight at 4°C with 10 µg H3.X/Y or 2.5 µg H3K9me3 antibody respectively (Active Motif-61161, Abcam-ab8898). Chromatin-antibody extracted with incubated with protein G Dynal beads (Thermo Fisher Scientific, 10003D) for an hour. The mixture was washed with Lithium Chloride 5 times and TE buffer once on Ice. Chromatin was eluted with 0.1 M NaHCO3 and 1% SDS. Proteinase K was added to chromatin and the mixture was proceeded to reverse-crosslinking for 2 hours at 55°C. DNA was purified with QIAquick PCR Purification Kit (Qiagen). ChIP DNA was quantified by qPCR with specific primers.

MeDIP

Cells were washed with PBS and harvested in SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris HCl, pH 8.1). Nuclear extract was sonicated with 3 cycles (30 seconds on and off). Chromatin extract was quickly spun down and supernatant was collected.

Chromatin was added with Proteinase K and incubated for 2 hours at 55°C. DNA was purified with QIAquick PCR Purification Kit (Qiagen). Between 0.5-1 µg of chromatin was proceeded to MeDIP (Methylated DNA Immunoprecipitation) according to the manufacture protocol (EpiMark® Methylated DNA Enrichment Kit, New England Biolabs). MeDIP DNA was quantified by qPCR with specific primers and sequencing.

RNA-seq and data processing

Total RNA was extracted by using the RNeasy kit (QIAGEN). Between 12-50ng of RNA was converted to cDNA using the Smart-Seq 2 protocol (38). DNA libraries were constructed using Nextera DNA Flex Library Prep Kit (Illumina). DNA samples were sequenced on the Illumina NextSeq500 platform using paired-end 43 bp mode with around 15 million reads per sample. RNA-seq raw reads were aligned with STAR (version 2.5.1b) using human genome reference hg38. Alignment default parameters were applied except with a maximum of 10 mismatches per pair, a ratio of mismatches to read length of 0.07, and a maximum of 10 multiple alignments. Read count was performed using RSEM (version 1.3) by defaults with gene annotations from GENCODE v28, and raw read counts were extracted for downstream analysis. Genes were filtered based on raw counts with at least 2 counts in at least 2 samples. Raw count was normalized by TMM in EdgeR and then converted to TPM (transcript per millions). Differential genes were calculated using the cut off P-values =< 0.01 and Log2 Fold Change >=1.5.

Statistical Analyses

Microsoft Excel software was used to perform statistical analyses on data from three independent experiments. Statistical comparisons were made using the unpaired Student's t-test and Wilcoxon's t-test. Statistical significance between two samples was determined by a p value of less than 0.05. Error bars shown mean ± SD.

2.6 Figures

Figure 2.1

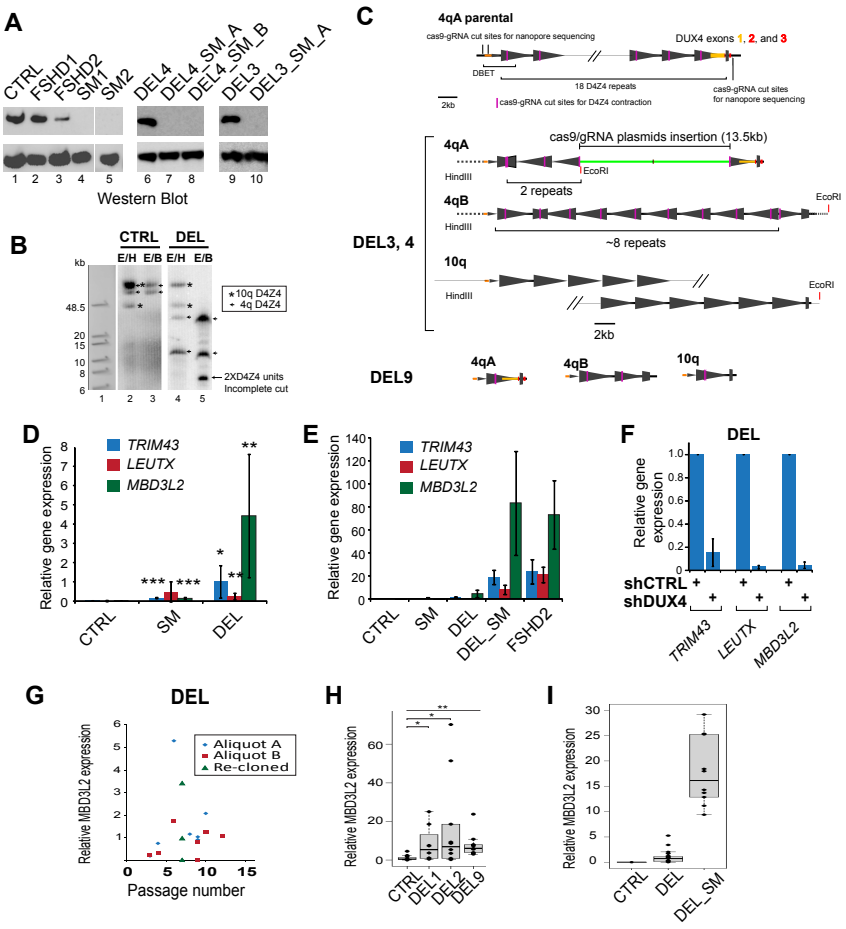


Figure 2.1. Generation of SMCHD1 and/or D4Z4 mutant cells from health permissive skeletal myoblast

A. Western blot analysis the SMCHD1 protein expression in the cell lines used in the study. Lysates of immortalized control and FSHD1 and FSHD2 patient myoblasts, SMCHD1 mutants (SM), D4Z4 deletion mutants (DEL) and double mutants (DEL_SM) were subjected to western blot analysis using antibody specific for SMCHD1. β-Tubulin serves as a loading control.

B. Determination of the 4q and 10q D4Z4 repeat number. Examples of control and DEL mutant clones are shown. Genomic DNA was digested with EcoRI/HindIII (E/H) or EcoRI/BlnI (E/B) and subjected to PFGE. They were then blot-hybridized with the 4q/10q specific "1-kb" D4Z4 probe. E/H digestion leaves intact two 4q and two 10q D4Z4 arrays, while BlnI in an E/B only cleaves 10qD4Z4 repeat units. Size markers (in kb) are shown on the left. Arrowheads and stars indicate 4q and 10q D4Z4, respectively. The arrow indicates a band around 6.6kb, which should be 2 D4Z4 repeat units caused by incomplete digestion. The two 4q D4Z4 repeat arrays are contracted, while the 10q D4Z4 bands size show no change.

C. D4Z4 gRNA targeting resulted in repeat contraction and recombination, leaving the last repeat with the *DUX4* gene intact at the 4qA allele in DEL mutant cells. Top: schematic diagram of D4Z4 array in 4qA allele of parental cell with gRNA target sites for D4Z4 deletion (purple bars) as well as crRNA target sites designed for nanopore sequencing were shown at the top panel. D4Z4 cluster in 4qA, 4qB and 10q alleles of DEL3 were shown below. 10q D4Z4 sequences were confirmed by SNP analysis. The large triangle represented a 3.3kb D4Z4 unit and its orientation. The small and partial triangle represented partial D4Z4 units and their orientation. The endonucleases (EcoRI/HindIII) cut sites, which generated the fragments detected in PFGE, are indicated.

D. Control, SMCHD1 mutant SM1 and D4Z4 contraction mutant DEL3 myoblasts were differentiated and analyzed for DUX4 target (*TRIM43, LEUTX* and *MBD3L2*) RNA expression levels. Data are expressed as relative expression (mean with standard deviation). The gene expression over *GAPDH* was normalized to the *TRIM43* value of DEL, which was set to be 1. * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$ vs. control.

E. Corresponding data from double mutant (DEL_SM) and FSHD2 were added to (D) for comparison.

F. DUX4 depletion by lentiviral shRNA abolished activation of DUX4 target genes (*TRIM43*, *LEUTX* and *MBD3L2*) compared to control shRNA (shCTRL). Cells were harvested at 4 days of myotube differentiation. DUX4 target expression levels were determined by RT-qPCR. Y-axis is relative expression (mean with standard deviation) with the expression in shCTRL-transduced samples as one. ** $P < 0.01$ and *** $P < 0.001$.

G. Stochasticity of target gene expression in DEL3. Two aliquots of the same clone at different passage numbers and replicates of a re-cloned cell line were differentiated for 4 days and analyzed for *MBD3L2* expression by RT-qPCR.

H. Variegation is also observed in multiple DEL mutant clones, including DEL9 as in (**G)**.

I. Comparison of *MBD3L2* expression level in early myotube of Control (n=11), single mutant DEL (n=23), and double mutant DEL_SM (n=10). The dots on each boxplot represent the individual data in each repeat.

Figure 2.2

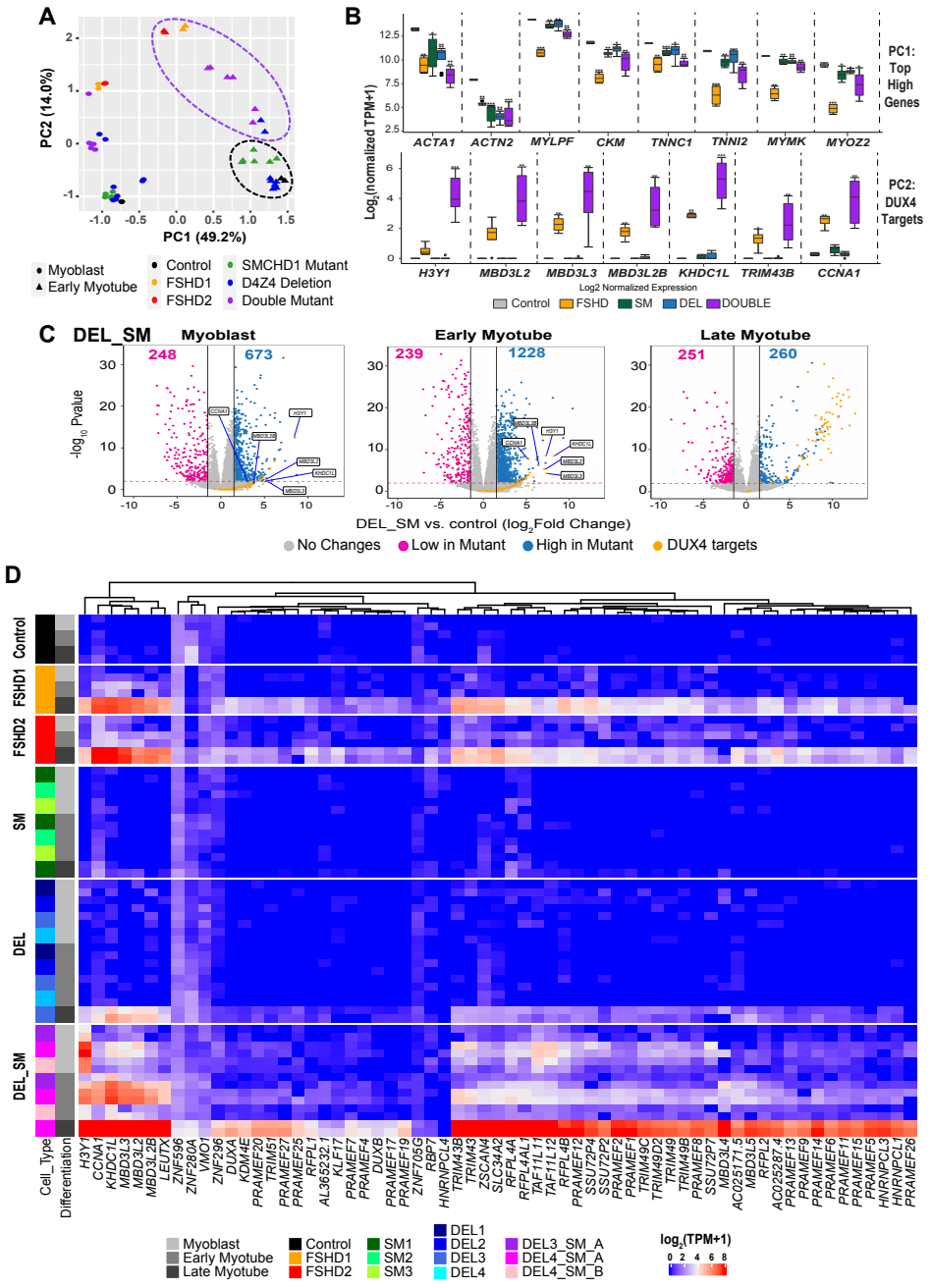

A
PC2 (14.0%)
PC1 (49.2%)
Myoblast
Early Myotube
Control
FSHD1
FSHD2
SMCHD1 Mutant
D4Z4 Deletion
Double Mutant
B
Log2(normalized TPM+1)
ACTA1
ACTN2
MYLPF
CKM
TNNC1
TNNI2
MYMK
MYOZ2
PC1: Top High Genes
H3Y1
MBD3L2
MBD3L3
MBD3L2B
KHDC1L
TRIM43B
CCNA1
PC2: DUX4 Targets
Log2 Normalized Expression
Control
FSHD
SM
DEL
DOUBLE
C
DEL_SM
Myoblast
Early Myotube
Late Myotube
248
673
239
1228
251
260
-log10 Pvalue
DEL_SM vs. control (log2Fold Change)
No Changes
Low in Mutant
High in Mutant
DUX4 targets
D
Control
FSHD1
FSHD2
SM
DEL
DEL_SM
Cell_Type
Differentiation
Myoblast
Early Myotube
Late Myotube
Control
FSHD1
FSHD2
SM1
SM2
SM3
DEL1
DEL2
DEL3
DEL4
DEL3_SM_A
DEL4_SM_A
DEL4_SM_B
log2(TPM+1)

Figure 2.2. Double mutants closely recapitulate patient cells

A. PCA analysis of myoblasts and early myotubes across all the cell types and clones. Top genes for each component are included in the Supplemental Table S2. Differentiation days are indicated by shapes and cell types are indicated by colors according to the label legend.

B. Expression comparison of selected genes from PC1 and PC2 from **(A)**. Top: box plots of the selected top high genes expression of PC1 in control, FSHD patients and 3 types of mutant early myotubes. Bottom: box plots of 7 DUX4 targets expression from the top 500 high genes of PC2. Expression values are in $\log_2$ (normalized TPM +1). Significant values were calculated by Wilcoxon t-test (**** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$).

C. Volcano plots of significance ($\log_{10}$ *P*-value) and $\log_2$ fold change of double mutants (DEL_SM) compared to control at myoblast, early and late myotube stages. Significantly upregulated (blue) and downregulated (pink) genes and DUX4 targets (orange) are shown. DUX4 target genes in PC2 **(B)** are indicated in myoblasts and early myotubes.

D. Hierarchical heatmap of DUX4 target gene expression. A total of 63 target genes were selected based on previous studies (38, 42). Expression values are in normalized TPM and log transformed. Grey shades indicate differentiation and colors indicate cell types.

Figure 2.3

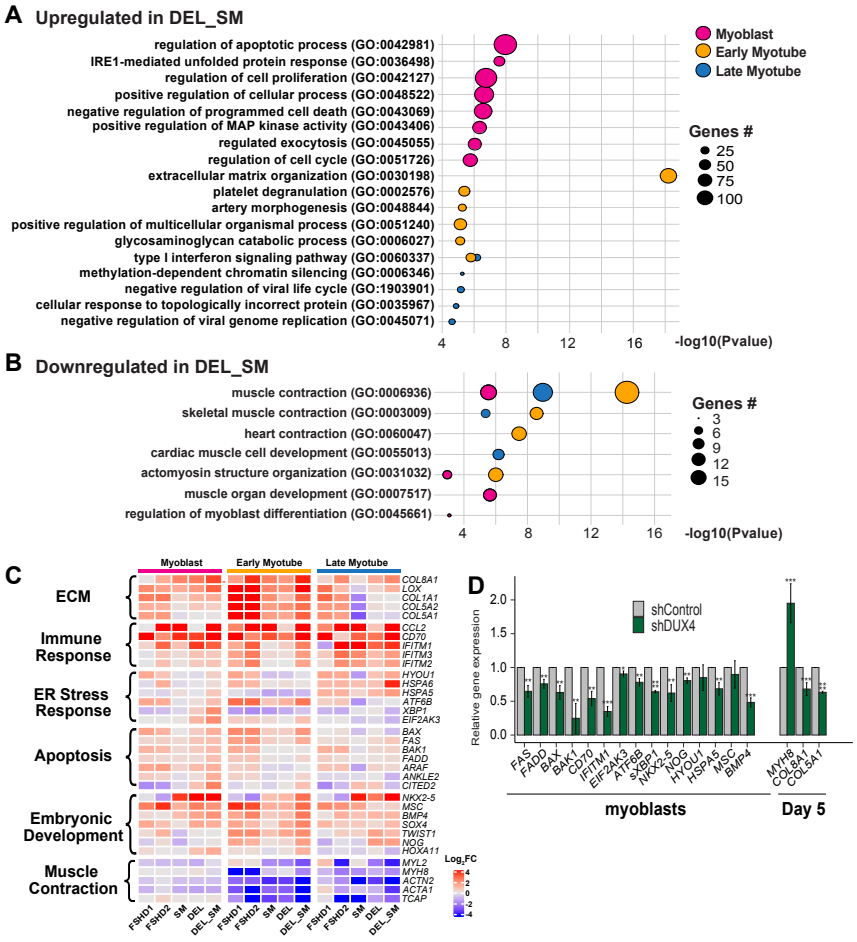

A
Upregulated in DEL_SM
regulation of apoptotic process (GO:0042981)
IRE1-mediated unfolded protein response (GO:0036498)
regulation of cell proliferation (GO:0042127)
positive regulation of cellular process (GO:0048522)
negative regulation of programmed cell death (GO:0043069)
positive regulation of MAP kinase activity (GO:0043406)
regulated exocytosis (GO:0045055)
regulation of cell cycle (GO:0051726)
extracellular matrix organization (GO:0030198)
platelet degranulation (GO:0002576)
artery morphogenesis (GO:0048844)
positive regulation of multicellular organismal process (GO:0051240)
glycosaminoglycan catabolic process (GO:0006027)
type I interferon signaling pathway (GO:0060337)
methylation-dependent chromatin silencing (GO:0006346)
negative regulation of viral life cycle (GO:1903901)
cellular response to topologically incorrect protein (GO:0035967)
negative regulation of viral genome replication (GO:0045071)
4
8
12
16
-log10(Pvalue)
Myoblast
Early Myotube
Late Myotube
Genes #
25
50
75
100
B
Downregulated in DEL_SM
muscle contraction (GO:0006936)
skeletal muscle contraction (GO:0003009)
heart contraction (GO:0060047)
cardiac muscle cell development (GO:0055013)
actomyosin structure organization (GO:0031032)
muscle organ development (GO:0007517)
regulation of myoblast differentiation (GO:0045661)
4
8
12
16
-log10(Pvalue)
Genes #
3
6
9
12
15
C
Myoblast
Early Myotube
Late Myotube
ECM
COL8A1
LOX
COL1A1
COL5A2
COL5A1
Immune Response
COL2
CD70
IFITM1
IFITM3
IFITM2
ER Stress Response
HYOU1
HSPA6
HSPA5
ATF6B
XBP1
EIF2AK3
Apoptosis
BAX
FAS
BAK1
FADD
ARAF
ANKLE2
CITED2
Embryonic Development
NKX2-5
MSC
BMP4
SOX4
TWIST1
NOG
HOXA11
Muscle Contraction
MYL2
MYH8
ACTN2
ACTA1
TCAP
FSHD1
FSHD2
SM
DEL
DEL_SM
Log2FC
4
2
0
-2
-4
D
shControl
shDUX4
Relative gene expression
2.0
1.5
1.0
0.5
0.0
FAS
FADD
BAX
BAK1
CD70
IFITM1
EIF2AK3
ATF6B
sXBP1
NKX2-5
NOG
HYOU1
HSPA6
MSC
BMP4
MYH8
COL8A1
COL5A1
myoblasts
Day 5

Figure 2.3. Ontology analyses of common and distinct gene expression in patient and double mutant cells

A. The bubble plot shows gene ontology enrichment analysis of upregulated genes in the double mutants in myoblasts (pink), early (yellow) and late (blue) myotubes. The plot shows the selected top terms for each differentiation stage. X-axis displays $\log_{10}$ *P*-value and bubble size indicates number of genes in each term as indicated.

B. Similar to **(A)**, bubble plot for downregulated genes.

C. Heatmap of $\log_2$ fold change expression for the selective genes in FSHD patients and mutants and their related pathway.

D. The expression level of selected genes from Figure 3C in double mutant myoblast (DEL4_SM_A) transduced with lentivirus carrying shControl or shDUX4. Real-time RT-qPCRs were performed for three biological replicates for each sample. Data are presented as mean ± SD; **p<0.01, ***p<0.001, by one-tailed student's t-test. Results presented as fold difference compared to shControl sample.

Figure 2.4

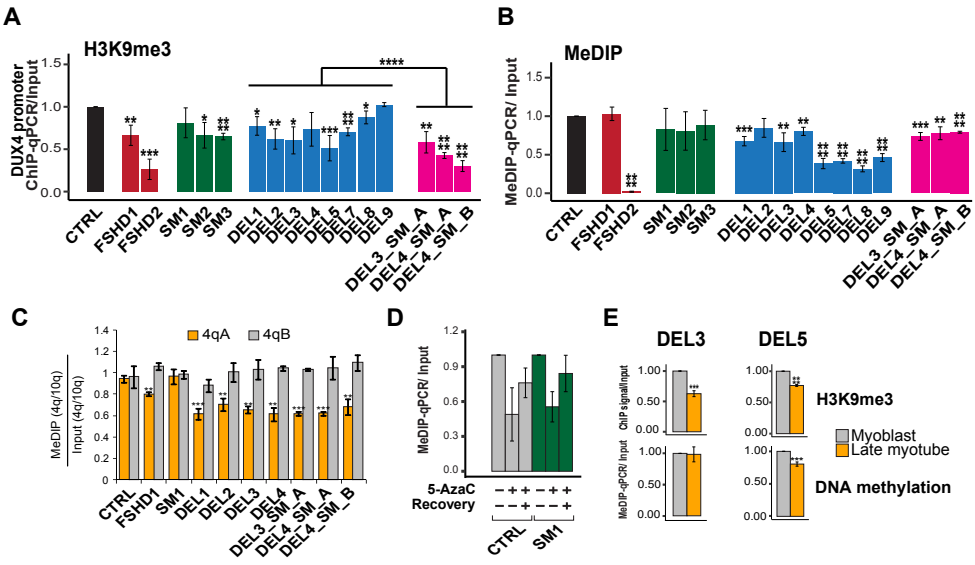

A
H3K9me3
DUX4 promoter ChIP-qPCR/Input
B
MeDIP
MeDIP-qPCR/ Input
CTRL
FSHD1
FSHD2
SM1
SM2
SM3
DEL1
DEL2
DEL3
DEL4
DEL5
DEL7
DEL8
DEL9
DEL3_SM_A
DEL4_SM_A
DEL4_SM_B
C
4qA
4qB
MeDIP (4q/10q)
Input (4q/10q)
D
MeDIP-qPCR/ Input
5-AzaC
Recovery
E
DEL3
DEL5
ChIP signal/Input
MeDIP-qPCR/ Input
H3K9me3
Myoblast
Late myotube
DNA methylation

Figure 2.4. Heterochromatin changes in mutant cells.

A. H3K9me3 ChIP-qPCR analysis of the *DUX4* promoter region in FSHD1, FSHD2 and mutant myoblasts. Reduction of H3K9me3 is enhanced in double mutant cells. For both **(A)** and **(B)**, signals were normalized to input. Significant comparisons to the control are shown with the asterisks calculated by student's t-test (**** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$).
B. DNA methylation from MeDIP analysis. No reduction of DNA methylation was observed in SMCHD1 only and no additional effect in double mutant myoblasts.

C. Comparison of DNA methylation levels at 4qA, 4qB, and 10q D4Z4 regions among the control, FSHD, and mutant cells. The MeDIP and input samples from **(B)** were amplified by using 4q/10q-D4Z4 specific PCR primers. The PCR products were sequenced and the 4qA, 4qB, and 10q D4Z4 specific sequence reads were analyzed. The 4qA(orange)/10q and 4qB(grey)/10q ratios of MeDIP were normalized with that of input. The data indicated relatively lower methylation at 4qA (but not 4qB) D4Z4 regions of FSHD1, D4Z4 deletion mutants and double mutants. *P*-values for significant differences versus the control sample are shown.

D. The effect of 5AzaC treatment on SMCHD1 mutants on DNA methylation. The control or SM1 cells were treated with 5AzaC for 24 hours and allowed for 48 hours of recovery before harvesting for MeDIP-qPCR. No significant differences were observed.

E. H3K9me3 ChIP-qPCR and MeDIP analysis at the DUX4 promoter region were performed to compare between Day 0 and Day 14 of DEL3 or Day 12 of DEL5 mutants. (**** $P < 0.0001$, *** $P < 0.001$).

Figure 2.5

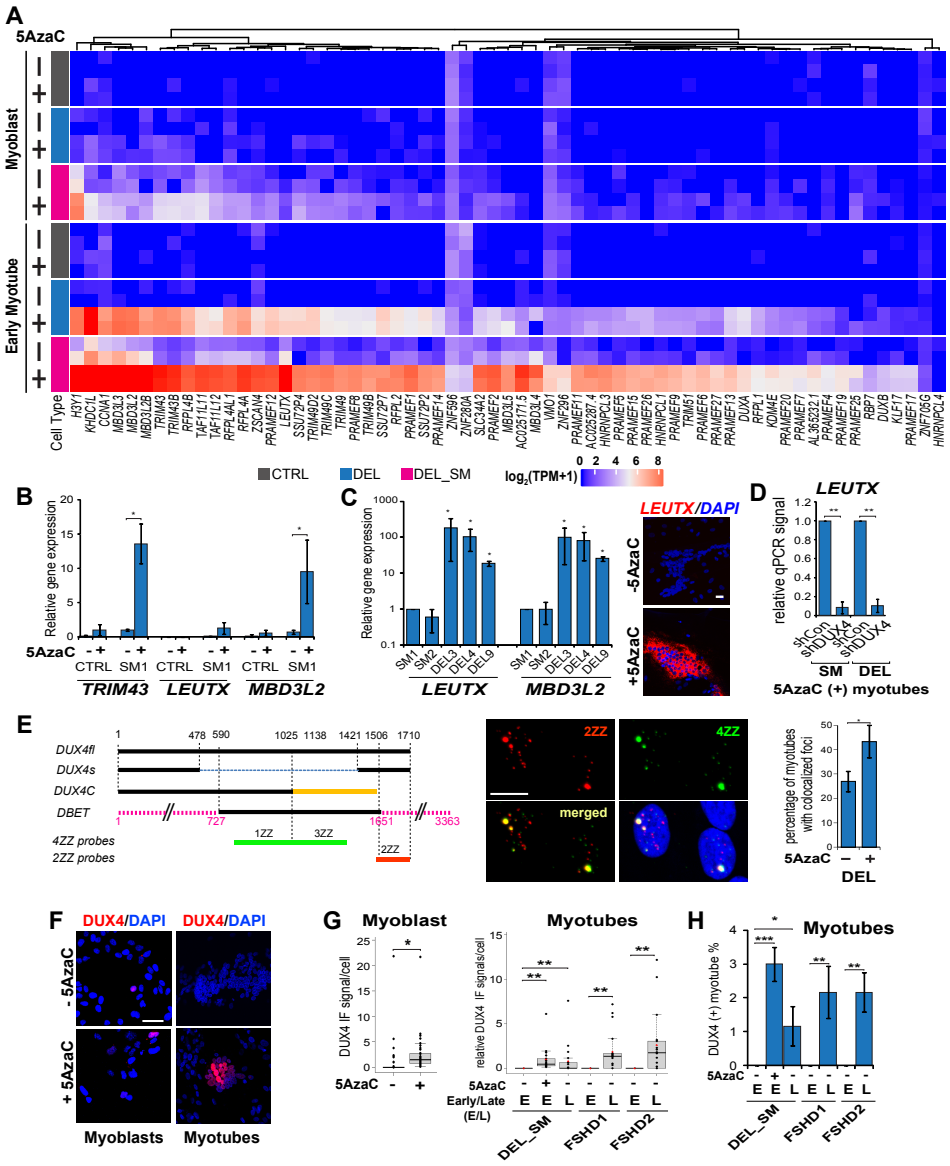

A
5AzaC
Myoblast
Early Myotube
Cell Type
CTRL
DEL
DEL_SM
log2(TPM+1)
B
Relative gene expression
5AzaC
CTRL SM1 CTRL SM1 CTRL SM1
TRIM43 LEUTX MBD3L2
C
Relative gene expression
SM1 SM2 DEL3 DEL4 DEL9
LEUTX MBD3L2
LEUTX/DAPI
-5AzaC
+5AzaC
D
LEUTX
relative qPCR signal
shCon shDUX4 shCon shDUX4
SM DEL
5AzaC (+) myotubes
E
DUX4fl
DUX4s
DUX4C
DBET
4ZZ probes
2ZZ probes
2ZZ
4ZZ
merged
percentage of myotubes with colocalized foci
5AzaC
DEL
F
DUX4/DAPI
DUX4/DAPI
-5AzaC
+5AzaC
Myoblasts
Myotubes
G
Myoblast
Myotubes
DUX4 IF signal/cell
relative DUX4 IF signal/cell
5AzaC
Early/Late (E/L)
DEL_SM
FSHD1
FSHD2
H
Myotubes
DUX4 (+) myotube %
5AzaC
DEL_SM
FSHD1
FSHD2

Figure 2.5. Inhibition of DNA methylation increases DUX4fl RNA and protein and robustly upregulates target genes in mutant cells.

A. Hierarchical heatmap of DUX4 target gene expression for control, DEL3, and DEL3_SM_A mutants with or without 5AzaC at myoblast and early myotube. A total of 63 target genes were selected based on previous studies (38, 42). Expression values are in normalized TPM and log transformed.

B. DUX4 target genes were greatly affected by 5AzaC treatment in mutant cells. Control and SM1 myoblasts were treated with or without 5AzaC for 48 hrs. Then 5AzaC was removed from the media, and differentiation was induced. 4 days later, RT-qPCR of DUX4 target genes were performed. The gene expression data were normalized to *GAPDH* level in each sample, and then normalized to the *LEUTX* value of 5AzaC treated SM1. Data are presented as mean ± SD; *p<0.05, by one-tailed student's t-test.

C. Comparison of DUX4 target genes level between early myotubes of 5AzaC treated SM and DEL mutants. SM and DEL myoblasts were treated with 5AzaC for 48 hrs right before differentiation. At day 5 of differentiation, the mRNA expression level of DUX4 target genes was assessed by real-time RT-PCR, relative to SM1. Data are presented as mean ± SD; *p<0.05, by one-tailed student's t-test. Representative images of in situ detection of *LEUTX* RNA (red) with or without 5AzaC treatment are shown on the right (blue: DAPI). Scale bar 10 μm.

D. DUX4 depletion inhibited DUX4 target gene upregulation induced by 5AzaC treatment in mutant cells. SM1 and DEL3 cells were treated with 5AzaC and induced differentiation same as **(B)**. During 5AzaC treatment, cells were infected with lentivirus containing shCTRL or shDUX4. For each cell line, LEUTX expression level after DUX4 depletion was shown as fold difference compared to the control. Data are presented as mean ± SD; **p<0.01, by one-tailed student's t-test.

E. 5AzaC facilitated DUX4fl expression in DEL3 early myotubes. Left: the schematic diagrams of mRNA transcripts for *DUX4fl,* the *DUX4s* isoform and *DUX4* homologs (*DUX4c* and *DBET*), the black regions, which represent >99% homology to *DUX4fl*, could be detected by corresponding *DUX4* 4ZZ probes or 2ZZ probes, but not by both. Therefore, the overlapping signals from 4ZZ and 2ZZ probes represent the *DUX4fl* transcripts. Middle panel, example images of the RNAScope results of the 4ZZ probes (green), 2ZZ probes (red) and the overlapping foci (yellow). DAPI is in blue. Scale bar = 10μm. Right, 5AzaC treatment increased the percentage of myotubes with overlapping foci of 4ZZ and 2ZZ probes. Data from 3 independent experiments are presented as means ± SD. *p<0.05, by one-tailed student's t-test.

F. Examples of DUX4 protein expression in double mutant myoblasts and myotubes. Immunofluorescence for DUX4 on DEL4_SM_A cells after 5 days of differentiation. Nuclei were counterstained with DAPI (blue). Scale bar 50 μm.

G. Quantification of DUX4 protein expression with and without 5AzaC in double mutant myoblasts (left) and early myotubes (right). Mutant late myotubes and FSHD1 and FSHD2 patient early and late myotubes are shown for comparison. The DUX4 integrated density values in myoblasts/myotubes were measured using ImageJ software (37). Top 3% values in each group were used for graph and data analysis. All the data were normalized to the corresponding mean value of the 5AzaC-treated DEL_SM samples. *p<0.05, ** p<0.01, ***p<0.001. (Totally 600 myotubes or 1200 myoblasts were observed in each group).

H. Replotting the data in (G, right panel) for the frequency of DUX4 IF staining positive myotubes. *p<0.05, ** p<0.01, ***p<0.001.

Figure 2.6

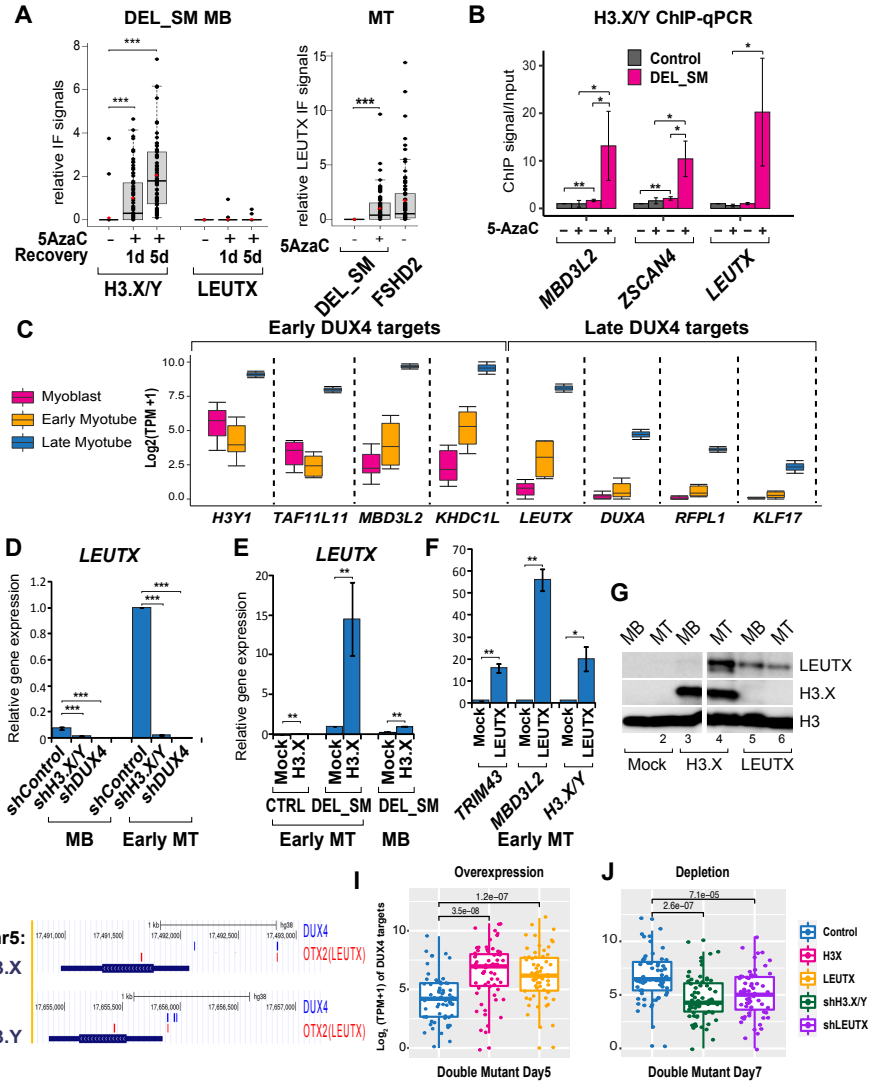

A
DEL_SM MB
MT
relative IF signals
relative LEUTX IF signals
5AzaC
Recovery
1d 5d
H3.X/Y
LEUTX
DEL_SM
FSHD2
B
H3.X/Y ChIP-qPCR
Control
DEL_SM
ChIP signal/Input
5-AzaC
MBD3L2
ZSCAN4
LEUTX
C
Early DUX4 targets
Late DUX4 targets
Myoblast
Early Myotube
Late Myotube
Log2(TPM+1)
H3Y1
TAF11L11
MBD3L2
KHDC1L
LEUTX
DUXA
RFPL1
KLF17
D
LEUTX
Relative gene expression
shControl
shH3.X/Y
shDUX4
MB
Early MT
E
LEUTX
Relative gene expression
Mock
H3.X
CTRL
DEL_SM
Early MT
MB
F
Mock
LEUTX
TRIM43
MBD3L2
H3.X/Y
Early MT
G
MB
MT
LEUTX
H3.X
H3
Mock
H3.X
LEUTX
H
chr5:
H3.X
H3.Y
DUX4
OTX2(LEUTX)
I
Overexpression
Log2(TPM+1) of DUX4 targets
Double Mutant Day5
J
Depletion
Double Mutant Day7
Control
H3X
LEUTX
shH3.X/Y
shLEUTX

Figure 2.6. Identification of early and late DUX4 target genes that are differentially sensitive to myotube differentiation.

A. Double mutant cells were treated with or without 5AzaC as indicated. IF signals of H3.X/Y and LEUTX in myoblasts (left panel) and LEUTX in early myotubes (right panel) were quantified as integrated intensity in each myoblast/myotube using ImageJ software. FSHD2 patient early myotubes are shown for comparison. The DUX4 integrated density values in each myoblast/myotube were measured using ImageJ software. Based on the highest positive myoblasts/myotubes number of all, same number of values in each group were used for graph and data analysis. All the data were normalized to the corresponding mean value of the 5AzaC treated samples (release day1 for the myoblasts). Red dots represent mean values. (N=300 myotubes or 1000 for myoblasts). ***p<0.001, by one-tailed student's t-test.

B. Incorporation of H3.X/Y into DUX4 targets in control and double mutant at Day 4 with or without 5AzaC. Cells were treated with 5AzaC for 48 hours before differentiation. Significant incorporation of H3.X/Y is shown by the asterisks with the indicated comparisons.

C. Box plots of representative early and late DUX4 target gene expression in double mutant cells were shown to compare myoblast, early and late myotube stages as indicated. Expression values are in $\log_2$ normalized TPM.

D. The expression level of DUX4 target gene LEUTX in double mutant DEL4_SM_A transduced with lentivirus carrying shControl, shH3.X/Y or shDUX4. Real-time RT-PCRs were carried out before or 5 days after the induction of differentiation. Three biological replicates for each sample were performed. Data are presented as mean ± SD; **p<0.01, ***p<0.001, by one-tailed student's t-test. Results presented as fold difference compared to shControl differentiated sample.

E. Control and DEL4_SM_A myoblasts were transduced with a lentiviral empty vector or a lentiviral vector expressing H3. X. Differentiation was induced at 48 hours after transduction. For myoblasts or early myotubes as indicated, the mRNA expression level of the downstream target genes was assessed by real-time RT-qPCR. Data are presented as mean ± SD; *p<0.05, **p<0.01, ***p<0.001, by one-tailed student's t-test. Results presented as fold difference compared to empty vector infected double mutant cells.

F. Similar experiments as in **(E)**, but DEL4_SM_A myoblasts were transduced with a lentiviral vector expressing LEUTX.

G. Overexpression of H3.X and LEUTX in MB or MT was assessed by western blot. Pan histone H3 antibody was used as control as indicated. Lanes 1 and 2: mock transfection. Lanes 3 and 4: H3.X OE. Lanes 5 and 6: LEUTX OE. The endogenous LEUTX is upregulated in H3.X OE myotubes (lane 4).

H. TF binding motifs at the promoter of H3.X/Y. Binding motifs for DUX4 and the putative LEUTX motif (OTX2) within 1 kb upstream and 0.5 kb downstream of the transcription start site for H3.X/Y were identified by using the MoLoTool provided in HOCOMOCO v11, with *P*-values less than or equal to 0.001. Visualization was done on the UCSC genome browser using GENCODE v36 for the H3.X/Y genes model.

I. The effects of control, H3.X/Y or LEUTX overexpression (as in **E-G**) on 63 DUX4 target genes in DEL_SM myotubes Day 5 are assessed by RNA-seq and displayed in box plots. *P*-values are calculated using Wilcoxon t-test indicated at the top.

J. Similar analysis was performed with H3.X/Y or LEUTX shRNA depletion compared to the same control as in **(I)** on DUX4 target gene expression in DEL_SM myotubes Day 7.

Figure 2.7

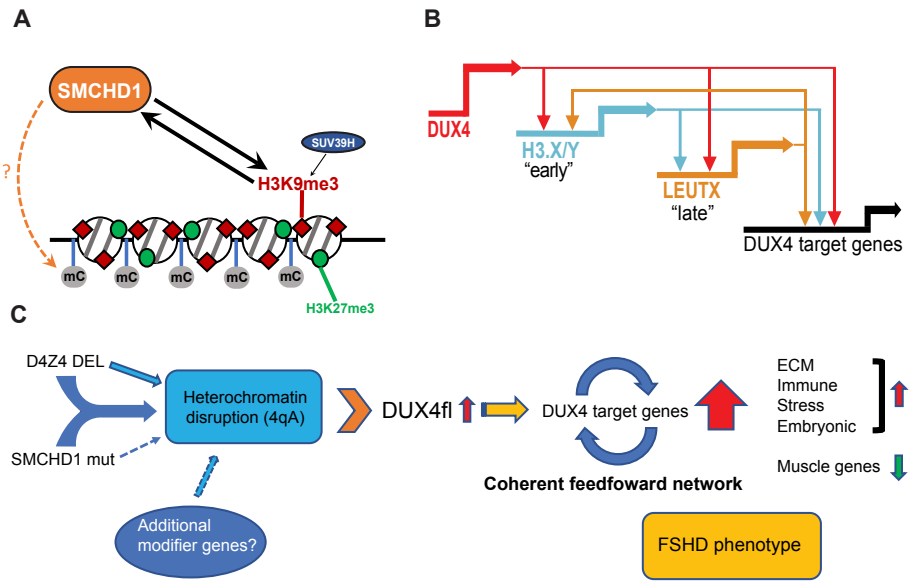

A
SMCHD1
SUV39H
H3K9me3
?
mC
mC
mC
mC
mC
H3K27me3
B
DUX4
H3.X/Y
"early"
LEUTX
"late"
DUX4 target genes
C
D4Z4 DEL
SMCHD1 mut
Heterochromatin disruption (4qA)
Additional modifier genes?
DUX4fl
DUX4 target genes
Coherent feedfoward network
ECM
Immune
Stress
Embryonic
Muscle genes
FSHD phenotype

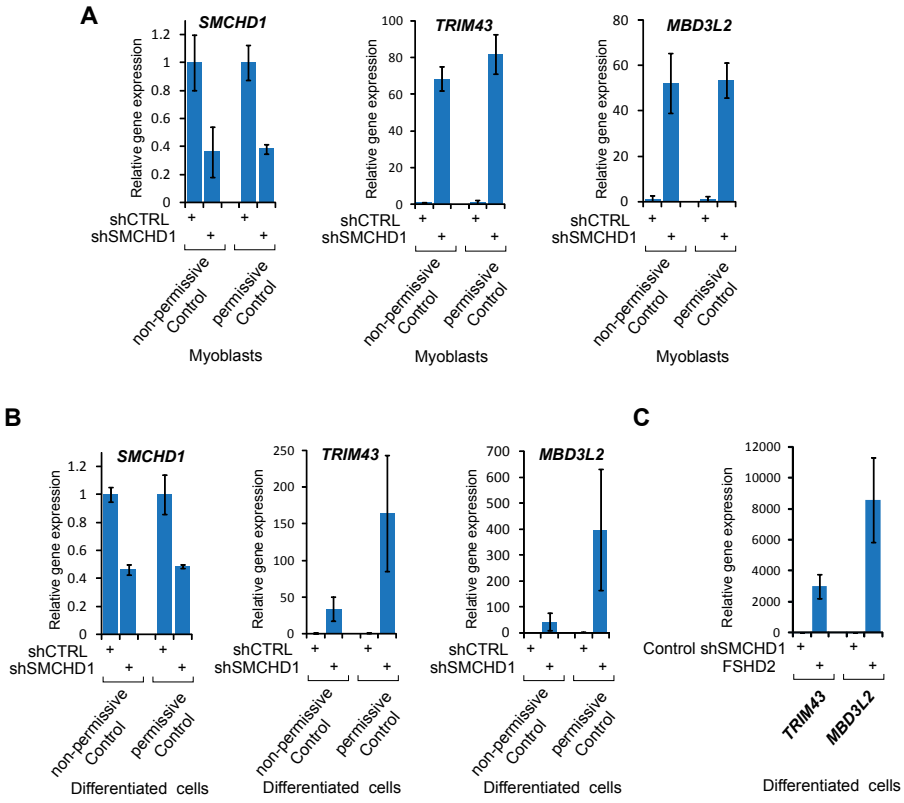
Figure 2.7. Models of the study

A. Positive feedback loop between H3K9me3 and SMCHD1. SMCHD1 interacts with D4Z4 chromatin in an H3K9me3-dependent manner (25) and also maintain H3K9me3.

B. Coherent feedforward mechanism of DUX4 and target gene expression. While DUX4 is critical for the initial activation of its target genes, the early target H3.X/Y expression is essential for efficient expression of other downstream target genes, including the late target TF, LEUTX. LEUTX in turn promotes further expression of H3.X/Y. H3.X/Y as well as LEUTX (and possibly other DUX4 target TFs) contribute significantly to the expression of other DUX4 target genes.

C. Two key processes in FSHD pathogenesis. D4Z4 heterochromatin disruption induced synergistically by D4Z4 and SMCHD1 mutations (and possibly other epigenetic modifiers) enables stabilization and enhancement of DUX4fl expression. Once activated by DUX4, DUX4 target genes undergo cross-regulation contributing to the establishment of the FSHD gene expression phenotype.

Figure S2.1:


A
SMCHD1
TRIM43
MBD3L2
Relative gene expression
shCTRL
shSMCHD1
non-permissive Control
permissive Control
Myoblasts
B
Differentiated cells
C
Control
shSMCHD1
FSHD2
TRIM43
MBD3L2

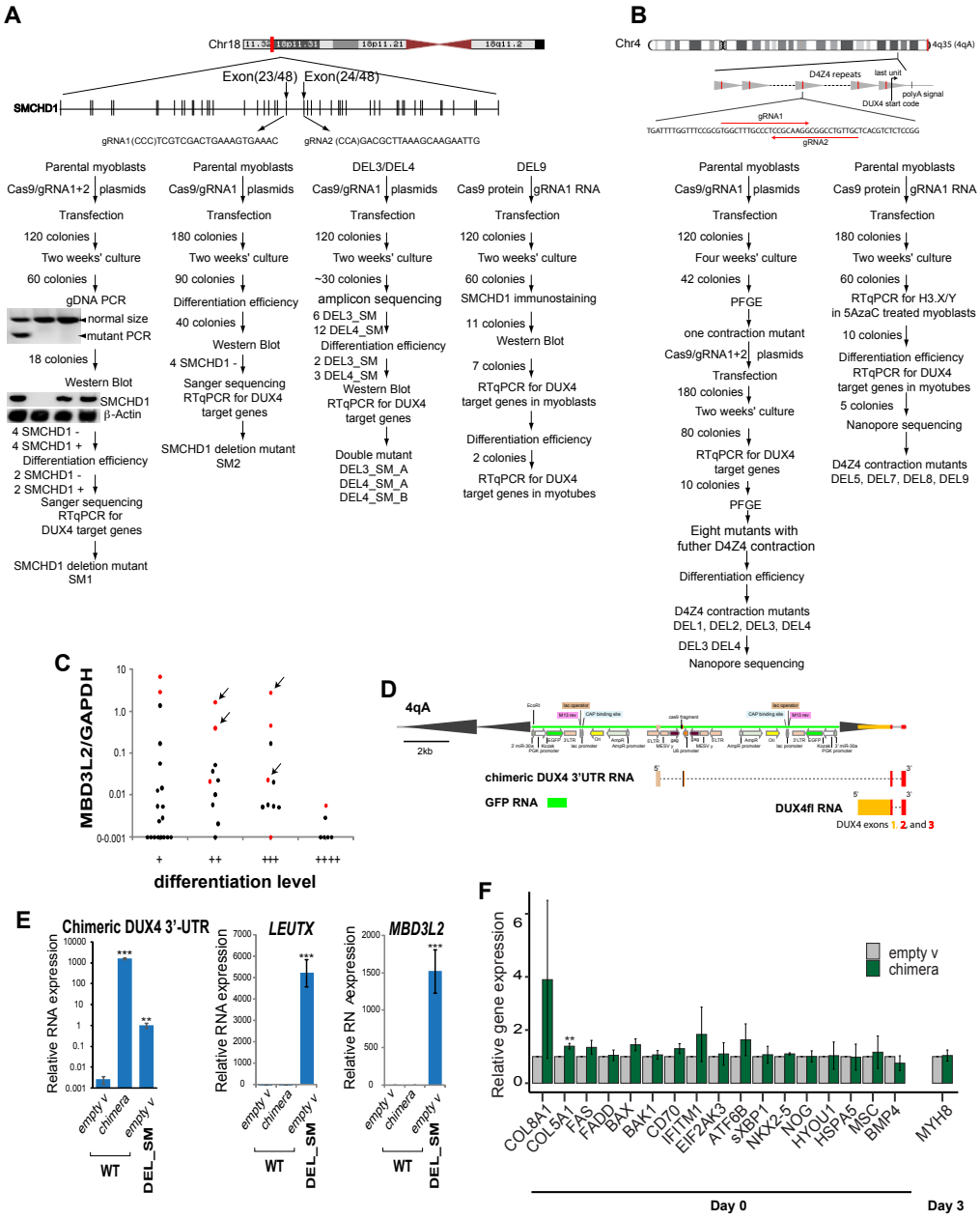
Supplemental Figure S2.1: Comparison of transient depletion of SMCHD1 in myoblasts with permissive or non-permissive haplotype.

A. Upregulation of DUX4 target genes are comparable in both haplotypes at the myoblast stage. RT-qPCR analyses of SMCHD1 depletion efficiency and expression of two DUX4 target genes (TRIM43 and MBD3L2) in permissive or non-permissive myoblasts treated with control shRNA and shRNA specific to SMCHD1. Data are presented as fold change in expression relative to the respective control shRNA-treated myoblasts.

B. Similar experiments as in (A) in myotubes. Following lentiviral shRNA infection, myoblasts were differentiated into myotubes for 3 days and RT-qPCR analyses were performed for *SMCHD1, TRIM43* and *MBD3L2* as indicated.

C. RT-qPCR comparison of shSMCHD1-treated myotubes and FSHD2 patient myotubes on day 3 of differentiation, demonstrating the vast overexpression of TRIM43 and MBD3L2 in patient myotubes.

Figure S2.2:


A
Chr18
Exon(23/48) Exon(24/48)
SMCHD1
gRNA1 (CCC)TCGTCGACTGAAAGTGAAAC
gRNA2 (CCA)GACGCTTAAAGCAAGAATTG
Parental myoblasts
Cas9/gRNA1+2 plasmids
Transfection
120 colonies
Two weeks' culture
60 colonies
gDNA PCR
normal size
mutant PCR
18 colonies
Western Blot
SMCHD1
β-Actin
4 SMCHD1 -
4 SMCHD1 +
Differentiation efficiency
2 SMCHD1 -
2 SMCHD1 +
Sanger sequencing
RTqPCR for DUX4 target genes
SMCHD1 deletion mutant SM1
Parental myoblasts
Cas9/gRNA1 plasmids
Transfection
180 colonies
Two weeks' culture
90 colonies
Differentiation efficiency
40 colonies
Western Blot
4 SMCHD1 -
Sanger sequencing
RTqPCR for DUX4 target genes
SMCHD1 deletion mutant SM2
DEL3/DEL4
Cas9/gRNA1 plasmids
Transfection
120 colonies
Two weeks' culture
~30 colonies
amplicon sequencing
6 DEL3_SM
12 DEL4_SM
Differentiation efficiency
2 DEL3_SM
3 DEL4_SM
Western Blot
RTqPCR for DUX4 target genes
Double mutant
DEL3_SM_A
DEL4_SM_A
DEL4_SM_B
DEL9
Cas9 protein
gRNA1 RNA
Transfection
120 colonies
Two weeks' culture
60 colonies
SMCHD1 immunostaining
11 colonies
Western Blot
7 colonies
RTqPCR for DUX4 target genes in myoblasts
Differentiation efficiency
2 colonies
RTqPCR for DUX4 target genes in myotubes
B
Chr4
4q35 (4qA)
D4Z4 repeats
last unit
polyA signal
DUX4 start code
gRNA1
gRNA2
Parental myoblasts
Cas9/gRNA1 plasmids
Transfection
120 colonies
Four weeks' culture
42 colonies
PFGE
one contraction mutant
Cas9/gRNA1+2 plasmids
Transfection
180 colonies
Two weeks' culture
80 colonies
RTqPCR for DUX4 target genes
10 colonies
PFGE
Eight mutants with futher D4Z4 contraction
Differentiation efficiency
D4Z4 contraction mutants
DEL1, DEL2, DEL3, DEL4
DEL3 DEL4
Nanopore sequencing
Parental myoblasts
Cas9 protein
gRNA1 RNA
Transfection
180 colonies
Two weeks' culture
60 colonies
RTqPCR for H3.X/Y in 5AzaC treated myoblasts
10 colonies
Differentiation efficiency
RTqPCR for DUX4 target genes in myotubes
5 colonies
Nanopore sequencing
D4Z4 contraction mutants
DEL5, DEL7, DEL8, DEL9
C
MBD3L2/GAPDH
differentiation level
D
4qA
2kb
chimeric DUX4 3'UTR RNA
GFP RNA
DUX4fl RNA
DUX4 exons 1, 2, and 3
E
Chimeric DUX4 3'-UTR
LEUTX
MBD3L2
Relative RNA expression
empty v
chimera
WT
DEL_SM
F
Relative gene expression
empty v
chimera
COL8A1
COL5A1
FAS
FADD
BAX
BAK1
CD70
IFITM1
EIF2AK3
ATF6B
sXBP1
NKX2-5
NOG
HYOU1
HSPA5
MSC
BMP4
MYH8
Day 0
Day 3

Supplemental Figure S2.2. Applications of CRISPR/Cas9 to generate FSHD modeling cells.

A. A scheme of the experimental procedure for generating and isolating single clones for *SMCHD1* knockout (SM) mutants. gRNA targeting sequences and PAM sequences(red) are shown. Two sets of screening for SM mutants and two sets of screening for double (DEL_SM) mutants were performed. *SMCHD1* mutants were confirmed by western blot and gRNA targeting region genomic DNA sequencing.

B. A scheme of the experimental procedure for screening and isolating single clones for D4Z4 deletion (DEL) mutants. For the D4Z4 repeat unit number and haplotype, the sizes of markers and their migration distance in the PFGE can be used to determine the size of the D4Z4 fragments using a standard curve Excel. Nanopore sequencing was also used to sequence chromosome 4 (chr4) D4Z4 region and chromosome 10 (chr10) D4Z4 region.

C. DUX4 target gene expression level and differentiation efficiency of potential single contraction mutant colonies. The cell lines marked with red dot were subjected to PFGE. Arrows indicate the cell lines used for other experiments in this paper. All the cell lines, which have the best differentiation efficiency, showed low MBD3L2 qPCR signal. One of them was confirmed no further contraction by using PFGE. All the tested cell lines with higher MBD3L2 signal, were deletion mutants.

D. Detailed schematic diagram of the Cas9/gRNA plasmids insertion in 4qA allele, the adjacent D4Z4 units, and the RNAs transcribed from this region.

E. Left: artificial overexpression of a chimeric DUX4 3'UTR RNA in parental WT cells and detection of the endogenous transcript in DEL_SM mutant cells by RT-qPCR. Middle and right: overexpression of chimeric DUX4 3'UTR induces no *LEUTX* or *MBD3L2* expression in WT cells. Expression of these genes in DEL_SM cells transfected with an empty vector is shown for comparison.

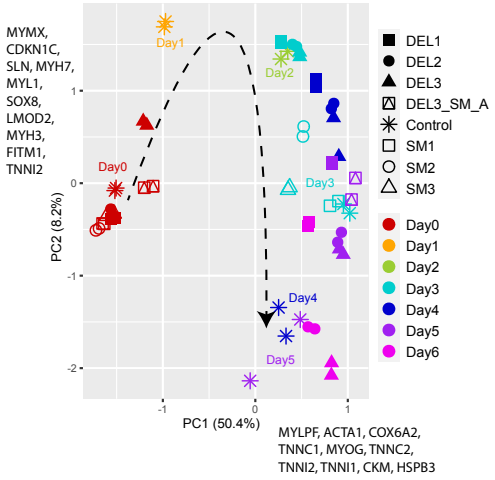
F. Similar overexpression of chimeric DUX4 3'UTR in myoblasts and myotubes had no effect on the genes found to be altered in mutant cells as in Figure 3C.

Figure S2.3

A

Cell line name	Doubling time (h)	Fusion index during differentiation				
		day1	day2	day3	day4	day5
Control	25.2	0	30%-60%	70%-80%	70%-90%	
FSHD1	24.5	0	0-20%	40%-60%	60%-80%	
FSHD2	28.5	0	10%-30%	60%-80%	60%-80%	
SM1	27.3	0	20%-40%	70%-80%	70%-90%	
DEL1	30.8	0	0	0%-30%	30%-60%	60%-90%
DEL2	29	0	0	0%-30%	30%-60%	60%-90%
DEL3	34.4	0	0	0%-30%	30%-50%	60%-90%
DEL3_SM_A	31.1	0	0	0%-30%	30%-50%	60%-90%
DEL4	24.8	0	0	0%-30%	40%-60%	70%-90%
DEL4_SM_A	25.9	0	0	0%-30%	40%-60%	70%-90%
DEL4_SM_B	26.9	0	0	0%-30%	30%-50%	60%-90%
DEL9	33.4	0	0-20%	60%-80%	70%-90%	

B

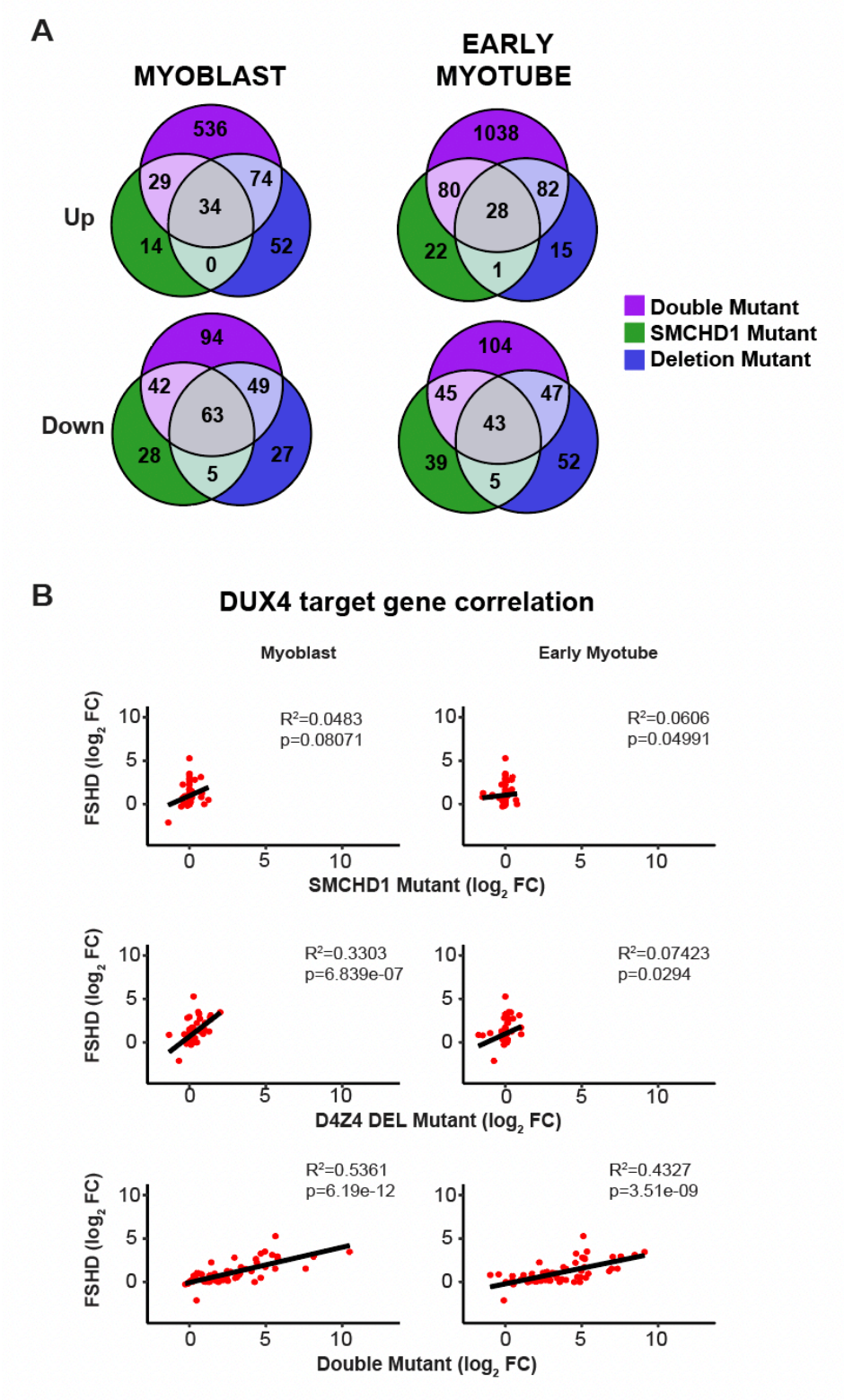


Supplemental Figure S2.3. Assessment of proliferation and differentiation of mutant myocytes.

A. A table summary for doubling time and differentiation efficiency. Total myoblast cell number in a field was counted when the cells were confluent one day before fusion. After fusion, the mononuclear cells number in this filed was counted at different day. Differentiation efficiency was then determined by [the initial cell number] – [the number of mononucleated cells left on each day] / [the initial cell number].

B. Principal component analysis of control and FSHD mutants along differentiation days from RNA-sequence experiment. The arrow indicates the trajectory of the control differentiation time course from Day 0 to Day 5. Both PCA1 and PCA2 explain variance across differentiation. Colors indicate days of differentiation and shapes indicate cell lines. The top 10 genes are shown for each component.

Figure S2.4


A
MYOBLAST
EARLY MYOTUBE
Up
Down
536
29
74
34
14
52
0
1038
80
82
28
22
15
1
94
42
49
63
28
27
5
104
45
47
43
39
52
5
Double Mutant
SMCHD1 Mutant
Deletion Mutant
B
DUX4 target gene correlation
Myoblast
Early Myotube
FSHD (log2 FC)
R²=0.0483
p=0.08071
R²=0.0606
p=0.04991
SMCHD1 Mutant (log2 FC)
R²=0.3303
p=6.839e-07
R²=0.07423
p=0.0294
D4Z4 DEL Mutant (log2 FC)
R²=0.5361
p=6.19e-12
R²=0.4327
p=3.51e-09
Double Mutant (log2 FC)

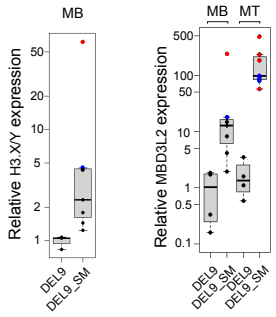
Supplemental Figure S2.4. RNA-seq data analyses of three types of mutants and patient cells.

A. Venn diagrams show the intersection of DEGs compared to control for each mutant type (SM, DEL and DEL_SM mutants) at myoblast and early myotube.

B. R square correlation plots for DUX4 targets (63 total) between both FSHD patient lines and each mutant type (SM, DEL, DEL_SM mutants) at myoblast and early myotube. *P*-values indicate significant correlations.

Figure S2.5

A



B

Myoblasts

Treatment		Total	H3X/Y (+)	LEUTX(+)
DEL4_SM_A no 5AzaC	repeat 1	312	1	0
	repeat 2	295	1	0
	repeat 3	389	1	0
DEL4_SM_A with 5AzaC recover 1 day	repeat 1	343	20	2
	repeat 2	312	16	0
	repeat 3	334	23	0
DEL4_SM_A with 5AzaC recover 5days	repeat 1	414	32	0
	repeat 2	320	25	0
	repeat 3	253	30	2

Myotubes

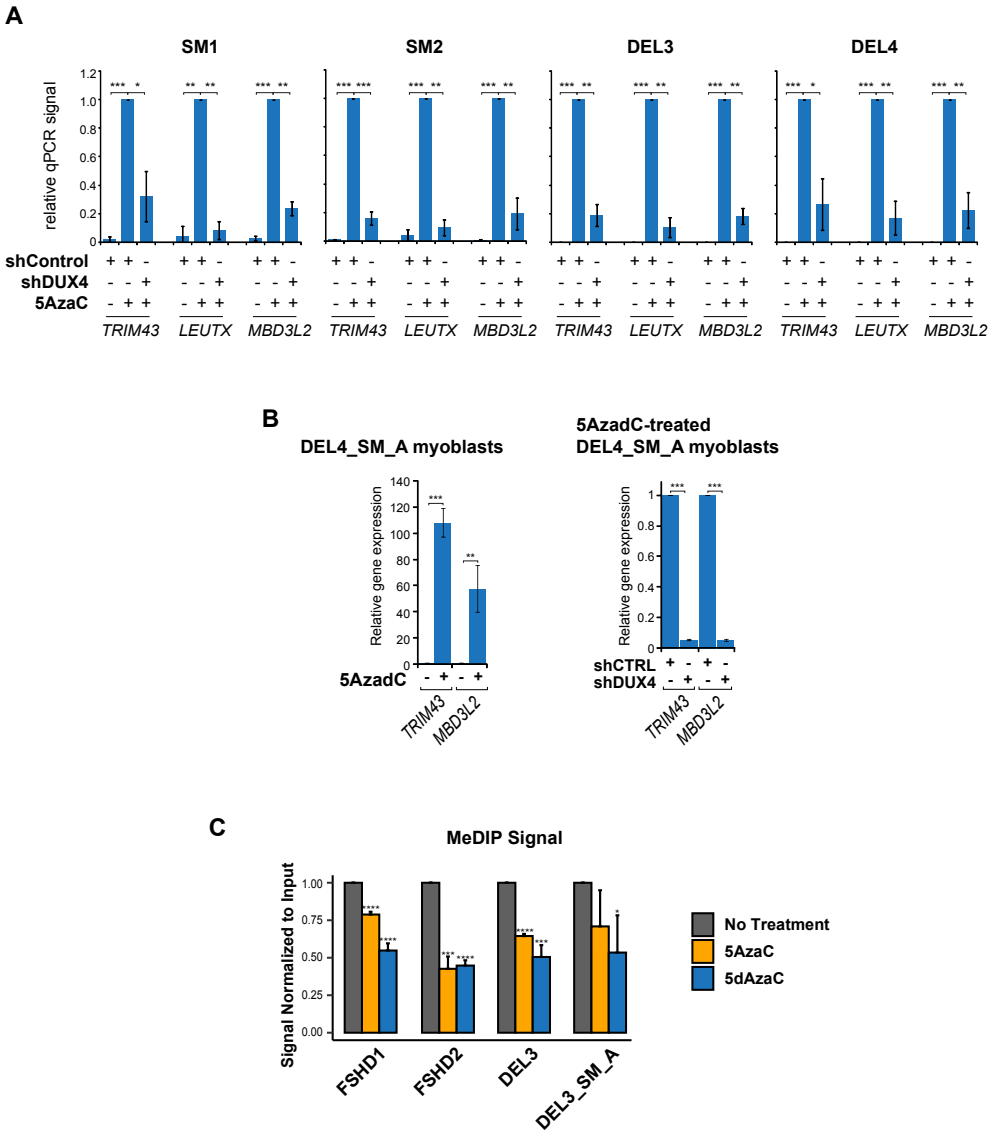
Cells and condition		Total	LEUTX(+)
DEL4_SM_A no 5AzaC differentiation day 5	repeat 1	100	0
	repeat 2	100	0
	repeat 3	100	0
DEL4_SM_A with 5AzaC differentiation day5	repeat 1	98	26
	repeat 2	78	13
	repeat 3	120	10
FSHD2 no 5AzaC differentiation day 4	repeat 1	105	23
	repeat 2	96	25
	repeat 3	102	11
FSHD2 with 5AzaC differentiation day 4	repeat 1	116	23
	repeat 2	89	31
	repeat 3	98	24

Supplemental Figure S2.5. Synergistic effect of double mutations and the effect of 5AzaC treatment.

A. The effect of SMCHD1 mutation on H3.X/Y and MBD3L2 expression in DEL9 mutant myoblasts (MB) and myotubes (MT) as indicated. Two double mutants (red and blue) with similar differentiation efficiency as DEL9, were used for the MT data.

B. The number of myotubes expressing H3.X/Y or LEUTX proteins without or with 5AzaC treated myoblasts or myotubes as indicated. Total number of cells counted as well as positive cells in triplicate experiments are shown.

Figure S2.6


A
SM1
SM2
DEL3
DEL4
relative qPCR signal
shControl
shDUX4
5AzaC
TRIM43
LEUTX
MBD3L2
B
DEL4_SM_A myoblasts
5AzadC-treated
DEL4_SM_A myoblasts
Relative gene expression
5AzadC
shCTRL
shDUX4
C
MeDIP Signal
Signal Normalized to Input
No Treatment
5AzaC
5dAzaC
FSHD1
FSHD2
DEL3
DEL3_SM_A

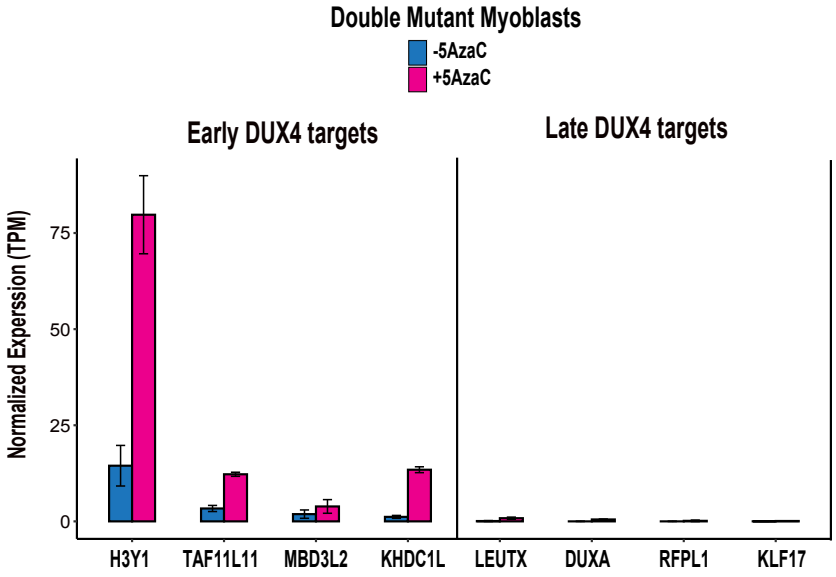
Supplemental Figure S6. Analysis of 5AzaC and 5AzadC treatment.

A. 5AzaC induced upregulation of DUX4 target genes in mutant cells was DUX4-dependent. Cells were treated with 5AzaC, infected with lentivirus containing shCTRL or shDUX4, and induced differentiation same as Figure 6D. For each cell line, DUX4 target genes expression level after DUX4 depletion was shown as fold difference compared to the control. Data are presented as mean ± SD; **p<0.01, by one-tailed student's t-test.

B. DUX4 target genes were upregulated by 5AzadC treatment in double mutant myoblasts (left panel). This upregulation was inhibited by DUX4 depletion (right panel). Cells were treated with 5AzadC for 24 h and harvested for gene expression analysis 2 days after release from 5AzadC. For the DUX4 depletion experiment, the cells were infected with shCTRL or shDUX4 lentivirus one day before 5AzadC treatment.

C. The effect of 5AzaC and 5dAzaC treatment on FSHD, single deletion and double mutant at Day 5 myotubes. Cells were treated with 5AzaC or 5dAzaC for 48 hours before differentiation. MeDIP analysis was performed to confirm the inhibition of DNA methylation at D4Z4. Significant inhibitions are indicated by the asterisks.

Figure S2.7


Double Mutant Myoblasts
-5AzaC
+5AzaC
Early DUX4 targets
Late DUX4 targets
Normalized Experssion (TPM)
75
50
25
0
H3Y1
TAF11L11
MBD3L2
KHDC1L
LEUTX
DUXA
RFPL1
KLF17

Supplemental Figure S7. Comparison of early and late DUX4 target genes in double mutant myoblasts with or without 5AzaC treatment.

In contrast to the early target genes, the late target genes are refractory to 5AzaC treatment at the myoblast stage.

2.7 References

1. J. C. Deenen, H. Arnts, S. M. van der Maarel, G. W. Padberg, J. J. Verschuuren, E. Bakker, S. S. Weinreich, A. L. Verbeek and B. G. van Engelen (2014) Population-based incidence and prevalence of facioscapulohumeral dystrophy. Neurol. 83:1056-9.

2. L. H. Wang and R. Tawil (2021) Current therapeutic approaches in FSHD. J. Neuromscul. Dis. 8:441-51.

3. S. M. van der Maarel and R. R. Frants (2005) The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. Am. J. Hum. Genet. 76:375-86.

4. S. M. van der Maarel, R. Tawil and S. J. Tapscott (2011) Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends Mol. Med. 17:252-8.

5. T. I. Jones, J. C. Chen, F. Rahimov, S. Homma, P. Arashiro, M. L. Beermann, O. D. King, J. B. Miller, L. M. Kunkel, C. P. J. Emerson, K. R. Wagner and P. L. Jones (2012) Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis. Hum. Mol. Genet. 21:4419-30.

6. R. J. L. F. Lemmers, R. Tawil, L. M. Petek, J. Balog, G. J. Block, G. W. E. Santen, A. M. Amell, P. J. van der Vliet, R. Almomani, K. R. Straasheijm, Y. D. Krom, R. Klooster, Y. Sun, J. T. den Dunnen, Q. Helmer, C. M. Donlin-Smith, G. W. Padberg, B. G. M. van Engelen, J. C. de Greef, A. M. Aartsma-Rus, R. R. Frants, M. de Visser, C. Desnuelle, S. Sacconi, G. N. Filippova, B. Bakker, M. J. Bamshad, S. J. Tapscott, D. G. Miller and S. M. van der Maarel (2012) Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat. Genet. 44:1370-4.

7. S. Sacconi, R. J. Lemmers, J. Balog, P. J. van der Vliet, P. Lahaut, M. P. van Nieuwenhuizen, K. R. Straasheijm, R. D. Debipersad, M. Vos-Versteeg, L. Salviati, A. Casarin, E. Pegoraro, R. Tawil, E. Bakker, S. J. Tapscott, C. Desnuelle and S. M. van der Maarel (2013) The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. Am. J. Hum. Genet. 93:744-51.

8. M. Larsen, S. Rost, N. El Hajj, A. Ferbert, M. Deschauer, M. C. Walter, B. Schoser, P. Tacik, W. Kress and C. R. Müller (2015) Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. Eur. J. Hum. Genet. 23:808-16.

9. J. Gabriëls, M. C. Beckers, H. Ding, A. De Vriese, S. Plaisance, S. M. van der Maarel, G. W. Padberg, R. R. Frants, J. E. Hewitt, D. Collen and A. Belayew (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene 236:25-32.

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

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

12. P. G. Hendrickson, J. A. Doráis, E. J. Grow, J. L. Whiddon, J. W. Lim, C. L. Wike, B. D. Weaver, C. Pflueger, B. R. Emery, A. L. Wilcox, D. A. Nix, C. M. Peterson, S. J. Tapscott, D. T. Carrell and B. R. Cairns (2017) Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet. 49:925-34.

13. J. L. Whiddon, A. T. Langford, C. J. Wong, J. W. Zhong and S. J. Tapscott (2017) Conservation and innovation in the DUX4-family gene network. Nat. Genet. 49:935-40.

14. R. J. L. F. Lemmers, P. J. van der Vliet, R. Klooster, S. Sacconi, P. Camaño, J. G. Dauwerse, L. Snider, K. R. Straasheijm, G. J. van Ommen, G. W. Padberg, D. G. Miller, S. J. Tapscott, R. Tawil, R. R. Frants and S. M. van der Maarel (2010) A unifying genetic model for facioscapulohumeral muscular dystrophy. Science 329:1650-3.

15. C. L. Himeda, T. I. Jones and P. L. Jones (2015) Facioscapulohumeral muscular dystrophy as a model for epigenetic regulation and disease. Antioxid. Redox. Signal. 22:1463-82.

16. K. Tsumagari, S. C. Chang, M. Lacey, C. Baribault, S. V. Chittur, J. Sowden, R. Tawil, G. E. Crawford and M. Ehrlich (2011) Gene expression during normal and FSHD myogenesis. BMC Med. Genomics 4:67.

17. N. Broucqsault, J. Morere, M. C. Gaillard, J. Dumonceaux, J. Torrents, E. Salort-Campana, A. Maues De Paula, M. Bartoli, C. Fernandez, A. L. Chesnais, M. Ferreboeuf, L. Sarda, H. Dufour, C. Desnuelle, S. Attarian, N. Levy, K. Nguyen, F. Magdinier and S. Roche (2013) Dysregulation of 4q35- and muscle-specific genes in fetuses with a short D4Z4 array linked to facio-scapulo-humeral dystrophy. Hum. Mol. Genet. 22:4206-14.

18. M. Ferreboeuf, V. Mariot, B. Bessières, A. Vasiljevic, T. Attié-Bitach, S. Collardeau, J. Morere, S. Roche, F. Magdinier, J. Robin-Ducellier, P. Rameau, S. Whalen, C. Desnuelle, S. Sacconi, V. Mouly, G. Butler-Browne and J. Dumonceaux (2014) DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. Hum. Mol. Genet. 23:171-81.

19. F. Rahimov, O. D. King, D. G. Leung, G. M. Bibat, C. P. J. Emerson, L. M. Kunkel and K. R. Wagner (2012) Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. Proc. Natl. Acad. Sci. 109:16234-9.

20. A. M. Rickard, L. M. Petek and D. G. Miller (2015) Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Hum. Mol. Genet. 24:5901-14.

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

22. J. C. de Greef, M. Wohlgemuth, O. A. Chan, K. B. Hansson, D. Smeets, R. R. Frants, C. M. Weemaes, G. W. Padberg and S. M. van der Maarel (2007) Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. Neurol. 69:1018-26.

23. J. C. de Greef, R. J. Lemmers, B. G. van Engelen, S. Sacconi, S. L. Venance, R. R. Frants, R. Tawil and S. M. van der Maarel (2009) Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Hum. Mol. Genet. 30:1449-59.

24. W. Zeng, J. C. de Greef, Y.-Y. Chen, R. Chien, X. Kong, H. C. Gregson, S. T. Winokur, A. Pyle, K. D. Robertson, J. A. Schmiesing, V. E. Kimonis, J. Balog, F. R. R., J. Ball, A. R., L. F. Lock, P. J. Donovan, S. van der Maarel and K. Yokomori (2009) 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.

25. W. Zeng, Y. Y. Chen, D. A. Newkirk, B. Wu, J. Balog, X. Kong, A. R. Ball, Jr., S. Zanotti, R. Tawil, N. Hashimoto, A. Mortazavi, S. M. van der Maarel and K. Yokomori (2014) 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.

26. W. Zeng, A. R. Ball and K. Yokomori (2012) The epigenetics of facioscapulohumeral muscular dystrophy. In: Appasani K (ed) Epigenomics: From Chromatin Biology to Therapeutics, Cambridge University Press, Cambridge pp. 347-361.

27. FA. Ashe, D. K. Morgan, N. C. Whitelaw, T. J. Bruxner, N. K. Vickaryous, L. L. Cox, N. C. Butterfield, C. Wicking, M. E. Blewitt, S. J. Wilkins, G. J. Anderson, T. C. Cox and E. Whitelaw (2008) A genome-wide screen for modifiers of transgene variegation identifies genes with critical roles in development. Genome Biol. 9:R182.

28. M. E. Blewitt, A. V. Gendrel, Z. Pang, D. B. Sparrow, N. Whitelaw, J. M. Craig, A. Apedaile, D. J. Hilton, S. L. Dunwoodie, N. Brockdorff, G. F. Kay and E. Whitelaw (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat. Genet. 40:663-9.

29. A.V. Gendrel, A. Apedaile, H. Coker, A. Termanis, I. Zvetkova, J. Godwin, Y. A. Tang, D. Huntley, G. Montana, S. Taylor, E. Giannoulatou, E. Heard, I. Stancheva and N. Brockdorff

(2012) Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. Dev. Cell 23:265-79.

30. V. Gendrel, Y. A. Tang, M. Suzuki, J. Godwin, T. B. Nesterova, J. M. Greally, E. Heard and N. Brockdorff (2013) Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes. Mol. Cell. Biol. 33:3150-65.

31. D. Bosnakovski, Z. Xu, E. J. Gang, C. L. Galindo, M. Liu, T. Simsek, H. R. Garner, S. Agha-Mohammadi, A. Tassin, F. Coppée, A. Belayew, R. R. Perlingeiro and M. Kyba (2008) An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. EMBO J. 27:2766-79.

32. Vanderplanck, E. Ansseau, S. Charron, N. Stricwant, A. Tassin, D. Laoudj-Chenivesse, S. D. Wilton, F. Coppée and A. Belayew (2011) The FSHD atrophic myotube phenotype is caused by DUX4 expression. PLoS One 6:e26820.

33. Q. Feng, L. Snider, S. Jagannathan, R. Tawil, S. M. van der Maarel, S. J. Tapscott and R. K. Bradley (2015) A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. Elife 4:doi: 10.7554/eLife.04996.

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

35. S. Jagannathan, Y. Ogata, P. R. Gafken, S. J. Tapscott and R. K. Bradley (2019) Quantitative proteomics reveals key roles for post-transcriptional gene regulation in the molecular pathology of facioscapulohumeral muscular dystrophy. Elife 8:e41740.

36. A.Lek, Y. Zhang, K. G. Woodman, S. Huang, A. M. DeSimone, J. Cohen, V. Ho, J. Conner, L. Mead, A. Kodani, A. Pakula, N. Sanjana, O. D. King, P. L. Jones, K. R. Wagner, M. Lek and L. M. Kunkel (2020) Applying genome-wide CRISPR-Cas9 screens for therapeutic discovery in facioscapulohumeral muscular dystrophy. Sci. Transl. Med. 12:eaay0271.

37. J. Chau, X. Kong, N. Nguyen, K. Williams, M. Ball, R. Tawil, T. Kiyono, A. Mortazavi and K. Yokomori (2021) Relationship of DUX4 and target gene expression in FSHD myocytes. Hum. Mutat. 42:421-33.

38. S. Jiang, K. Williams, X. Kong, W. Zeng, X. Ma, R. Tawil, K. Yokomori and A. Mortazavi (2020) Single-nucleus RNA-seq identifies divergent populations of FSHD2 myotube nuclei. PLoS Genet. 16:e1008754. doi: doi: 10.1371/journal.pgen.1008754.

39. S. Homma, M. L. Beermann, F. M. Boyce and J. B. Miller (2015) Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. Ann. Clin. Transl. Neurol. 2:151-66.

40. Bosnakovski, S. S. K. Chan, O. O. Recht, L. M. Hartweck, C. J. Gustafson, L. L. Athman, D. A. Lowe and M. Kyba (2017) Muscle pathology from stochastic low level DUX4 expression in an FSHD mouse model. Nat. Commun. 8:550.

41. T. Jones and P. L. Jones (2018) A cre-inducible DUX4 transgenic mouse model for investigating facioscapulohumeral muscular dystrophy. PLoS One 13:e0192657.

42. Z. Yao, L. Snider, J. Balog, R. J. Lemmers, S. M. Van Der Maarel, R. Tawil and S. J. Tapscott (2014) DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. Hum. Mol. Genet. 23:5342-52.

43. Dion, S. Roche, C. Laberthonnière, N. Broucqsault, V. Mariot, S. Xue, A. D. Gurzau, A. Nowak, C. T. Gordon, M. C. Gaillard, C. El-Yazidi, M. Thomas, A. Schlupp-Robaglia, C. Missirian, V. Malan, L. Ratbi, A. Sefiani, B. Wollnik, B. Binetruy, E. Salort Campana, S. Attarian, R. Bernard, K. Nguyen, J. Amiel, J. Dumonceaux, J. M. Murphy, J. Déjardin, M. E. Blewitt, B. Reversade, J. D. Robin and F. Magdinier (2019) SMCHD1 is involved in de novo methylation of the DUX4-encoding D4Z4 macrosatellite. Nucleic Acids Res. 47:2822-39.

44. R. J. M. Goselink, T. H. A. Schreuder, N. van Alfen, I. J. M. de Groot, M. Jansen, R. Lemmers, P. J. van der Vliet, N. van der Stoep, T. Theelen, N. C. Voermans, S. M. van der Maarel, B. G. M. van Engelen and C. E. Erasmus (2018) Facioscapulohumeral dystrophy in childhood: a nationwide natural history study. Ann. Neurol. 84:627-637. doi: 10.1002/ana.25326

45. S. T. Winokur, Y. W. Chen, P. S. Masny, J. H. Martin, J. T. Ehmsen, S. J. Tapscott, S. M. van der Maarel, Y. Hayashi and K. M. Flanigan (2003) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. Hum. Mol. Genet. 12:2895-907.

46. M. Jouhilahti, E. Madissoon, L. Vesterlund, V. Töhönen, K. Krjutškov, A. Plaza Reyes, S. Petropoulos, R. Månsson, S. Linnarsson, T. Bürglin, F. Lanner, O. Hovatta, S. Katayama and J. Kere (2016) The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. Development 143:3459-69.

47. R. Resnick, C. J. Wong, D. C. Hamm, S. R. Bennett, P. J. Skene, S. B. Hake, S. Henikoff, S. M. van der Maarel and S. J. Tapscott (2019) DUX4-induced histone variants H3.X and H3.Y mark DUX4 target genes for expression. Cell Rep. 29:1812-1820.e5.

48. H. Kim, S. Ham, M. Jo, G. H. Lee, Y. S. Lee, J. H. Shin and Y. Lee (2017) CRISPR-Cas9 mediated telomere removal leads to mitochondrial stress and protein aggregation. Int. J. Mol. Sci. 18:2093. doi: 10.3390/ijms18102093

49. J. C. de Greef, Y. D. Krom, B. den Hamer, L. Snider, Y. Hiramuki, R. F. P. van den Akker, K. Breslin, M. Pakusch, D. C. F. Salvatori, B. Slütter, R. Tawil, M. E. Blewitt, S. J. Tapscott and S. M. van der Maarel (2018) Smchd1 haploinsufficiency exacerbates the phenotype of a

transgenic FSHD1 mouse model. Hum. Mol. Genet. 27:716-731. doi: 10.1093/hmg/ddx437

50. M. Sasaki-Honda, T. Jonouchi, M. Arai, A. Hotta, S. Mitsuhashi, I. Nishino, R. Matsuda and H. Sakurai (2018) A patient-derived iPSC model revealed oxidative stress increases facioscapulohumeral muscular dystrophy-causative DUX4. Hum. Mol. Genet. 27:4024-35.

51. R. S. Nozawa, K. Nagao, K. T. Igami, S. Shibata, N. Shirai, N. Nozaki, T. Sado, H. Kimura and C. Obuse (2013) Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. Nat. Struc. Mol. Biol. 20:566-73.

52. N. J. Brideau, H. Coker, A. V. Gendrel, C. A. Siebert, K. Bezstarosti, J. Demmers, R. A. Poot, T. B. Nesterova and N. Brockdorff (2015) Independent Mechanisms Target SMCHD1 to Trimethylated Histone H3 Lysine 9-Modified Chromatin and the Inactive X Chromosome. Mol. Cell. Biol. 35:4053-68.

53. K. Hamanaka, D. Šikrová, S. Mitsuhashi, H. Masuda, Y. Sekiguchi, A. Sugiyama, K. Shibuya, R. Lemmers, R. Goossens, M. Ogawa, K. Nagao, C. Obuse, S. Noguchi, Y. K. Hayashi, S. Kuwabara, J. Balog, I. Nishino and S. M. van der Maarel (2020) Homozygous nonsense variant in LRIF1 associated with facioscapulohumeral muscular dystrophy. Neurol. 94:e2441–e2447.

54. Z. Huang, J. Yu, W. Cui, B. K. Johnson, K. Kim and G. P. Pfeifer (2021) The chromosomal protein SMCHD1 regulates DNA methylation and the 2c-like state of embryonic stem cells by antagonizing TET proteins. Sci. Adv. 7. doi: 10.1126/sciadv.abb9149

55. A.E. Campbell, A. E. Belleville, R. Resnick, S. C. Shadle and S. J. Tapscott (2018) Facioscapulohumeral dystrophy: activating an early embryonic transcriptional program in human skeletal muscle. Hum. Mol. Genet. 27:R153-r162. doi: 10.1093/hmg/ddy162

56. E. Campbell, S. C. Shadle, S. Jagannathan, J. W. Lim, R. Resnick, R. Tawil, S. M. van der Maarel and S. J. Tapscott (2018) NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. Elife 7:e31023. doi: 10.7554/eLife.31023

57. S. G. Jin, C. L. Jiang, T. Rauch, H. Li and G. P. Pfeifer (2005) MBD3L2 interacts with MBD3 and components of the NuRD complex and can oppose MBD2-MeCP1-mediated methylation silencing. J. Biol. Chem. 280:12700-9. doi: 10.1074/jbc.M413492200

58. Peter and E. Davidson (2015) Genomic Control Process: Development and Evolution. Elsevier.

59. M. Ehrlich, K. Jackson, K. Tsumagari, P. Camaño and R. J. Lemmers (2007) Hybridization analysis of D4Z4 repeat arrays linked to FSHD. Chromosoma 116:107-16.

60. Shiomi, T. Kiyono, K. Okamura, M. Uezumi, Y. Goto, S. Yasumoto, S. Shimizu and N. Hashimoto (2011) CDK4 and cyclin D1 allow human myogenic cells to recapture growth property without compromising differentiation potential. Gene Ther. 18:857-66.

61. J.-S. Kim, T. B. Krasieva, V. J. LaMorte, A. M. R. Taylor and K. Yokomori (2002) Specific recruitment of human cohesin to laser-induced DNA damage. J. Biol. Chem. 277:45149-153. doi: 10.1074/jbc.M209123200

62. O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. Mikkelson, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench and F. Zhang (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343:84-87.

Chapter 3

The crossregulation of H3.X/Y and LEUTX in the DUX4 gene network and the role of LEUTX in early embryonic development

3.1 Abstract

Within the past decades, DUX4 target genes have been focused and characterized in order to define FSHD pathology. Though misexpression of DUX4 in FSHD patients is rare but its burst-like expression is sufficient to activate and maintain the downstream target expression. One of the DUX4 target, histone variant H3.X (or Y) was recently found to enhance DUX4 activation in FSHD patients. Depletion of DUXA, another DUX4 target, was also shown to reduce the expression of other targets such as ZSCAN4 and LEUTX. We proposed previously that DUX4 target gene expression consists of two-steps cross-regulation: (1) early/differentiation-independent targets such as H3.X/Y and MBD3 family can positively feedback into DUX4 network to enhance (2) late/differentiation-dependent targets such as LEUTX and DUXA which can also feedback and further enhance the DUX4 network. In this study, we performed overexpression and depletion of the DUX4 targets, H3.X/Y, LEUTX and DUXA, and determined the cross-interaction between H3.X/Y and LEUTX in regulating the DUX4 gene network. Particularly, we discovered that overexpressed LEUTX activated a specific set of embryonic genes along with DUX4 targets in FSHD muscle cells. Together, these results provided further insight into the mechanism of FSHD pathology and also delineate the role of LEUTX in the DUX4-triggered embryonic program in FSHD myocytes.

3.2 Introduction

FSHD is one of the most common forms of muscular dystrophies. The disease is linked to the misexpression of DUX4 from the last repeat of D4Z4 macrosatellite array of the sub-telomeric 4qA(*1*, *2*). DUX4 is a transcription factor which belongs to the PRD-like class family and plays a central role in EGA program(*3*–*7*). The expression of DUX4 is generally low and mostly undetectable in bulk RNA-seq sequencing. However, DUX4 target genes are highly expressed in FSHD patients and often used as biomarker genes for FSHD(*4*, *8*–*10*). Some DUX4 target genes were shown to act in positive feedback manner which maintains and enhances DUX4 activation(*11*–*13*). H3.X and H3.Y which are histone variants were demonstrated to be ones of the earliest targets of DUX4. Both H3.X and H3.Y are known to incorporated into other DUX4 target genes, increasing their expression(*11*). MBD3L family were also demonstrated to inhibit MBD3, releasing DUX4 repression. Depletion of PRD-like class DUX4 target genes, LEUTX and DUXA, also inhibited other DUX4 targets(*13*, *14*). These findings demonstrated that DUX4 relies on its own target genes to maintain and enhance the network. Therefore, DUX4 and its target gene expression continue to sustain in FSHD patients without any obvious level of DUX4 protein. This observation suggests that DUX4 inhibition alone might not an effective strategy to block DUX4 activation as the target genes can still carry out the function of DUX4 downstream. Additional inhibition of DUX4 target genes such as H3.X/Y and LEUTX, for example, could be a better approach to completely inactivate DUX4 network.

DUXA and LEUTX which are double homeobox proteins like DUX4 are recently shown to be involved in EGA(*15*–*17*). They are expressed from 4-cell to 8-cell stage after the peak of DUX4 (4-cell stage). *LEUTX* is suggested to be one of the earliest expressed genes in embryo preimplantation. DPRX, a novel PRD-like class protein, is upregulated at 8-cell stage and suggested to act as a repressor to balance *LEUTX* expression(*15*). *LEUTX* is normally only expressed in pluripotent cells with the exception of FSHD patients where DUX4 is present(*15*). Therefore, misexpression of a potent transcription factor like LEUTX in myocytes is likely to disrupt the muscle differentiation program. However, our understanding of the effect of LEUTX in myocytes has been limited.

In this study, we overexpressed or depleted three DUX4 target genes (H3.X/Y, LEUTX and DUXA) in parental control, FSHD double mutant (Chapter2) and FSHD2, and examined their effect by RNA-seq analysis. The results revealed that both H3.X/Y and LEUTX overexpression further increased DUX4 target genes whereas their depletions decreased the target genes in FSHD double mutant. Both H3.X/Y overexpression and depletion have little effect on DUX4 targets in FSHD2, suggesting the variable effect of H3.X/Y due to cell line differences. Surprisingly, both depletion and overexpression of DUXA inhibit the DUX4 network in FSHD2 and double mutant. Furthermore, overexpression of LEUTX in FSHD myotubes upregulated genes involved in embryo implantation. These results highlight the significant role of LEUTX in regulating the DUX4 gene network, contributing to FSHD pathology.

3.3 Results

Differential gene expression of H3.X/Y, LEUTX, and DUXA by overexpression and depletion in control, FSHD2 and double mutant myotubes

We chose a control healthy myoblast or parental wildtype (B0) with a normal D4Z4 repeat size, a FSHD2 cell line (B2) and a double mutant (DEL4_SM_A) from my chapter 2 and examined the effect of depletion and overexpression of H3.X/Y, LEUTX, and DUXA. All cell lines were infected with lentivirus expressing either shRNA or a corresponding recombinant protein for 9 hours, 2 days prior to differentiation. We also included two empty lentiviral vectors as controls (one for shRNA knockdown and the other for overexpression condition). The cells were allowed to differentiate after 48 hours as depicted in the scheme (Figure 3.1A). The parental control and FSHD2 differentiated myotube detached at around Day 5 while the double mutant differentiated myotubes detached at around Day 7. DUX4 target gene expression generally decreased following myotube detachment. Therefore, we wanted to harvest the cells right before the detachment to obtain the optimal DUX4 target gene expression. The control and FSHD2 samples were harvested at Day 5. Meanwhile, H3.X/Y, LEUTX, and DUXA were expressed significantly at the early myotube (Day 4-6) of the double mutants. Therefore, to better examine the effect of overexpression and depletion of these targets in the double mutant, we harvested the samples at day 5 for the overexpression of H3.X/Y and LEUTX while the depletion samples were harvested at Day 7 right before the myotube detachment. DUXA overexpression and depletion samples were both harvested at Day 7 because DUXA expression occurs later than H3.X/Y and LEUTX (Chapter 2). The harvested samples were

proceeded to DNA library building and sequencing using Smart-Seq 2 protocol (*13*). All the samples collected for this study were summarized in Table S3.1.

We confirmed the increased expression of H3.X/Y and LEUTX in parental control and double mutant by lentiviral transduction of the recombinant H3.X and LEUTX, respectively, compared to the control virus treatment (Figure 3.1B). However, H3.X/Y expression did not increase in FSHD2. This may be due to the already high basal level of H3.X in this FSHD2 cell line. H3.X and LEUTX overexpression reciprocally upregulated each other only in double mutant but not in control cell line. Furthermore, the increased expression of either H3.X/Y or LEUTX was more pronounced in double mutant than control. This result demonstrated that DUX4 is necessary for enhanced expression of H3.X/Y, LEUTX and DUXA. To our surprise, DUXA overexpression decreased DUXA instead in both FSHD2 and double mutant. Depletion of H3.X/Y, LEUTX and DUXA also cross-downregulate each other in FSHD2 and double mutant (Figure 3.1C). Depletion of H3.X/Y and LEUTX appeared to have a stronger inhibitory effect than depletion of DUXA on the other two target genes, suggesting the important role of early targets in setting up DUX4 activation as DUXA is activated later than H3.X/Y and LEUTX.

Overexpression of LEUTX has a stronger global impact than overexpression of H3.X or DUXA

We selected 63 DUX4 target genes for hierarchically clustered heatmap based on previous paper comparing between double mutant and control cell line (Figure 3.2A) (*13*, *18*). Similar to the result in Figure 3.1, overexpression of H3.X, LEUTX, or DUXA in parental

control had no global effect on DUX4 target genes. In contrast, overexpression of H3.X/Y and LEUTX further increased global DUX4 target genes at Day 5 while DUXA overexpression had the opposite effect at Day 7. The effect of the overexpression (normalized $\log_2$ TPM) on DUX4 target genes was significant in double mutant (Figure 3.2B). The result indicated that H3.X/Y and LEUTX act as positive feedback loop to boost DUX4 activation while DUXA might act as a repressor of DUX4.

EdgeR analysis was performed to determine differential expressed genes (DEGs) with *P*-values =< 0.01 and Log_2 Fold Change >= 1. Volcano plot showed the DEGs in each comparison between double mutant and parental control, DUX4 targets were also indicated (Figure 3.2C). H3.X overexpression had smaller impact on global differential gene expression compared to LEUTX and DUXA in both double mutant and parental control. Overexpression of LEUTX induced more upregulated genes than downregulated genes in both double mutant and parental control. Additionally, overlapped DEGs between double mutant and parental control were more pronounced in LEUTX and DUXA overexpression (only 1 for H3.X/Y in upregulated and downregulated DEGs) (Figure S3.1A). The result implied that H3.X/Y alone has very little impact on the global expression and might only work in conjunction with the presence of DUX4 to induce noticeable DEGs. Meanwhile, LEUTX and DUXA induced clear DEGs in parental control without activated DUX4 though DUX4 target genes were not upregulated. This demonstrated that LEUTX and DUXA activated some set of genes that are not DUX4 targets. Gene ontology (GO) analysis of upregulated DEGs of LEUTX and DUXA overexpression in control cell was performed with the indicated categories (Figure S3.1B). Taken together, overexpression of LEUTX significantly upregulated DUX4 target genes and also induced non-DUX4 target genes.

The global effect of depletion on of H3.X/Y, LEUTX, and DUXA on DUX4 target genes

We performed similar analysis as described above to examine the effect of depletion of H3.X/Y, LEUTX, and DUXA on double mutant and FSHD2. Hierarchical clustered heatmap of 63 DUX4 targets showed a global reduction of DUX4 target gene expression for all shRNA depletions in double mutant (Figure 3.3A). Boxplot of expressed DUX4 target genes (normalized $\log_2$ TPM) showed that H3.X/Y depletion had the most significant effect on DUX4 target gene expression followed by LEUTX and DUXA (Figure 3.3B). However, only depletion of LEUTX and DUXA were able to inhibit DUX4 target genes in FSHD2 (Figure S3.2A,B). The result further confirmed that H3.X/Y depletion or overexpression had no effect on FSHD2 patients while both H3.X/Y depletion and overexpression affected DUX4 target genes significantly in double mutant. This finding suggested the variable effect of H3.X/Y due to cell line differences (Figure 3.2A,B and Figure 3.3A,B). Interestingly, both depletion and overexpression of DUXA triggered inhibition of DUX4 target genes (Figure 3.2, Figure 3.3, and Figure S3.2). This observation implied that DUXA can both activate and inhibit DUX4 target genes and optimized expression of DUXA might be required to achieve sustainable DUX4 target expression.

Volcano plot showed global DEGs in comparison between double mutant and FSHD2 under depletion condition (Figure 3.3C). The global effect of LEUTX and DUXA depletion was significantly stronger than H3.X/Y depletion in FSHD2. DUX4 target genes were also more reduced by depletion of LEUTX and DUXA in FSHD2 (orange color). Meanwhile, H3.X/Y depletion reduced DUX4 target genes significantly in double mutant. This result

also suggested cell line variability of H3.X/Y, LEUTX and DUXA shRNA depletion in terms of global effect. Nevertheless, in addition to overexpression results, the depletion results further confirmed that H3.X/Y, LEUTX and DUXA cross-regulate each other in a positive feedback loop which affects the rest of DUX4 target gene expression in double mutant.

WGCNA analysis reveals 5 gene clusters

To explore the correlation of genes affected by different treatments, we performed WGCNA (Weighted correlation network analysis) on all the samples(*19*). We characterized 5 clusters/modules of genes: blue, orange, purple, turquoise, gold (Figure 3.4A). Blue module correlated positively with FSHD2 and double mutant in general. The eigenvalues decreased with shH3.X/Y, shLEUTX and shDUXA depletion but increased with H3.X and LEUTX overexpression (Figure 3.4A). Module blue contained 55 DUX4 target genes of 64 selected for the study (Table S3.2). Hierarchical heatmap of module blue showed gene expression of DUX4 targets and other genes (Figure 3.4B). The majority of genes in the blue module are non-DUX4 target genes (266 genes). Interestingly, DUXA overexpression in double mutant drastically rescued most of the genes in the blue module close to parental control (Figure 3.4B). Gene ontology (GO) analysis of blue module indicated the genes in the module are associated with ubiquitin-related pathway and innate immune response. Ubiquitin proteasome pathway was also identified in FSHD proteasome study(*20*).

Orange module correlated negatively with parental control but positively with LEUTX overexpression in double mutant and parental control (Figure 3.4A). It also correlated positively with FSHD2 overall. GO analysis indicated that orange module is

associated with RNA processing (Figure 3.4C). Purple module is correlated with DUXA and LEUTX overexpression in parental control and also associated with RNA processing (Figure 3.4A, C). This suggested that LEUTX and DUXA might be involved in rRNA processing in embryogenesis. Past findings also indicated the central role of rRNA genes during preimplantation embryo development[18]. Turquoise module showed increase in eigenvalues of parental control and double mutant with H3.X/Y overexpression compared to control treatment (Figure 3.4A). Gene ontology indicated that the turquoise module was associated with stress/apoptotic responses.

Gold module correlated positively with parental control. It also correlated positively with LEUTX depletion in FSHD2 and both LEUTX and H3.X/Y depletion in double mutant (Figure 3.4A). GO analysis indicated that genes in gold module were associated with muscle related genes (Figure 3.4C). Depletion of H3.X/Y and LEUTX in double mutant was likely to reverse the delayed differentiation compared to control treatment and also parental control (Figure 3.4B). The rescued muscle genes included TNNI and TNNC family (Table S3.2). However, the rescue effect was only pronounced for LEUTX depletion in FSHD2 but not H3.X/Y depletion. This observation suggested that inhibition of DUX4 network by shRNA depletion of H3.X/Y and LEUTX improved muscle differentiation in FSHD. Non-muscle genes that were rescued by depletion of H3X./Y and LEUTX were also shown (Figure 3.4B and Table S3.2).

LEUTX regulates embryo preimplantation in myotubes

Because LEUTX was found to be involved in preimplantation embryo development, we investigated the role of LEUTX in embryonic regulation(*15*). There were 115 DEGs overlapped between LEUTX overexpression and LEUTX depletion in double mutant, and of those overlapped were 58 DUX4 target genes (Figure 3.5A). We selected 1373 upregulated genes by LEUTX overexpression in double mutant to determine transcription factors (TFs) binding at the promoter regions of these genes by homer motif enrichment analysis(*21*). The result showed enrichment of TFs belong to PRD-like class such as GSC, DPRX, OTX1, OTX2, DUX4, and PITX1 (Figure 3.5B, Table S3.3). KFL3 which is associated with 8-cell embryonic stage was also enriched(*22*). Other TFs from KFL family also enriched at these promoter regions (Table S3.3). Interestingly, DPRX, LEUTX, OTX1 and OTX2 belong to K50-type PRD-like class that contains TATTCC-containing promoter(*23*). GSC (Goosecoid) which is a homeobox gene and also contained TATTCC sequence in the motif analysis result (Figure 3.5B)(*24*). We scanned the motifs over H3.X and H3.Y gene regions and observed that GSC, DPRX, PITX1, and OTX2 shared common genomic binding sites, suggesting the potential genomic binding competition among these TFs (Figure S.3.3). Together, the results of motif analysis implied that LEUTX orchestrates with its related genes that drive the early embryonic program.

We observed that the upregulated genes by LEUTX overexpression are involved in embryo preimplantation pathway (Figure 5C, Figure S3.4A). However, these genes were not activated at all in the parental control with either H3.X or LEUTX overexpression indicating that DUX4 was required as an upstream initiator for their activation. Furthermore, depletion of LEUTX and overexpression of DUXA decreased expression of *ARGFX*, *DPPA3*, *TPRX1*, *ZSCAN4*, and *DDIT3* in FSHD2 and double mutant. In addition to

LEUTX overexpression, H3.X overexpression also upregulated some level of embryonic gene expression in double mutant (Figure 3.5C). We also compared our genes affected by LEUTX overexpression or depletion with genes induced by overexpression of LEUTX in human embryonic stem cells (hESCs) from Jouhilahti et al., 2016 (Figure S3.4B)(*15*). There were 105 overlapped DEGs in upregulated genes by LEUTX overexpression between double mutant and hESC while there were 161 overlapped DEGs in downregulated genes of FSHD2 by LEUTX depletion (Figure S3.4C). Our results altogether indicated that LEUTX activated by DUX4 continued to further regulate downstream embryonic genes in muscles, potentially disrupting myogenesis.

3.4 Discussion

LEUTX and H3XY effect on DUX4 network:

We previously found that H3.X/Y are one of the early DUX4 targets and that H3.X/Y and LEUTX (late target) cross-regulate their expression in FSHD cells. Our current analysis explored the DUX4 network with the depletion and/or overexpression of LEUTX (for the first time) and H3.X/Y in differentiated muscle cells in the context of FSHD. Importantly, both H3.Y and LEUTX were indicated to be primate-specific (*23*, *25*). Overexpression of neither H3.X nor LEUTX in parental control had any effect on DUX4 network, demonstrating the essential requirement of DUX4 for initiating the pathway (Figure 3.2A). In contrast, overexpression of H3.X/Y and LEUTX significantly upregulated DUX4 target genes in double mutant while their depletion significantly suppressed the network (Figures

3.2 and 3.3). Therefore, both H3.X/Y and LEUTX expression play significant roles in activation of the DUX4 gene network, possibly carrying out some of DUX4 biological roles.

Cell line variance from H3X.Y effect:

H3.X overexpression in our FSHD2 line failed to enhance DUX4 target genes further and depletion effect of H3.X/Y was also not obviously significant (Figures 3.1B and S3.2). H3.X/Y depletion were shown to reduce DUX4 target expression in FSHD1 and FSHD2 in the previous finding which contrasts with our FSHD2 line (*11*). This indicated that effect of H3.X/Y depletion could be variable and depends on other genetic or epigenetic factors that synergize with H3.X/Y. H3.X and H3.Y were discovered to be incorporated into DUX4 target gene regions and increased their expression via chromatin relaxation(*11*, *25*). However, constitutively expressed genes that also incorporate with H3.X and H3.Y are not affected by H3.X/Y depletion, suggesting the gene-specific role of H3.X/Y (*11*). H3.Y depletion was indicated to reduce gene expression involved in cell cycle (*25*). Yet, it is not clear how H3.X/Y specifically activated the DUX4 target genes in muscle cells though appeared to have little impact on global gene expression changes compared to LEUTX and DUXA (Figures 3.2C and D, Figures 3.3C and D). More studies are needed to examine the mechanism of gene regulation by H3.X/Y.

DUXA effect:

DUXA and DUX4 both belong to the PRD-like class family and have a very similar binding motif(*13*). Depletion of DUXA was shown to reduce other DUX4 target genes such as LEUTX and ZSCAN4(*13*). However, overexpression of DUXA did not increase DUX4 target genes unexpectedly but rather decreased them significantly (Figure 3.2A and B). DUXA was found to be expressed at 8-cell stage and downregulated a large number of genes in hESCs during early embryonic development(*26*). Consequently, DUXA might act as a repressor to DUX4 to set up a different network downstream of DUX4 though it is still possible that DUXA initially acts to positively feedback the DUX4 network similar to H3.X/Y and LEUTX. Therefore, we hypothesize that over a period of time during which DUX4 network activates such high expression of DUXA which turns to repress DUX4 instead, shutting down the DUX4 network and turning on a different program. DUXA must inactivate DUX4 at an appropriate time where other DUX4 targets such as LEUTX already accomplish its biological goals. Future studies should explore effect of DUXA in a dose-dependent manner to determine the turning point where DUXA negatively feedback the DUX4 network.

Embryonic genes induced by LEUTX:

Similar to DUXA, LEUTX also belongs to the PRD-like class family and is expressed from 4-cell to 8-cell stage (*15*). Both LEUTX and DUXA were suggested to have a key role in EGA but the exact mechanism has not been studied extensively. In this study, we showed that LEUTX is able to activate a set of genes that are not DUX4 targets in double mutant. Motif enrichment result showed that genes upregulated by LEUTX have promoter regions bound by other PRD-like class proteins (GSC, DPRX, OTX1, OTX2, DUX4) that are mostly

involved in early embryonic program (Figure 3.5B)(*15*, *17*, *26–28*). Particularly, LEUTX activated genes that are involved in embryo preimplantation (Figure 3.5C, Figure S3.3A). These genes cannot be expressed in the parent control line even with LEUTX or H3.X overexpression. This indicates that DUX4 still has a crucial role in initiating further downstream networks that are regulated by DUX4 target genes. LEUTX depletion of FSHD2 strongly decreased the expression of these genes, further confirming their expression by LEUTX regulation (Figure 3.5B). H3.X overexpression also upregulated expression of embryo preimplantation in double mutant though the effect is weaker than LEUTX. This suggested that H3.X/Y also cross-regulates with LEUTX to enhance non-DUX4 gene network. It is not clear whether downstream LEUTX network could impact muscle cell differentiation though we showed that LEUTX depletion in double mutant rescued myogenesis (Figure 3.4B). Future studies should examine LEUTX network to determine its effect on muscle differentiation. In conclusion, our study has provided more insight into LEUTX as one of the major DUX4 targets, and we also demonstrated that LEUTX cross-regulates with H3.X/Y in a feedback loop to orchestrate a program that partially resembles early embryonic development.

3.5 Method

3.5.1 Human myoblast culture and differentiation

Immortalized parental control, FSHD2 and double mutant myoblast cells were grown in high glucoseDMEM (Gibco) supplemented with 20% FBS (Omega Scientific, Inc.), 1%Pen-Strep (Gibco), and 2% Ultrasor G (Crescent Chemical Co.). Immortalization and single cell

clone isolation of primary myoblasts were performed as previously described for Control and FSHD2 myoblasts(*29*).Upon reaching 80% confluence, myoblast differentiation was induced by using high glucose DMEM medium supplemented with 2% FBS and ITS supplement (insulin 0.1%, 0.000067% sodium selenite, 0.055% transferrin; Invitrogen). Fresh differentiation media was changed every day.

3.5.2 Overexpression and Depletion transfection

The shRNA pLVshH3.X/Y (5'-GCGGGAAATCAGAAAGTAC-3', the siH3.X/Y targeting sequence in previous paper (*11*)), and corresponding pLVshControl (5'-CCTAAGGTTAAGTCGCCCTCG-3') were synthesized and cloned in pLV(shRNA)-Puro-U6 vector by Vector Builder. To construct the H3.X and LEUTX overexpression lentiviral plasmids pLVX_H3.X and pLVX_LEUTX, human H3.X and LEUTX ORF sequences were amplified from FSHD1 myotube cDNA using primer pairs H3.X_PCR and LEUTX_PCR respectively, and cloned into pLVX Lentiviral vector (Addgene, plasmid #135182). An empty pLVX vector was used as a negative control. Lentivirus packaging, transduction and puromycin selection were performed as previously described with slight modifications(*14*). Briefly myoblasts were infected twice at 48 hour and 24 hours prior to differentiation. The differentiated cells were harvested at days 5-7 right before cells started fusion to form myotubes.

3.5.3 RNA sequencing and processing

Total RNA was extracted by using the RNeasy kit (QIAGEN). Between 12-50 ng of RNA was converted to cDNA using the Smart-Seq 2 protocol (*13*). DNA libraries were constructed using Nextera DNA Flex Library Prep Kit (Illumina). DNA samples were sequenced on the Illumina NextSeq2000 platform using paired-end 43 bp mode with around 15 million reads per sample. RNA-seq raw reads were aligned with STAR (version 2.5.1b) using human genome reference hg38. Alignment default parameters were applied except with a maximum of 10 mismatches per pair, a ratio of mismatches to read length of 0.07, and a maximum of 10 multiple alignments. Read count was performed using RSEM (version 1.3) by defaults with gene annotations from GENCODE v28, and raw read counts were extracted for downstream analysis. Genes were filtered based on raw counts with at least 2 counts in at least 2 samples. Raw count was normalized by TMM in EdgeR and then converted to TPM (transcript per millions). Differential genes were calculated using the cut off *P*-values=< 0.01 and Log_2 Fold Change >=1.0.

3.5.4 WGCNA analysis

WGCNA analysis was done in R (Version 1.71) (*19*, *30*, *31*). Normalized TPM values from all of the samples in the study were used as input matrix for WGCNA. The following parameters were selected for the analysis: power = 9, deepSplit = 2, networkType = "signed", minModuleSize = 20, maxBlockSize = 4000, mergeCutHeight = 0.2. Five modules (Blue, Orange, Purple, Turquoise, and Gold) were chosen for further analysis. Gene list from each of the five modules was used as input for gene ontology analysis (GO).

3.5.5 Motif enrichment analysis

Lists of DEGs were selected for homer motif enrichment at the promoter regions (homer v4.11.1). Enrichment analysis was done by findMotifsGenome.pl with default parameter and motif database HOCOMOCO v11. Function scanMotifGenomeWide.pl was also used to determine the binding sites at genomic specific regions provided motif files.

3.5.6 Other bioinformatic tools

GO analysis was performed using EnrichR with the following databases: GO_Biological_Process_2018, KEGG_2019_Human, Reactome_2016, GO_Molecular_Function_2018, and GO_Cellular_Component_2018. GO results were plotted as bubble chart and hierarchical heatmaps were generated using ComplexHeatmap. Volcano plot and boxplot were done using ggplot2. Genome track was visualized using IGV software. Embryo implantation embryo pathway was extracted from cytospace.

3.7 Figure

Figure 3.1

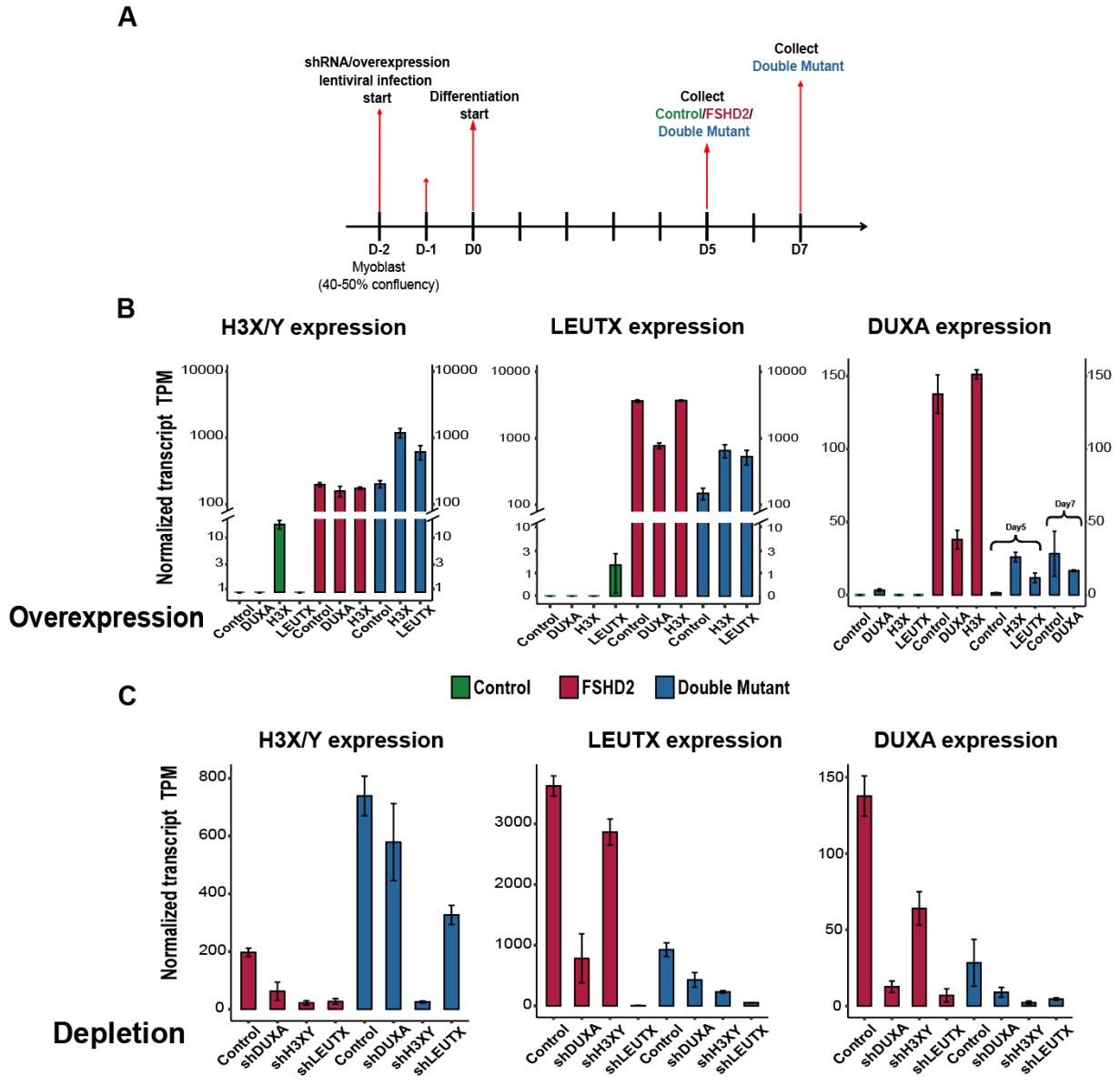


Figure 3.1: Effect of overexpression/depletion of H3.X/Y, LEUTX and DUXA

(A): Schematic diagram of overexpression and depletion by lentiviral vector and days of collection

(B): Bar plots of H3.X/Y, LEUTX, and DUXA expression under overexpression and control treatment in parental control, FSHD2 and double mutant from bulk RNA-seq analysis.

(C): Similar to (B), bar plots show the depletion and control treatment from bulk RNA-seq analysis

Figure 3.2

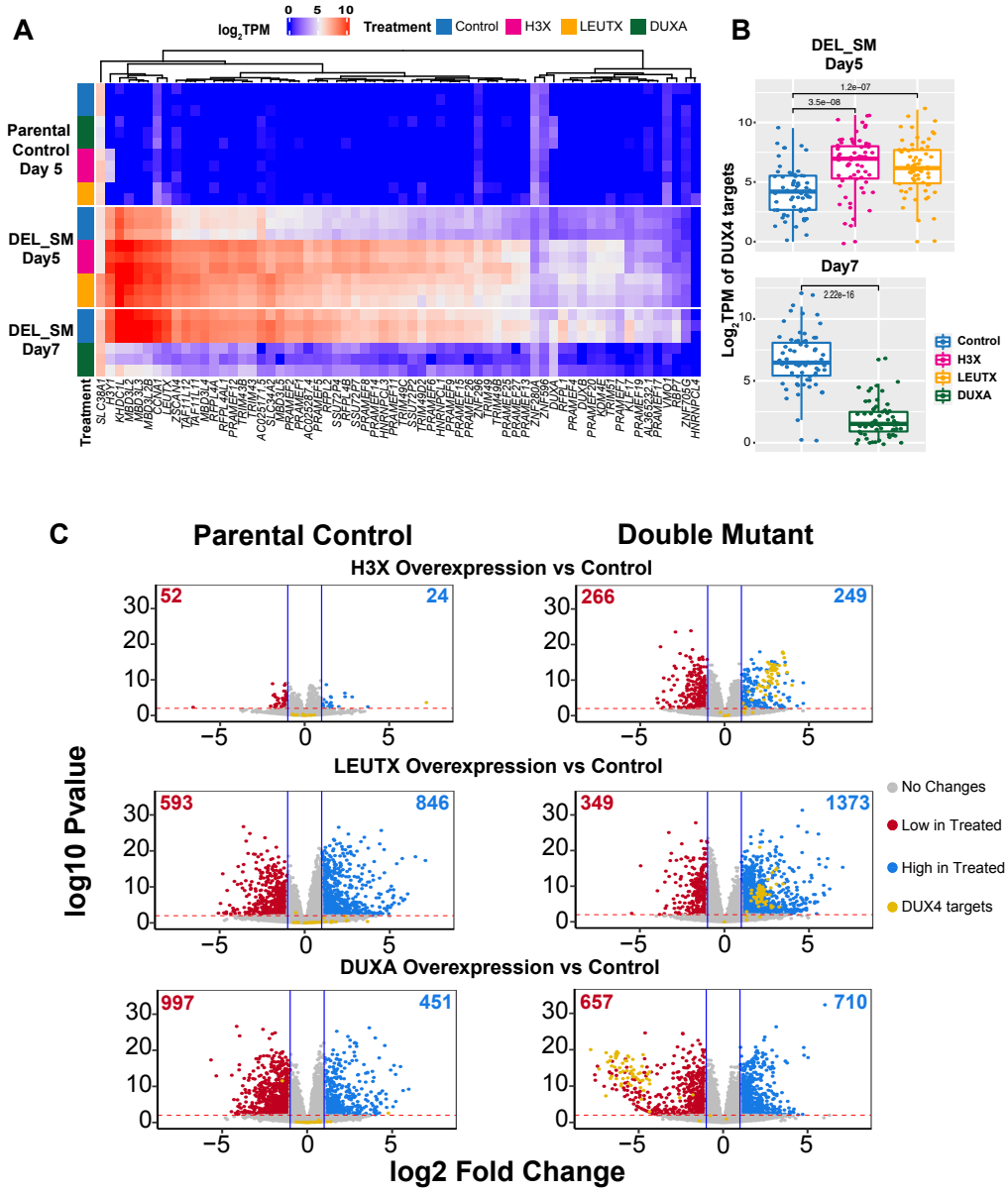

A
log2TPM
0 5 10
Treatment
Control
H3X
LEUTX
DUXA
Parental Control Day 5
DEL_SM Day5
DEL_SM Day7
Treatment
B
DEL_SM Day5
1.2e-07
3.5e-08
Day7
2.22e-16
Log2TPM of DUX4 targets
Control
H3X
LEUTX
DUXA
C
Parental Control
Double Mutant
H3X Overexpression vs Control
52
24
266
249
LEUTX Overexpression vs Control
593
846
349
1373
DUXA Overexpression vs Control
997
451
657
710
log10 Pvalue
log2 Fold Change
No Changes
Low in Treated
High in Treated
DUX4 targets

Figure 3.2: Differential gene expression by overexpression of DUX4 targets

(A): Hierarchically clustered heatmap showing the $\log_2$ normalized TPM of DUX4 target genes in parental control and double mutant. DUX4 target genes were selected based on previous papers (*13*, *18*). Colors indicate different overexpression treatment.

(B): Boxplot of DUX4 target genes from **(A)**, each dot represents an average of ($\log_2$ normalized TPM +1) of a DUX4 target across 3 replicates. Significant comparisons were indicated by Wilcoxon t-test.

(C): Volcano plots showing global DEGs by overexpression in parental control and double mutant comparing to control treatment. DUX4 target genes were indicated by orange.

Figure 3.3

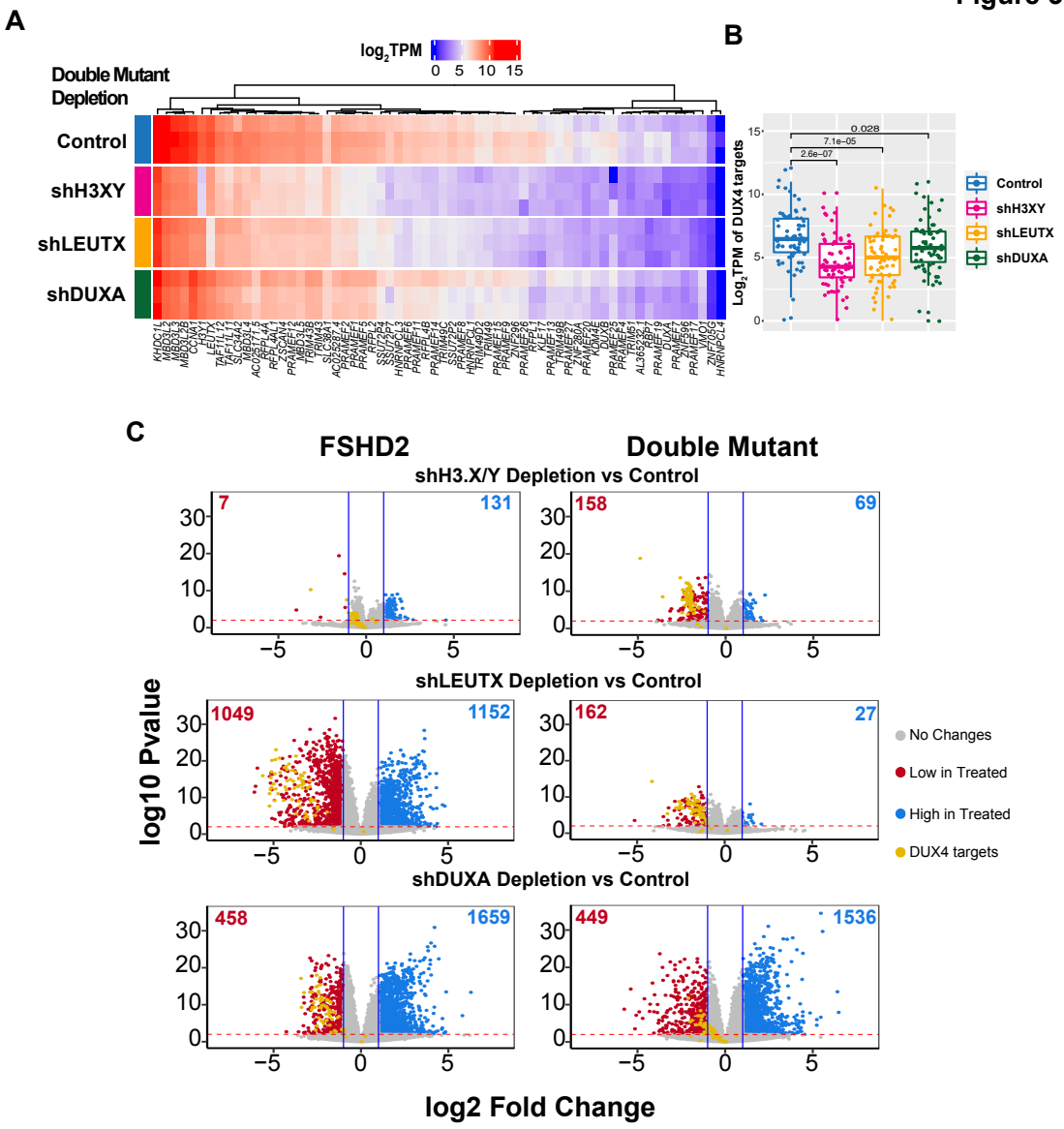

A
log2TPM
0 5 10 15
Double Mutant Depletion
Control
shH3XY
shLEUTX
shDUXA
B
Log2TPM of DUX4 targets
0.028
7.1e−05
2.6e−07
Control
shH3XY
shLEUTX
shDUXA
C
FSHD2
Double Mutant
shH3.X/Y Depletion vs Control
7
131
158
69
shLEUTX Depletion vs Control
1049
1152
162
27
shDUXA Depletion vs Control
458
1659
449
1536
log10 Pvalue
log2 Fold Change
No Changes
Low in Treated
High in Treated
DUX4 targets

Figure 3.3: Differential gene expression by depletion of DUX4 targets

(A): Similar to **Figure 3.2A**, hierarchically clustered heatmap showing the log_2 normalized TPM of DUX4 target genes in double mutant. Colors indicate different shRNA treatment.

(B): Boxplot of DUX4 target genes from **(A)**, each dot represents an average of (log_2 normalized TPM +1) of a DUX4 target across 3 replicates. Significant comparisons were indicated by Wilcoxon t-test.

(C): Volcano plots showing global DEGs by shRNA depletion in FSHD2 and double mutant comparing to control treatment. DUX4 target genes were indicated by orange.

Figure 3.4

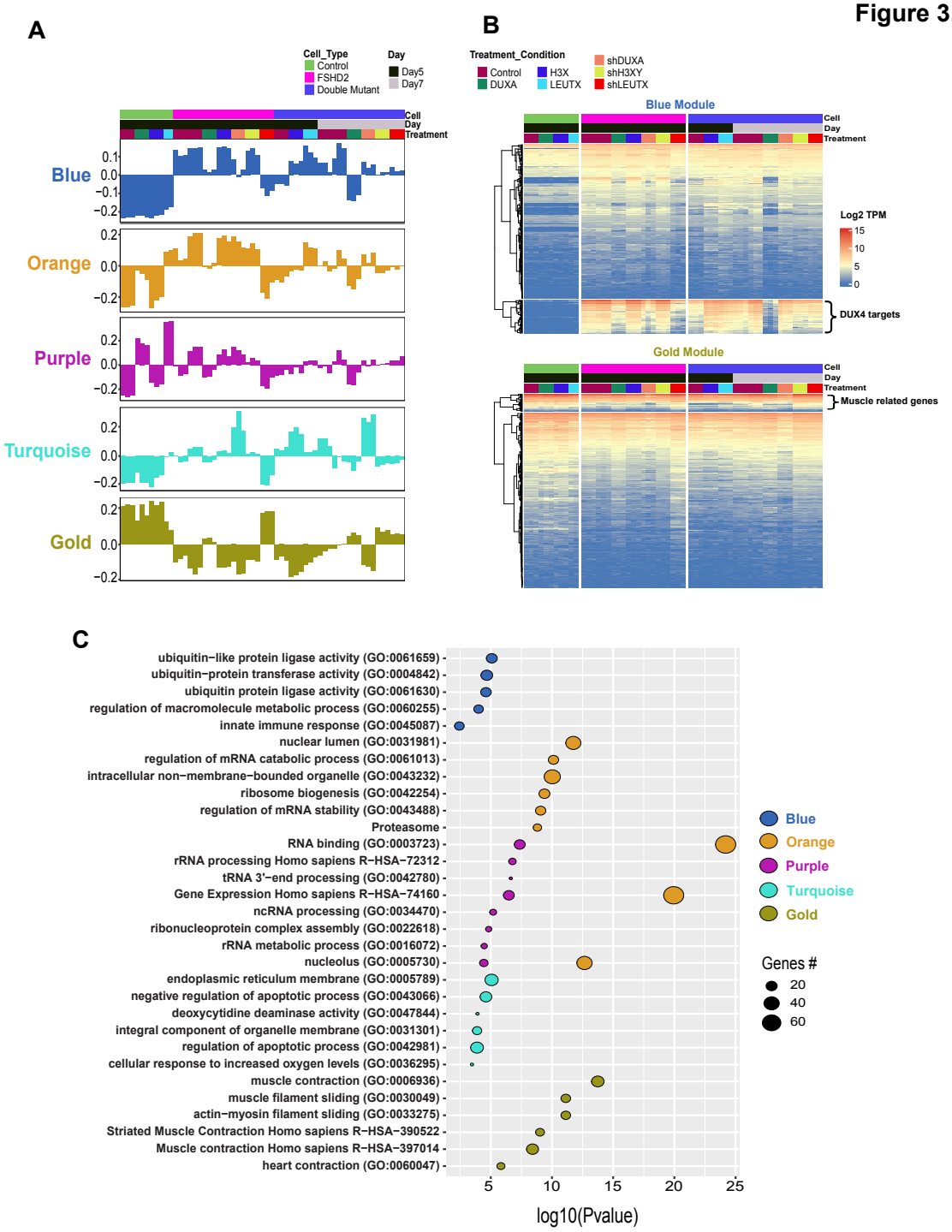

A
Cell_Type
Control
FSHD2
Double Mutant
Day
Day5
Day7
Cell
Day
Treatment
Blue
Orange
Purple
Turquoise
Gold
B
Treatment_Condition
Control
DUXA
H3X
LEUTX
shDUXA
shH3XY
shLEUTX
Blue Module
Log2 TPM
DUX4 targets
Gold Module
Muscle related genes
C
ubiquitin−like protein ligase activity (GO:0061659)
ubiquitin−protein transferase activity (GO:0004842)
ubiquitin protein ligase activity (GO:0061630)
regulation of macromolecule metabolic process (GO:0060255)
innate immune response (GO:0045087)
nuclear lumen (GO:0031981)
regulation of mRNA catabolic process (GO:0061013)
intracellular non−membrane−bounded organelle (GO:0043232)
ribosome biogenesis (GO:0042254)
regulation of mRNA stability (GO:0043488)
Proteasome
RNA binding (GO:0003723)
rRNA processing Homo sapiens R−HSA−72312
tRNA 3'−end processing (GO:0042780)
Gene Expression Homo sapiens R−HSA−74160
ncRNA processing (GO:0034470)
ribonucleoprotein complex assembly (GO:0022618)
rRNA metabolic process (GO:0016072)
nucleolus (GO:0005730)
endoplasmic reticulum membrane (GO:0005789)
negative regulation of apoptotic process (GO:0043066)
deoxycytidine deaminase activity (GO:0047844)
integral component of organelle membrane (GO:0031301)
regulation of apoptotic process (GO:0042981)
cellular response to increased oxygen levels (GO:0036295)
muscle contraction (GO:0006936)
muscle filament sliding (GO:0030049)
actin−myosin filament sliding (GO:0033275)
Striated Muscle Contraction Homo sapiens R−HSA−390522
Muscle contraction Homo sapiens R−HSA−397014
heart contraction (GO:0060047)
log10(Pvalue)
Blue
Orange
Purple
Turquoise
Gold
Genes #

Figure 3.4: WGCNA analysis

(A): 5 clusters/modules derived from WGCNA analysis. Y-axis shows eigenvalues across all samples indicated by the colored annotation.

(B): Hierarchically clustered heatmap from blue and gold module in $\log_2$ normalized TPM values

(C): Gene ontology analysis of all the genes in each module from WGCNA results

Figure 3.5

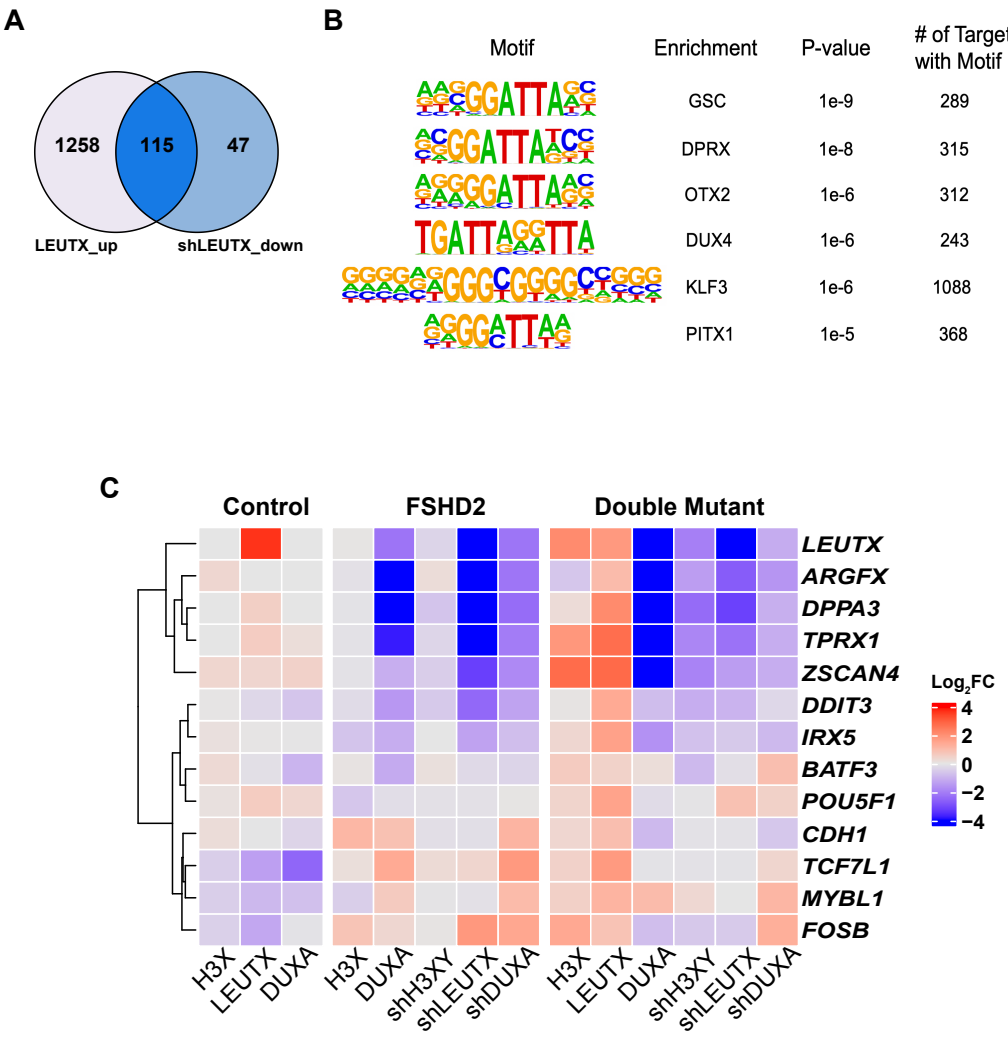

A
1258
115
47
LEUTX_up
shLEUTX_down
B
Motif
Enrichment
P-value
# of Targets with Motif
GSC
1e-9
289
DPRX
1e-8
315
OTX2
1e-6
312
DUX4
1e-6
243
KLF3
1e-6
1088
PITX1
1e-5
368
C
Control
FSHD2
Double Mutant
LEUTX
ARGFX
DPPA3
TPRX1
ZSCAN4
DDIT3
IRX5
BATF3
POU5F1
CDH1
TCF7L1
MYBL1
FOSB
H3X
LEUTX
DUXA
H3X
DUXA
shH3XY
shLEUTX
shDUXA
H3X
LEUTX
DUXA
shH3XY
shLEUTX
shDUXA
Log2FC
4
2
0
−2
−4

Figure 3.5: LEUTX upregulated a set of genes involved in embryo implantation

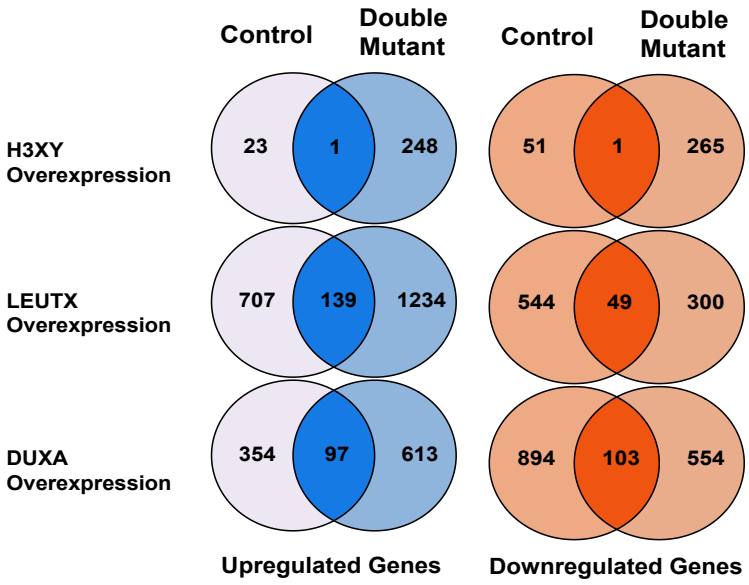
(A): Venn diagram shows overlapped upregulated genes from LEUTX overexpression and depletion

(B): Homer motif enrichment analysis at the promoters of genes upregulated by LEUTX overexpression

(C): Heatmap of $\log_2$ fold change (normalized TPM) of genes involved in embryo implantation

Figure S3.1

A



B

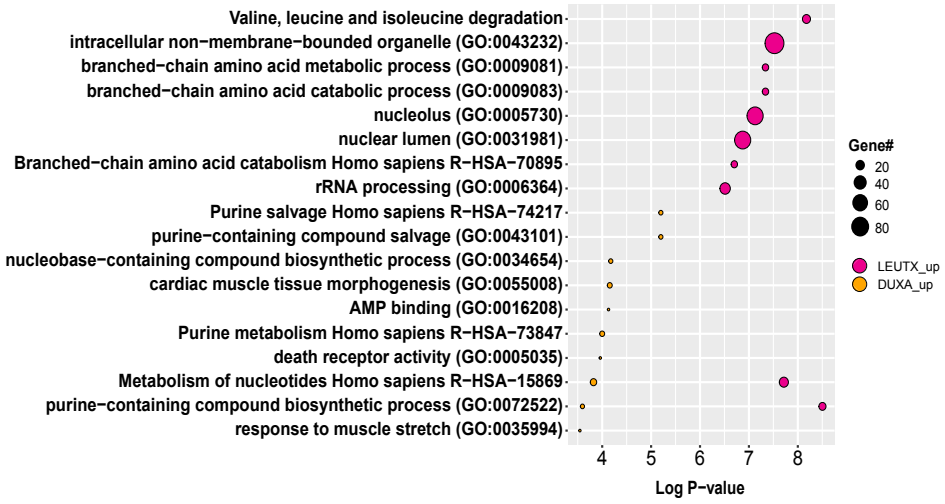


Figure S3.1: Comparison of DUX4 target overexpression

(A): Venn diagram shows overlapped DEGs from overexpression of H3.X/Y, LEUTX and DUXA in parental control and double mutant

(B): Gene ontology analysis showed unique/overlapped terms by genes upregulated by LEUTX and DUXA overexpression

Figure S3.2

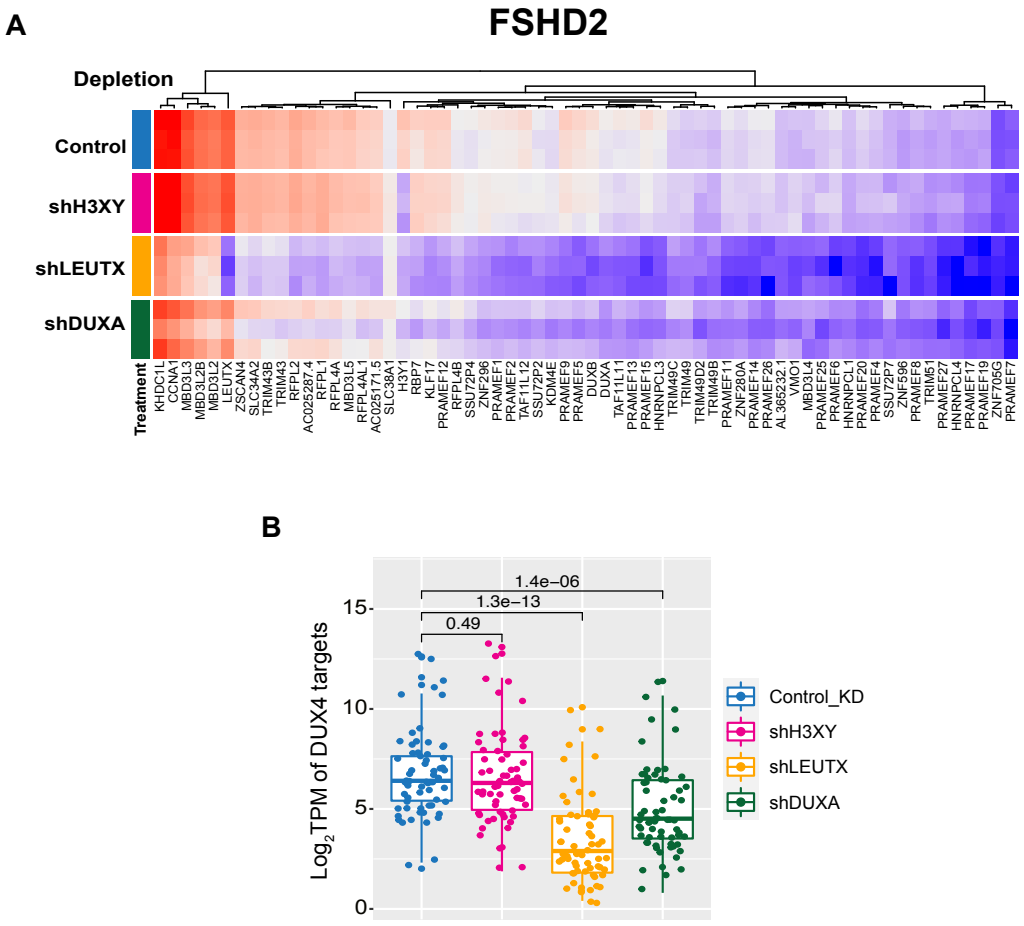

A
FSHD2
Depletion
Control
shH3XY
shLEUTX
shDUXA
Treatment
KHDC1L
CCNA1
MBD3L3
MBD3L2B
MBD3L2
LEUTX
ZSCAN4
SLC34A2
TRIM43B
TRIM43
RFPL2
AC025287.4
RFPL1
RFPL4A
MBD3L5
RFPL4AL1
AC025171.5
SLC38A1
H3Y1
RBP7
KLF17
PRAMEF12
RFPL4B
SSU72P4
ZNF296
PRAMEF1
PRAMEF2
TAF11L12
SSU72P2
KDM4E
PRAMEF9
PRAMEF5
DUXB
DUXA
TAF11L11
PRAMEF13
PRAMEF15
HNRNPCL3
TRIM49C
TRIM49
TRIM49D2
TRIM49B
PRAMEF11
ZNF280A
PRAMEF14
PRAMEF26
AL365232.1
VN0O1
MBD3L4
PRAMEF25
PRAMEF6
HNRNPCL1
PRAMEF20
PRAMEF4
SSU72P7
ZNF596
PRAMEF8
TRIM51
PRAMEF27
HNRNPCL4
PRAMEF17
PRAMEF19
ZNF705G
PRAMEF7
B
1.4e−06
1.3e−13
0.49
15
10
5
0
Log2 TPM of DUX4 targets
Control_KD
shH3XY
shLEUTX
shDUXA

Figure S3.2: Differential gene expression by overexpression of DUX4 targets in FSHD2

(A): Hierarchically clustered heatmap showing the $\log_2$ normalized TPM of DUX4 target genes in FSHD2. Colors indicate different shRNA treatment.

(B): Boxplot of DUX4 target genes from **(A)**, each dot represents an average of ($\log_2$ normalized TPM +1) of a DUX4 target across 3 replicates. Significant comparisons were indicated by Wilcoxon t-test.

Figure S3.3

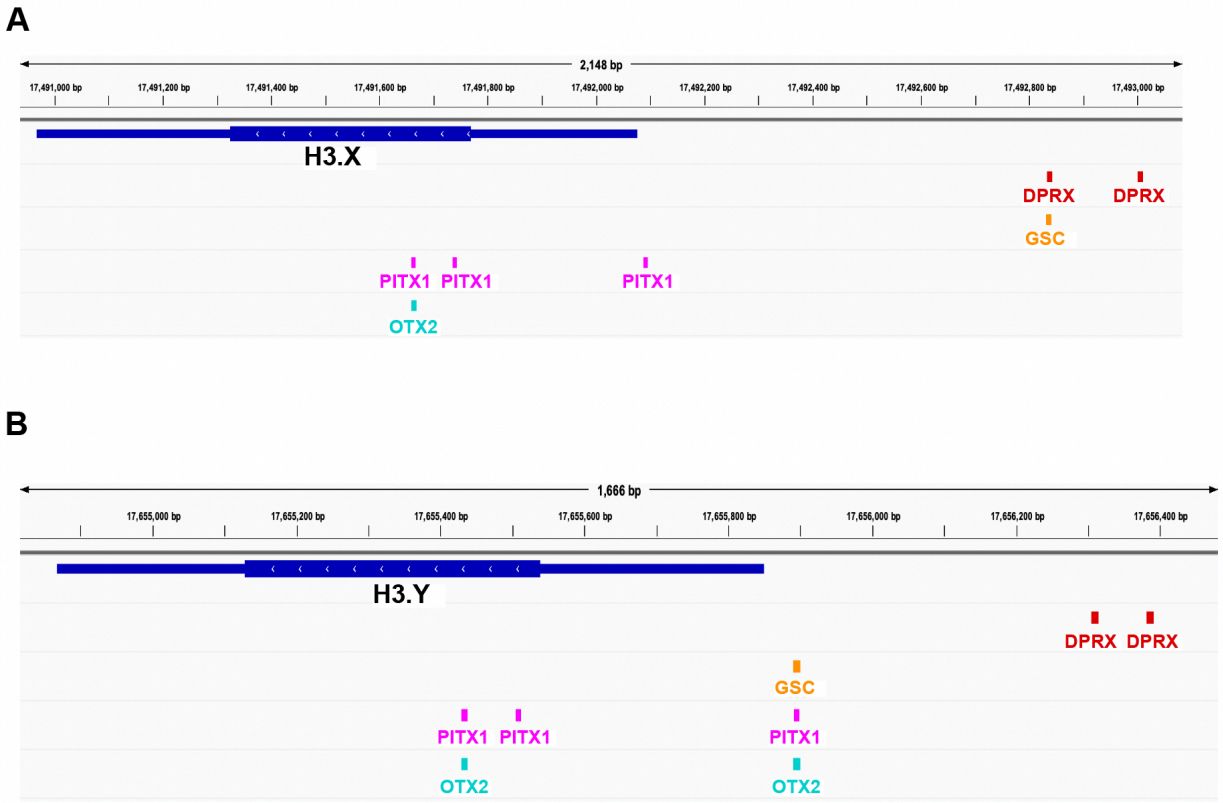

A
2,148 bp
17,491,000 bp
17,491,200 bp
17,491,400 bp
17,491,600 bp
17,491,800 bp
17,492,000 bp
17,492,200 bp
17,492,400 bp
17,492,600 bp
17,492,800 bp
17,493,000 bp
H3.X
DPRX
DPRX
GSC
PITX1
PITX1
PITX1
OTX2
B
1,666 bp
17,655,000 bp
17,655,200 bp
17,655,400 bp
17,655,600 bp
17,655,800 bp
17,656,000 bp
17,656,200 bp
17,656,400 bp
H3.Y
DPRX
DPRX
GSC
PITX1
PITX1
PITX1
OTX2
OTX2

Figure S3.3: Transcription factors binding at H3.X and H3.Y

(A): Motif scanning result from homer at H3.X gene region with the indicated motifs by colors.

(B): Similar to **(A)** shown for H3.Y gene region instead

Figure S3.4

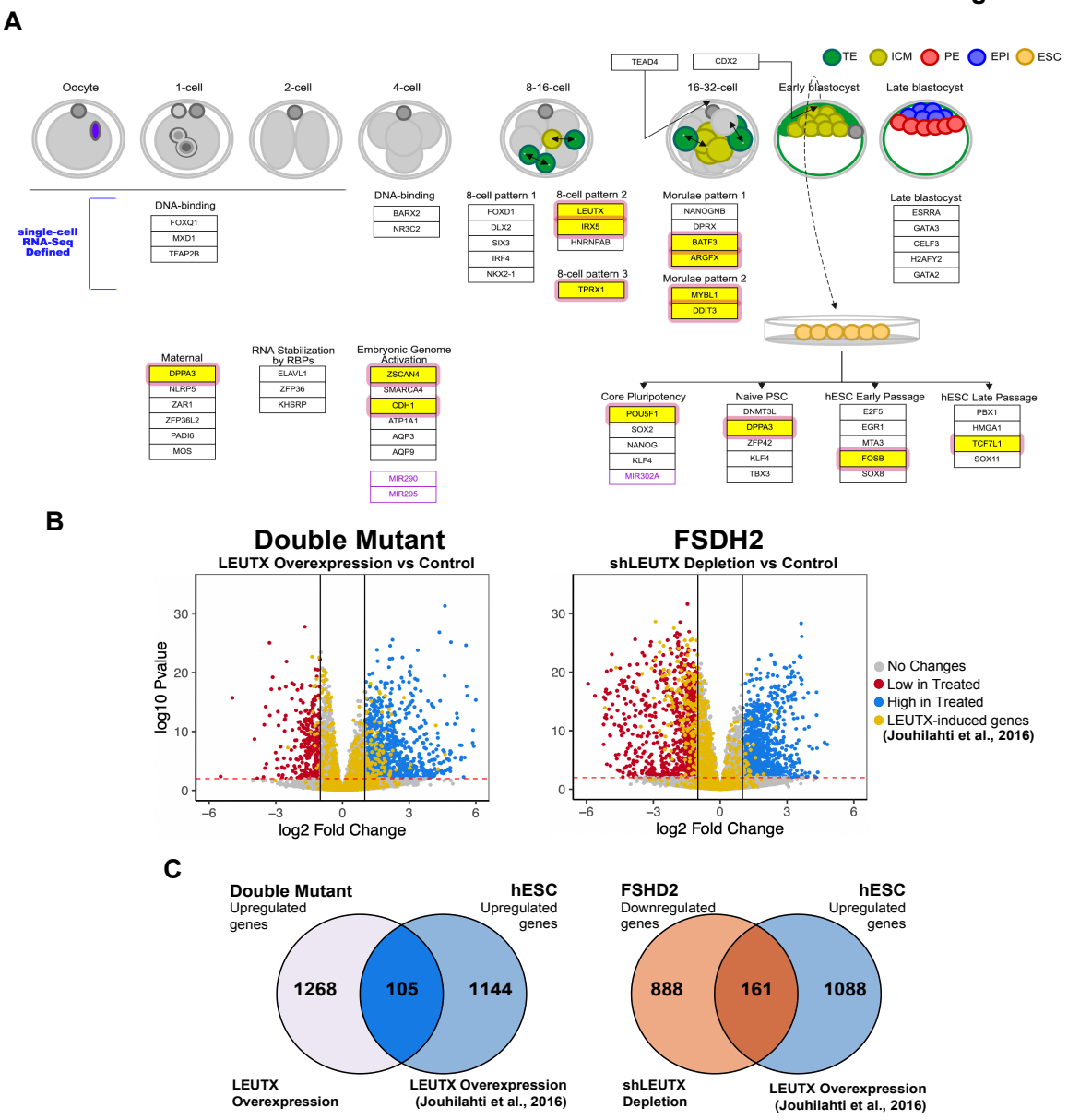

A
TE
ICM
PE
EPI
ESC
Oocyte
1-cell
2-cell
4-cell
8-16-cell
16-32-cell
Early blastocyst
Late blastocyst
TEAD4
CDX2
single-cell RNA-Seq Defined
DNA-binding
FOXQ1
MXD1
TFAP2B
DNA-binding
BARX2
NR3C2
8-cell pattern 1
FOXD1
DLX2
SIX3
IRF4
NKX2-1
8-cell pattern 2
LEUTX
IRX5
HNRNPAB
8-cell pattern 3
TPRX1
Morulae pattern 1
NANOGNB
DPRX
BATF3
ARGFX
Morulae pattern 2
MYBL1
DDIT3
Late blastocyst
ESRRA
GATA3
CELF3
H2AFY2
GATA2
Maternal
DPPA3
NLRP5
ZAR1
ZFP36L2
PADI6
MOS
RNA Stabilization by RBPs
ELAVL1
ZFP36
KHSRP
Embryonic Genome Activation
ZSCAN4
SMARCA4
CDH1
ATP1A1
AQP3
AQP9
MIR290
MIR295
Core Pluripotency
POU5F1
SOX2
NANOG
KLF4
MIR302A
Naive PSC
DNMT3L
DPPA3
ZFP42
KLF4
TBX3
hESC Early Passage
E2F5
EGR1
MTA3
FOSB
SOX8
hESC Late Passage
PBX1
HMGA1
TCF7L1
SOX11
B
Double Mutant
LEUTX Overexpression vs Control
FSDH2
shLEUTX Depletion vs Control
log10 Pvalue
log2 Fold Change
No Changes
Low in Treated
High in Treated
LEUTX-induced genes
(Jouhilahti et al., 2016)
C
Double Mutant
Upregulated genes
hESC
Upregulated genes
1268
105
1144
LEUTX Overexpression
LEUTX Overexpression
(Jouhilahti et al., 2016)
FSHD2
Downregulated genes
hESC
Upregulated genes
888
161
1088
shLEUTX Depletion
LEUTX Overexpression
(Jouhilahti et al., 2016)

Figure S3.4: LEUTX activated a set of embryo preimplantation genes

(A): Pathway of embryo preimplantation with highlighted genes indicated their upregulation by LEUTX

(B): Volcano plots showiing global DEGs by LEUTX overexpression in double mutant (left) and LEUTX depletion in FSHD2 (right) comparing to control treatment. LEUTX-induced genes in Jouhilahti et al, 2016 indicated by orange color

(C): Venn diagram showing either genes upregulated by LEUTX overexpression in double mutant (left) or genes downregulated by LEUTX depletion in FSHD2 (right) overlaps with genes upregulated by LEUTX overexpression in hESCs from Jouhilahti et al, 2016

Table S3.1: Summary of cell lines, treatment, day of differentiation of the study

Cell line	Rep	Day	Cell Type	Gene Target	Knockdown/Overexpression
B0	1	5	Control	Control_Puro	Control_Expression
B0	2	5	Control	Control_Puro	Control_Expression
B0	3	5	Control	Control_Puro	Control_Expression
B0	1	5	Control	LEUTX	Overpression
B0	2	5	Control	LEUTX	Overpression
B0	1	5	Control	H3X	Overpression
B0	2	5	Control	H3X	Overpression
B0	3	5	Control	H3X	Overpression
B0	1	5	Control	DUXA	Overpression
B0	2	5	Control	DUXA	Overpression
B0	3	5	Control	DUXA	Overpression
B2	1	5	FSHD2	shControl	shControl
B2	2	5	FSHD2	shControl	shControl
B2	3	5	FSHD2	shControl	shControl
B2	1	5	FSHD2	LEUTX	Knockdown
B2	2	5	FSHD2	LEUTX	Knockdown
B2	3	5	FSHD2	LEUTX	Knockdown
B2	1	5	FSHD2	H3XY	Knockdown
B2	2	5	FSHD2	H3XY	Knockdown
B2	3	5	FSHD2	H3XY	Knockdown
B2	1	5	FSHD2	DUXA	Knockdown
B2	2	5	FSHD2	DUXA	Knockdown
B2	3	5	FSHD2	DUXA	Knockdown
B2	1	5	FSHD2	Control_Puro	Control_Expression

B2	2	5	FSHD2	Control_Puro	Control_Expression
B2	3	5	FSHD2	Control_Puro	Control_Expression
B2	1	5	FSHD2	DUXA	Overpression
B2	2	5	FSHD2	DUXA	Overpression
B2	3	5	FSHD2	DUXA	Overpression
B2	1	5	FSHD2	H3X	Overpression
B2	2	5	FSHD2	H3X	Overpression
B2	3	5	FSHD2	H3X	Overpression
DEL4_SM_A	1	5	Double Mutant	Control_Puro	Control_Expression
DEL4_SM_A	2	5	Double Mutant	Control_Puro	Control_Expression
DEL4_SM_A	3	5	Double Mutant	Control_Puro	Control_Expression
DEL4_SM_A	1	5	Double Mutant	LEUTX	Overpression
DEL4_SM_A	2	5	Double Mutant	LEUTX	Overpression
DEL4_SM_A	3	5	Double Mutant	LEUTX	Overpression
DEL4_SM_A	1	5	Double Mutant	H3X	Overpression
DEL4_SM_A	2	5	Double Mutant	H3X	Overpression
DEL4_SM_A	3	5	Double Mutant	H3X	Overpression
DEL4_SM_A	1	7	Double Mutant	shControl	shControl
DEL4_SM_A	2	7	Double Mutant	shControl	shControl

DEL4_SM_A	3	7	Double Mutant	shControl	shControl
DEL4_SM_A	1	7	Double Mutant	LEUTX	Knockdown
DEL4_SM_A	2	7	Double Mutant	LEUTX	Knockdown
DEL4_SM_A	3	7	Double Mutant	LEUTX	Knockdown
DEL4_SM_A	1	7	Double Mutant	H3XY	Knockdown
DEL4_SM_A	2	7	Double Mutant	H3XY	Knockdown
DEL4_SM_A	3	7	Double Mutant	H3XY	Knockdown
DEL4_SM_A	1	7	Double Mutant	DUXA	Knockdown
DEL4_SM_A	2	7	Double Mutant	DUXA	Knockdown
DEL4_SM_A	3	7	Double Mutant	DUXA	Knockdown
DEL4_SM_A	1	7	Double Mutant	Control_Puro	Control_Expression
DEL4_SM_A	2	7	Double Mutant	Control_Puro	Control_Expression
DEL4_SM_A	3	7	Double Mutant	Control_Puro	Control_Expression
DEL4_SM_A	1	7	Double Mutant	DUXA	Overpression
DEL4_SM_A	2	7	Double Mutant	DUXA	Overpression
DEL4_SM_A	3	7	Double Mutant	DUXA	Overpression

Table S3.2: Gene list of each module from WGCNA analysis

Blue Module Genes						
AC000372.1	AL138955.1	DERL1	IBA57	NAA38	RAB20	TET2.AS1
AC003092.1	AL161421.1	DPPA3	IFT22	NAPA	RASIP1	THAP1
AC004540.1	AL161716.1	DSP	IL23A	NCOA7	RBBP6	TMEM160
AC004908.2	AL162231.4	DTX2	ILKAP	NEPRO	RBMXL1	TMEM222
AC004917.1	AL163051.1	DUXB	KCNK15.AS1	NLRP2	REEP6	TPBG
AC005083.1	AL355075.3	DYNAP	KDM4E	NOX4	RFFL	TPPP
AC005280.2	AL365214.2	ELN	KDM4F	NPIPB11	RFPL2	TPRX1
AC007250.1	AL365232.1	ELP6	KDM5A	NPIPB12	RFPL4A	TRIB3
AC008870.5	AL731684.1	ENSG00000276182	KHDC1L	NT5C1B.RDH14	RFPL4AL1	TRIM43
AC011990.1	ALOXE3	EXTL2	KLHL32	PELO	RFPL4B	TRIM43B
AC012676.3	ALPG	FAM122A	KMT2E.AS1	PGBD2	RND1	TRIM48
AC013400.1	ANKFN1	FAM166B	LEUTX	PIAS3	RNF20	TRIM49
AC016629.2	ANKRD54	FAM200B	LGMN	PLRG1	RRP8	TRIM49B
AC020915.1	ANXA2R	FAM90A1	LINC00115	PNMA1	RSBN1L	TRIM49C
AC021242.2	AP000459.2	FBXL13	LINC00159	PNMA2	SF3B4	TRIM49D1
AC022001.3	AP001042.1	FCRLA	LINC00327	PNP	SLC25A13	TRIM49D2
AC023157.3	AP005482.2	FGF14.AS2	LINC00423	POP5	SLC25A16	TRIM51
AC025171.5	APH1B	FRG2C	LINC00456	PPP1R21	SLC25A38	TRIM64
AC025176.1	APTR	FUT8.AS1	LINC00958	PRAMEF1	SLC34A2	TRIM64B
AC025287.4	ARHGAP29.AS1	GATA6	LINC00973	PRAMEF10	SLC35B1	TRNAU1AP
AC025419.1	ART3	GCNT1	LINC01088	PRAMEF11	SLC35E4	TRPC5OS
AC034213.1	ASB3	GNA14	LINC01283	PRAMEF12	SLC3A1	TRPC6
AC044849.1	ASH1L.AS1	GPBAR1	LINC01619	PRAMEF13	SLC49A4	TTI2
AC068506.1	ASPHD2	GPR161	LINC01828	PRAMEF14	SLC7A9	UBE3D
AC083799.1	BARX1	GRAMD1C	LINC02177	PRAMEF15	SMG6	UBTFL1
AC090692.1	BORCS8	GRK5	LINC02575	PRAMEF17	SNRNP25	UBXN10.AS1
AC092650.1	C12orf50	GSC	LINC02608	PRAMEF18	SOCS4	UBXN8
AC093484.3	C4orf17	GSTCD	LINC02623	PRAMEF19	SPECC1L	UFD1
AC096531.2	CALCOCO2	GTF2B	LINC02861	PRAMEF2	SRFBP1	USP29
AC103923.1	CASC19	GTF2F1	LRRC38	PRAMEF20	SRSF8	UTP25
AC104126.1	CCDC148	GTPBP8	LRRC42	PRAMEF25	SSU72P2	ZDHHC14

AC106712.2	CCL2	GUK1	MARCHF5	PRAMEF26	SSU72P3	ZHX1
AC123595.1	CCL26	H2AC19	MBD3L2	PRAMEF27	SSU72P4	ZNF17
AC132872.1	CCNA1	H3Y1	MBD3L2B	PRAMEF33	SSU72P5	ZNF217
AC139530.1	CD53	H4C15	MBD3L3	PRAMEF4	SSU72P7	ZNF235
AC243829.2	CD7	HBA2	MBD3L4	PRAMEF5	SSU72P8	ZNF296
AC243960.3	CDK2AP2	HCFC2	MBD3L5	PRAMEF6	STIM2	ZNF408
ACSM5	CHAC1	HDC	MBIP	PRAMEF7	STX6	ZNF596
ACTR10	CHIC2	HEY2	METTL14	PRAMEF8	SVBP	ZNF668
ACTR5	CLU	HHLA3.AS1	MFSD11	PRAMEF9	SYNJ2BP.COX16	ZNF705A
ADAMTSL3	COPG2	HNRNPCL1	MFSD4B	PRR3	TAF11L11	ZNF705E
AKR1B10	CPHXL	HNRNPCL2	MINDY3	PRRG4	TAF11L12	ZNHIT6
AKR1D1	CRTC2	HNRNPCL3	MTA3	PTRHD1	TAF11L13	ZSCAN4
AL023803.3	CYP2C18	HSPA1A	MTHFD2	PUS3	TAF11L14	ZSCAN5B
AL121658.1	DEDD	HSPA1B	MYC	R3HCC1L	TBC1D25	ZSCAN5C
AL132639.3	DEFB119	HUS1B	MYCT1	RAB11B	TBCB	
Orange Module Genes						
AAR2	CDKN2AIP	FAM104A	MRPL27	PRTG	SGTA	TXLNG
AASDH	CLDN4	FAM117B	MRPL35	PSMA1	SLC25A14	TXNL4B
AATF	COPS5	FAM222B	MRPL4	PSMC3	SMYD5	TYW1
ABI3	CPSF4	FLAD1	MRPS11	PSMC4	SNHG11	UBE2L3
ABT1	CSTF1	GABRA3	MSS51	PSMC5	SS18	UBE2N
AC004151.1	CSTF2	GADD45A	MYO5B	PSMD1	SSBP1	UBIAD1
AC005632.3	CTBP2	GEMIN5	NDUFAF7	PSMD13	STOML2	USP45
AC016542.1	CTNNBL1	GOSR1	NEK10	PSMD6	SUGT1	UTP18
AC023043.4	CXorf40B	GPATCH2	NFYA	PSME3	SUMO1	UTP3
AC025171.1	DAP3	GTF2F2	NOL10	PTPN2	SUPV3L1	VCP
AC087501.4	DCP1A	GTF2H1	NOL12	PWP1	SURF2	VTA1
AC106028.3	DDX20	GTF2H2C	NPIPB5	RARS1	SURF6	WDR75
AC132192.2	DDX27	H4.16	NR2C1	RBIS	SYF2	YAE1
ACKR4	DDX28	HACL1	NRBP1	RBM39	TARBP2	YTHDF2
ADCK1	DDX31	HARS1	NUP50.DT	RBM4	TAS2R14	YY2
ADSL	DDX49	HKDC1	NUP88	RBM48	TBCC	ZGPAT
AL590399.1	DDX50	HSPD1	NVL	RBMX2	TEN1	ZNF14
ANKHD1.EIF4EBP3	DDX55	INTS4	NXF1	RCL1	TERF1	ZNF174

AP003068.1	DNAJA1	ISG20L2	ORMDL1	RNF113A	TIGAR	ZNF215
AP4B1	DNAJB7	KIAA1614.AS1	OSGEP	RNF25	TMEM199	ZNF326
APEX1	DPM1	KPNA1	OTUB1	RPF1	TMEM44.AS1	ZNF35
ARID4A	EIF3B	KRBOX4	PARL	RPL7L1	TMEM69	ZNF547
ARMCX5	EIF3D	KRR1	PDE11A.1	RRN3	TOE1	ZNF559
B3GALNT2	EIF3I	LSG1	PEX3	RRP7A	TOM1L1	ZNF586
BBS5	EIF4A1	MAPK8IP2	PHAX	RTCB	TOP1	ZNF589
C12orf43	EIF4A3	MED10	POLG2	RUVBL2	TRA2A	ZNF597
CCAR1	ELOF1	MED7	POLR2F	SAFB2	TRIM27	ZNF790.AS1
CCDC84	ERVK3.1	MEPCE	POLR3H	SAMD12	TRIM28	ZW10
CCNC	ESAM	MID1	POMP	SEC23B	TRMT6	
CCT5	EXO5	MPDU1	POP1	SELENOT	TRMT61A	
CDC123	EXOSC4	MRPL19	PPWD1	SENP3.EIF4A1	TSFM	
CDK7	EXOSC6	MRPL20	PRPF31	SET	TTC19	
Purple Module Genes						
AAMP	CDC37	GNL3L	LTF	POLR1C	RRP1	TRMU
AC117503.3	DIMT1	GPATCH4	NARS1	PPME1	SCFD2	TRUB2
ACAA2	DNAJA3	GTF3C6	NLE1	PRMT5	SCO1	WDR12
AK2	ELAC2	HSD17B10	NOB1	PSMA4	SPRYD4	WDR77
ARMC6	FARSA	HSP90AB1	PARK7	PTDSS1	SSB	
BMS1	FMC1.LUC7L2	IPPK	PLIN3	PWP2	SYNCRIP	
C1orf109	FP565260.1	KDM5B	PNPLA3	RPS19BP1	TRAPPC4	
Turquoise Module Genes						
ABCF3	BRMS1	DPF2	HPSE	NMD3	RPN2	TMCO1
AC004922.1	BROX	DRAM1	HSPA5	NMT2	RTF2	TMEM115
AC008894.2	BSG	DUSP11	HTR1A	NQO1	SAAL1	TMEM138
AC009163.5	C11orf58	DUSP6	IFNAR1	NRP2	SARS1	TMEM165
AC011451.1	C11orf86	DZIP1	IL24	NUDT5	SCAMP3	TMEM39A
AC020929.1	C1orf35	EDEM2	ILK	OR2B6	SCAMP4	TMX1
AC037459.1	C5orf15	EEF1AKMT1	INTS11	ORMDL2	SCG5	TNKS2
AC092135.2	CALM2	EIF2B4	INTS14	OSR1	SDF2	TOR1A
AC096887.2	CARD16	EIF3G	ITPR3	P2RX7	SESN2	TOR1AIP2
AC099524.1	CASP8	EIF4EBP1	KLHL11	PABPC5.AS1	SHPK	TPGS2
AC115618.1	CCDC115	EIF6	L3HYPDH	PAFAH1B2	SLC1A5	TRIM32

ADI1	CCDC134	ERG28	LAT2	PARN	SLC25A39	TRIP4
ADPGK	CCN3	ERLIN1	LINC00525	PCSK5	SLC39A3	TSN
AKAP12	CD320	ERP44	LINC01586	PDCD10	SLC39A6	TTC26
AL024497.1	CDC26	ESM1	LINC02223	PDE9A	SLC3A2	TTC39C
AL049795.1	CDC42EP4	EVI2B	LINC02535	PIH1D1	SLC66A3	TTPAL
AL132780.2	CDCA4	FAF2	LPCAT2	PRCP	SMIM15	TWIST1
AL132857.2	CHKA	FAM111A.DT	LRPAP1	PRELID2	SMIM7	TXNDC9
ALDH2	COMMD1	FAM131C	LURAP1	PROCR	SNX1	UBE2Z
ALKBH3	COMMD4	FAM174C	MAEA	PRPS2	SNX6	UGDH
ANGPTL4	COPS6	FAM32A	MAP1B	PSEN1	SNX8	UHMK1
AP2S1	COX20	FAS	MAP4K2	PSMB1	SOAT1	UROD
AP3S1	CPNE3	FBXO8	MAPK3	PTPN9	SORD	VANGL1
AP5Z1	CPPED1	FUT10	MED27	PYCARD	SPA17	VKORC1
APOBEC3D	CRACR2A	GALNT4	MED28	PYGB	SPCS2	VMP1
APOBEC3F	CROT	GANAB	MED29	QTRT2	SPOUT1	VPS16
APOBEC3G	CRYBG1	GET3	MLF2	RAB8A	SPTLC1	VPS26A
APOL1	CYREN	GNAI3	MMP24	RAET1G	STAMBPL1	VPS4B
ARL6IP1	DCUN1D3	GNG11	MOXD1	RAP1B	STXBP5.AS1	VSIG10
ASH2L	DDX41	GPR137	MPZL1	RAPGEF3	SYAP1	WDR18
ATF6	DEGS1	GPR87	MRPL28	RBM7	SYS1	XYLB
ATG5	DGUOK	GPX1	MRPS18A	REEP3	TBPL1	YIPF5
ATG7	DNAI1	GSTO1	MTMR14	RER1	TCP11L1	ZBTB9
AUP1	DNAJB1	H3.3B	MTX1	RGS10	TIMM10B	ZCCHC4
BAG4	DNAJC3	HDAC1	MYDGF	RHBDF2	TLK1	ZNF816
BCL10	DNMBP	HDAC3	NCR3	RNF149	TM2D2	
BLZF1	DOLPP1	HFE	NDUFA12	RPE	TM4SF1	
BRK1	DPAGT1	HLA.B	NMB	RPGRIP1L	TM4SF18	
Gold Module Genes						
ABRA	APOBEC2	CXorf40A	HDHD5.AS1	MLLT3	PPP2R3B	SOX11
AC002398.2	ARL8A	CYC1	HERC3	MMP15	PPP2R3B_PARY	SPTB
AC004130.2	ARX	DCLK1	HHIPL1	MPC1	PRDX6	SRL
AC004816.2	ASB12	DDI1	HOTAIRM1	MPST	PRICKLE1	SRP68
AC007036.3	ATG4B	DDN	HOXC.AS1	MRLN	PRKAG2.AS1	ST8SIA2
AC008781.2	ATP1B4	DES	HRAS	MRPL41	PRKAR2A	STARD7

AC011603.4	ATP5F1D	DHRS7C	HSPB2.C11orf52	MRPS36	PROB1	STXBP6
AC012085.2	ATP5F1E	DOCK8	HSPB3	MS4A15	PRSS50	SYPL2
AC018926.3	ATP5MF	DOK5	HYAL3	MT3	PRSS56	TAL2
AC020907.1	ATP5PF	DPP3	ICAM4	MTUS2	PRXL2C	TARS3
AC020907.5	ATP5PO	DPYSL4	ICOSLG	MUSTN1	PSMA2	TATDN1
AC021087.4	BAG2	DRP2	IDH2	MYADM	PSMA7	TBC1D10C
AC022167.3	BAIAP2	DUSP28	IFFO2	MYADML2	PTGDS	TCAF2
AC026471.5	BAIAP3	EAF1.AS1	IL20RA	MYBPC1	PYGM	TCP10L
AC055822.1	BCAM	EEF1A2	INHBE	MYBPC2	RASGRP2	THOC7
AC061999.1	BEST3	EHBP1L1	IQANK1	MYH2	RASL10A	THPO
AC068658.1	BIN3	ENDOG	IRF5	MYH8	RBFOX1	TMC6
AC068792.1	BRI3	ENOX1	JPT1	MYL10	RBM20	TMEM117
AC073332.1	BTF3	ENSG00000235475	KATNB1	MYL2	RBM38	TMEM121B
AC079089.1	BX539320.1	ENSG00000263812	KBTBD13	MYL3	RBP1	TMEM220.AS1
AC079336.5	C16orf46	ENSG00000269946	KCNJ11	MYLK2	RELL1	TNFRSF21
AC090360.1	C1orf105	ENTPD2	KCNJ12	MYLPF	RFX3	TNNC1
AC090617.2	C1orf116	EPHB3	KIAA1671.AS1	MYOM1	RGS14	TNNC2
AC091488.1	C4orf47	EPM2A	KIF1C	MYOM2	RHCE	TNNI2
AC092687.3	C6orf118	EXOC6	KIF2A	MYOT	RHOBTB3	TNNT2
AC097375.3	CA7	FABP3	KLC1	MYOZ1	RIIAD1	TOMM7
AC105219.4	CACNG4	FAHD2B	KLF5	MYOZ2	RINL	TP53INP2
AC105942.1	CADM2	FAM180B	KLHDC3	NAMA	RIPOR2	TPTEP2.CSNK1E
AC106886.3	CALML3.AS1	FAM228B	KLHL30.AS1	NAT14	RNF103	TRAF4
AC107294.1	CAMK2B	FAM78A	KLHL40	NAT8L	RNF41	TRIB1
AC112236.1	CASQ2	FAM83H	LANCL1.AS1	NAV3	RPRML	TRIM45
AC112236.2	CASTOR2	FBXW12	LDB3	NDRG2	RTN4R	TRIM72
AC120498.1	CASTOR3	FGF13	LGI4	NDUFAF2	S100A7	TSLP
AC127029.2	CCDC113	FGF7	LINC00163	NDUFS3	SAMD4A	TSPAN13
AC131097.3	CCDC88C	FHL3	LINC00957	NDUFS7	SBDS	TSPAN7
AC132217.2	CD163	FHOD3	LINC01136	NDUFV3	SBK1	TTC24
AC136475.2	CDC37L1	FITM1	LINC01208	NES	SBK2	TUBA8
AC137834.1	CDHR1	FN3K	LINC01405	NEXN	SBK3	UBE2E1
ACAP3	CDK18	FP565260.3	LINC01438	NFIL3	SCN4A	UBQLN4
ACSS2	CERS1	FSD2	LINC01534	NKAIN1	SELENOW	USP13

ACTC1	CERT1	FUNDC2	LINC01714	NKAIN2	SEPTIN9	USP2.AS1
ACTR3B	CES4A	FXYD4	LINC01825	NMRK2	SGTB	UTS2R
ADAMTS9	CFD	FYN	LINC01827	NPM3	SIK1	VDAC1
ADAMTSL4	CFL2	FZD9	LINC02004	NR4A1	SIK1B	VPREB3
ADAMTSL5	CHN2	GAS2L2	LINC02641	NTF4	SLC16A4	VWA5A
AFAP1L1	CHPT1	GDF1	LRRC14B	P2RX5	SLC25A3	WIPF3
AFAP1L2	CHST10	GDF5	LRRC24	P2RY2	SLC29A2	WNT4
AGMAT	CKM	GID4	LRRC73	PABPC4	SLC35D2	XKRX
AKR7A3	CLDN5	GNLY	MACROD1	PACSIN3	SLC66A2	ZC3H8
AL023806.2	CNOT6L	GOLT1A	MAFA	PANO1	SLC6A12	ZFAT
AL096865.1	CNPY1	GPAT3	MAP1LC3C	PDE7A	SLC6A13	ZFHX4.AS1
AL109955.1	COA4	GPR157	MAP6D1	PDZRN3.AS1	SMDT1	ZNF106
AL158070.2	COL11A2	GPT	MCAM	PEBP4	SMIM2.IT1	ZNF30
AL158071.4	COX5A	GPX7	MEF2C	PHETA1	SMOC1	ZNRF1
AL499627.1	COX6A2	GRB14	MEG9	PITX2	SNAI3	
AL928596.1	CST6	GYG1	METTL8	PLA2G3	SNAI3.AS1	
AMOTL1	CTDNEP1	HACD1	MID1IP1	PLEKHA4	SNHG21	
AP002387.2	CTXN1	HADH	MINDY4	PLPP1	SNN	
AP003068.2	CXCL13	HAVCR2	MITF	PLXNB1	SNX20	

Table S3.3: TF enrichment of upregulated genes by LEUTX

Rank	Motif	Name	P-value	# of Target Sequences with Motif	% of Target Sequences with Motif
1	GSC	HUMAN.H11MO.0.D	1.00E-09	289	24.79%
2	DPRX	HUMAN.H11MO.0.D	1.00E-08	315	27.02%
3	ZN219	HUMAN.H11MO.0.D	1.00E-07	1046	89.71%
4	NFYA	HUMAN.H11MO.0.A	1.00E-07	496	42.54%
5	NFYC	HUMAN.H11MO.0.A	1.00E-07	520	44.60%
6	OTX2	HUMAN.H11MO.0.A	1.00E-06	312	26.76%
7	DUX4	HUMAN.H11MO.0.A	1.00E-06	243	20.84%
8	SUH	HUMAN.H11MO.0.A	1.00E-06	931	79.85%
9	SP1	HUMAN.H11MO.1.A	1.00E-06	1092	93.65%
10	KLF15	HUMAN.H11MO.0.A	1.00E-06	1084	92.97%
11	KLF12	HUMAN.H11MO.0.C	1.00E-06	1079	92.54%
12	KLF3	HUMAN.H11MO.0.B	1.00E-06	1088	93.31%
13	OTX1	HUMAN.H11MO.0.D	1.00E-05	283	24.27%
14	SP4	HUMAN.H11MO.1.A	1.00E-05	1078	92.45%
15	SP2	HUMAN.H11MO.0.A	1.00E-05	1096	94.00%
16	PITX1	HUMAN.H11MO.0.D	1.00E-05	368	31.56%
17	EGR2	HUMAN.H11MO.1.A	1.00E-05	1067	91.51%
18	EGR4	HUMAN.H11MO.0.D	1.00E-05	1081	92.71%
19	E2F5	HUMAN.H11MO.0.B	1.00E-05	882	75.64%
20	KLF4	HUMAN.H11MO.0.A	1.00E-05	1039	89.11%
21	EGR1	HUMAN.H11MO.0.A	1.00E-05	1069	91.68%
22	SP4	HUMAN.H11MO.0.A	1.00E-05	1082	92.80%

References:

1. S. Re, M. van der Maarel, R. R. Frants, "The D4Z4 Repeat-Mediated Pathogenesis of Facioscapulohumeral Muscular Dystrophy" (2005).

2. S. M. van der Maarel, R. Tawil, S. J. Tapscott, Facioscapulohumeral muscular dystrophy and DUX4: Breaking the silence. *Trends Mol Med*. **17** (2011), pp. 252–258.

3. J. L. Whiddon, A. T. Langford, C.-J. Wong, J. W. Zhong, S. J. Tapscott, Conservation and innovation in the DUX4-family gene network. *Nat Genet*. **49**, 935–940 (2017).

4. M. Ferreboeuf, V. Mariot, B. Bessières, A. Vasiljevic, T. Attié-Bitach, S. Collardeau, J. Morere, S. Roche, F. Magdinier, J. Robin-Ducellier, DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles, doi:10.1093/hmg/ddt409ï.

5. L. N. Geng, Z. Yao, L. Snider, A. P. Fong, J. N. Cech, J. M. Young, S. M. van der Maarel, W. L. Ruzzo, R. C. Gentleman, R. Tawil, S. J. Tapscott, DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev Cell*. **22**, 38–51 (2012).

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

7. P. G. Hendrickson, J. A. Doráis, E. J. Grow, J. L. Whiddon, J.-W. Lim, C. L. Wike, B. D. Weaver, C. Pflueger, B. R. Emery, A. L. Wilcox, D. A. Nix, C. M. Peterson, S. J. Tapscott, D. T. Carrell, B. R. Cairns, Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet*. **49**, 925–934 (2017).

8. F. Rahimov, O. D. King, D. G. Leung, G. M. Bibat, C. P. Emerson, L. M. Kunkel, K. R. Wagner, Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers. *Proc Natl Acad Sci U S A*. **109**, 16234–16239 (2012).

9. N. Broucqsault, J. Morere, M.-C. Gaillard, J. Dumonceaux, J. Torrents, E. Salort-Campana, A. Maues De Paula, M. Bartoli, C. Fernandez, A. L. Chesnais, M. Ferreboeuf, L. Sarda, H. Dufour, C. Desnuelle, S. Attarian, N. Levy, K. Nguyen, F. Magdinier, S. Roche, Dysregulation of 4q35- and muscle-specific genes in fetuses with a short D4Z4 array linked to facio-scapulo-humeral dystrophy. *Hum Mol Genet*. **22**, 4206–4214 (2013).

10. T. I. Jones, J. C. J. Chen, F. Rahimov, S. Homma, P. Arashiro, M. lou Beermann, O. D. King, J. B. Miller, L. M. Kunkel, C. P. Emerson, K. R. Wagner, P. L. Jones, 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).

11. R. Resnick, C. J. Wong, D. C. Hamm, S. R. Bennett, P. J. Skene, S. B. Hake, S. Henikoff, S. M. van der Maarel, S. J. Tapscott, DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. *Cell Rep*. **29**, 1812-1820.e5 (2019).

12. A. E. Campbell, S. C. Shadle, S. Jagannathan, J.-W. Lim, R. Resnick, R. Tawil, S. M. van der Maarel, S. J. Tapscott, NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *Elife*. **7**, e31023 (2018).

13. S. Jiang, K. Williams, X. Kong, W. Zeng, N. V. Nguyen, X. Ma, R. Tawil, K. Yokomori, A. Mortazavi, Single-nucleus RNA-seq identifies divergent populations of FSHD2 myotube nuclei. *PLoS Genet*. **16**, 1–26 (2020).

14. J. Chau, X. Kong, N. Viet Nguyen, K. Williams, M. Ball, R. Tawil, T. Kiyono, A. Mortazavi, K. Yokomori, Relationship of DUX4 and target gene expression in FSHD myocytes. *Hum Mutat*. **42**, 421–433 (2021).

15. E.-M. Jouhilahti, E. Madissoon, L. Vesterlund, V. Töhönen, K. Krjutškov, A. Plaza Reyes, S. Petropoulos, R. Månsson, S. Linnarsson, T. Bürglin, F. Lanner, O. Hovatta, S. Katayama, J. Kere, The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. *Development*. **143**, 3459–3469 (2016).

16. A. Leidenroth, J. E. Hewitt, A family history of DUX4: Phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evol Biol*. **10** (2010), doi:10.1186/1471-2148-10-364.

17. V. Töhönen, S. Katayama, L. Vesterlund, E.-M. Jouhilahti, M. Sheikhi, E. Madissoon, G. Filippini-Cattaneo, M. Jaconi, A. Johnsson, T. R. Bürglin, S. Linnarsson, O. Hovatta, J. Kere, Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat Commun*. **6**, 8207 (2015).

18. Z. Yao, L. Snider, J. Balog, R. J. L. F. Lemmers, S. M. van der Maarel, R. Tawil, S. J. Tapscott, DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum Mol Genet*. **23**, 5342–5352 (2014).

19. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. **9**, 559 (2008).

20. S. Jagannathan, Y. Ogata, P. R. Gafken, S. J. Tapscott, R. K. Bradley, Quantitative proteomics reveals key roles for post-transcriptional gene regulation in the molecular pathology of facioscapulohumeral muscular dystrophy. *Elife*. **8**, 1–15 (2019).

21. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. **38**, 576–589 (2010).

22. J. Hao, X. Yang, C. Zhang, X.-T. Zhang, M. Shi, S.-H. Wang, L. Mi, Y.-T. Zhao, H. Cao, Y. Wang, KLF3 promotes the 8-cell-like transcriptional state in pluripotent stem cells. *Cell Prolif*. **53**, e12914–e12914 (2020).

23. S. Katayama, V. Ranga, E.-M. Jouhilahti, T. T. Airenne, M. S. Johnson, K. Mukherjee, T. R. Bürglin, J. Kere, Phylogenetic and mutational analyses of human LEUTX, a homeobox gene implicated in embryogenesis. *Sci Rep*. **8**, 17421 (2018).

24. F. Liu, Q. Fu, Y. Li, K. Zhang, M. Tang, W. Jiang, B. Bo, Y. Cui, L. Kong, USP21 modulates Goosecoid function through deubiquitination. *Biosci Rep*. **39**, BSR20182148 (2019).

25. T. Kujirai, N. Horikoshi, K. Sato, K. Maehara, S. Machida, A. Osakabe, H. Kimura, Y. Ohkawa, H. Kurumizaka, Structure and function of human histone H3.Y nucleosome. *Nucleic Acids Res*. **44**, 6127–6141 (2016).

26. E. Madissoon, E.-M. Jouhilahti, L. Vesterlund, V. Töhönen, K. Krjutškov, S. Petropoulos, E. Einarsdottir, S. Linnarsson, F. Lanner, R. Månsson, O. Hovatta, T. R. Bürglin, S. Katayama, J. Kere, Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. *Sci Rep*. **6**, 28995 (2016).

27. P. P. Boyl, M. Signore, A. Annino, J. P. M. Barbera, D. Acampora, A. Simeone, Otx genes in the development and evolution of the vertebrate brain. *International Journal of Developmental Neuroscience*. **19**, 353–363 (2001).

28. B. Ulmer, M. Tingler, S. Kurz, M. Maerker, P. Andre, D. Mönch, M. Campione, K. Deißler, M. Lewandoski, T. Thumberger, A. Schweickert, A. Fainsod, H. Steinbeißer, M. Blum, A novel role of the organizer gene Goosecoid as an inhibitor of Wnt/PCP-mediated convergent extension in Xenopus and mouse. *Sci Rep*. **7**, 43010 (2017).

29. K. Shiomi, T. Kiyono, K. Okamura, M. Uezumi, Y. Goto, S. Yasumoto, S. Shimizu, N. Hashimoto, CDK4 and cyclin D1 allow human myogenic cells to recapture growth property without compromising differentiation potential. *Gene Ther*. **18**, 857–866 (2011).

30. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. **9**, 559 (2008).

31. B. Zhang, S. Horvath, A General Framework for Weighted Gene Co-Expression Network Analysis. **4** (2005), doi:doi:10.2202/1544-6115.1128.

Chapter 4

Summary and Conclusion

4.1 Road to FSHD discoveries

In 1884, the first case of FSHD was described by two French physicians Louis Landouzy and Joseph Dejerine. The formal definition of clinical features of FSHD was not defined until 1952. It took 40 more years to discover that FSHD is associated with the 4q35 region of chromosome 4 and D4Z4 repeat contraction in the early 1990s. Shortly after that, DUX4 was finally discovered in 1999. Within the past two decades, discoveries on FSHD and DUX4 have accelerated owing to the advancement in high-throughput sequencing technologies as we are entering an exciting era of big data. Recent findings in FSHD field have shined a light on our understanding of human biology both in diseases and development. The biological role of DUX4 has been unveiled across multiple fields including early embryonic development, muscular dystrophy, cancer progression. FSHD and DUX4 have been a subject of scientific research debate and many more studies are being conducted to dismantle the disease mechanism.

4.2 Heterochromatin disruption in FSHD

Previously, the Yokomori laboratory found that D4Z4 repeats contain the heterochromatin structure consisting of H3K9me3 that recruits cohesin and HP1γ; and they also found loss of H3K9me3 in both FSHD1 and FSHD2. These alterations enable DUX4 de-repression/reactivation of DUX4 in FSHD patients, which was linked to ineffective

muscle function (*1*). However, it was unclear what causes H3K9me3 loss. More recently, SMCHD1, a known epigenetic modifier, has been implicated in the heterochromatin establishment at D4Z4 repeats. SMCHD1 mutations are linked to most of FSHD2 population (*2*, *3*). The Yokomori laboratory demonstrated that SMCHD1 is recruited to D4Z4 repeats in an H3K9me3-dependent manner (*4*), and in turn promotes H3K9me3 (Chapter 2). Additionally, other epigenetic modifiers, LRIF1 and DNM3TB were also found to be mutated in FSHD2 patients, implicated in heterochromatin loss and DUX4 de-repression (*5*, *6*). Therefore, multiple factors, possibly including some additional undiscovered epigenetic/gene modifiers, may be involved in the regulation of D4Z4 heterochromatin and DUX4 repression which are critical for FSHD pathogenesis.

4.3 D4Z4 contracted mutants recapitulate FSHD phenotypes

Cumulative evidence indicated that degree of FSHD penetrance varies across population and depends on many factors (*7*, *8*). However, it is clear that repeat sizes are highly associated with the severity of FSHD symptoms (*9*). Our analysis of FSHD mutants generated from artificial deletion of D4Z4 repeats by CRIPSR-Cas9 confirmed that DUX4 can be re-activated from the artificial deletion of D4Z4 repeats in adult muscles (Figure 2.2) (*10*). Our SMCHD1 mutants, however, did not effectively activate DUX4 and its target genes unlike the contracted mutants. One possibility is that epigenetic modifiers such as SMCHD1 are more crucially involved during the early development when cells first begin to silence D4Z4 repeats (*11*). Meanwhile, D4Z4 repeat contraction can cause reactivation of DUX4 even in adult muscles. This is consistent with the fact that some individuals without the

inherited contracted repeats appear to develop de novo D4Z4 mutations, leading to FSHD phenotypes (*12*). Interestingly, however, SMCHD1 mutations in our study (induced in adult myoblasts) still synergize quite effectively with D4Z4 contraction mutants and exponentially increase DUX4 target genes' expression, recapitulating the severe cases of FSHD1 with SMCHD1 mutations. We also found that SMCHD1 mutations reduce H3K9me3. Our results suggest that SMCHD1 is still important in regulating DUX4 expression in adults, possibly through H3K9me3 establishment (*10*). The detailed mechanism of how SMCHD1 is involved in heterochromatic establishment of D4Z4 repeats and/or DUX4 expression are yet to be determined. Nonetheless, our FSHD mutants engineered from a healthy permissive myoblast line closely recapitulate FSHD patient phenotype.

4.4 DUX4 coherent feedforward loop

When H3.X and H3.Y were discovered to increase perdurance and enhance expression of DUX4 targets, it is clear that DUX4 targets can act to self-enhance the DUX4 target gene network (*13*). In fact, reduced expression of LEUTX and DUXA by shRNA in FSHD2 significantly decreased other DUX4 targets (*14*). Overexpression of H3.X/Y and LEUTX significantly increased other DUX4 targets more than 10-fold in FSHD differentiated myotube mutants (Figure 2.6 and 2.7) (*10*). These findings explain why little to no detectable gene expression of DUX4 is sufficient to activate DUX4 network and inhibit myogenesis because these DUX4 targets can upregulate other targets and carry on expanding the network (*15*, *16*). Moreover, H3.X/Y are early target genes which are activated as early as myoblast while LEUTX is activated in later differentiation. LEUTX

expression requires myotube differentiation because it remained to be lowly expressed in myoblasts even with the 5AzaC treatment or H3.X overexpression. Thus, H3.X/Y may be activated first and promote the expression of later DUX4 targets such as LEUTX and DUXA (*10*). LEUTX then further enhances the DUX4 network and also stimulates H3.X/Y in a positive feedback loop in differentiated myocytes. It is not clear what mechanism allows DUX4 target to be efficiently upregulated during differentiation. One possibility is the ability of fused myotubes to spread transcripts across multiple nuclei; and therefore, this process might enable DUX4 and its targets' transcripts to expand their influences efficiently within the same myotube (*17*). Alternatively, additional differentiation-specific cofactor(s) may be required in addition to DUX4 for LEUTX and other late target gene expression. Altogether, our results raise the possibility that these DUX4 targets coordinate on a timeline to upregulate the genes in the network without the need of continuous DUX4 gene expression.

4.5 DUX4 and LEUTX role in early embryonic program

The earliest role of DUX4 begins during embryonic genome activation (EGA)(*18*, *19*). The biological function of DUX family is relatively conserved across mammalian species and was suggested to diverge from a single homeobox sDUX (*20*). This family of developmental genes is believed to evolve relatively fast to eliminate infertile off-spring early during development and preserve energy for the survival of the species. Therefore, understanding the biological role of DUX family and DUX4 could provide more insight into evolution and development. Indeed, other homologs of DUX4 such as DUXC and Dux were

discovered to play a similar role in other mammalian species (*21*, *22*). DUX4 is expressed during 2-cell to 4-cell stage while LEUTX, DUXA, DUXB (DUX4 targets) and other PRD-like transcription factors are expressed during 4-cell to 8-cell stage (*23*, *24*). Therefore, DUX4 may initiate directly or indirectly the expression of downstream PRD-like genes, and some of which such as LEUTX and DUXA are found to further enhance the DUX4 network activation and continue to possibly regulate the embryonic program. It is not completely clear how other PRD-like genes coordinate with PRD-like DUX4 targets to regulate embryonic development. However, it is certain that most of these PRD-like transcription factors are strictly expressed at early embryonic stage and their functions are the key to developing embryo as they could dynamically upregulate or downregulate each other to balance the overall transcriptional regulation of EGA (*23*). PRD-like genes as well as other genes in the homeobox family must coordinate their expression to regulate early embryonic development and further understanding of the relationship of these genes could provide more insight on embryonic cell fate.

Indeed, from the overexpression and depletion of LEUTX experiment, I found that LEUTX regulated a subset of embryonic genes that are not considered DUX4 targets (Figure 3.5 and S3.4). For example, DPRX is differentially expressed when knocking down or overexpressing LEUTX. DPRX was previously found to be upregulated at 8-cell stage and suggested to suppress LEUTX (*25*). Therefore, there seems to be a pattern of cross inhibition and activation among the developmental genes. Like DUX4 and other homeobox genes, LEUTX emerged from duplication and divergence event (*26*). Perhaps, their similar origin reflects in their close relationship among the homeobox genes and also in their conserved biological functions across species in regulating the embryonic program.

4.6 Future Direction

It has been an honor for me to join the FSHD research community and participated in the effort to understand human diseases applying the next generation technologies. During my time as a graduate student in the Yokomori lab and also Mortazavi lab, I had the chance to contribute to a larger understanding of DUX4 mechanism and FSHD pathology. FSHD is a complex disease and is associated with DUX4 which is also a complex transcription factor. Future studies should focus in both areas of muscle regeneration and embryonic development to characterize how FSHD patients develop muscle dystrophy overtime. An important research direction is to further investigate the feedback regulation of the DUX4 target gene network, which may reveal a potentially more effective target(s) for FSHD therapy development. LEUTX and DUXA are good candidate DUX4 target genes as they are found to be restricted to early embryonic development. I hope future members of the project will continue the work I leave here.

Reference:

1. W. Zeng, J. C. de Greef, Y.-Y. Chen, R. Chien, X. Kong, H. C. Gregson, S. T. Winokur, A. Pyle, K. D. Robertson, J. A. Schmiesing, V. E. Kimonis, J. Balog, R. R. Frants, A. R. Ball Jr., L. F. Lock, P. J. Donovan, S. M. van der Maarel, K. Yokomori, 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).

2. R. J. L. F. Lemmers, R. Tawil, L. M. Petek, J. Balog, G. J. Block, G. W. E. Santen, A. M. Amell, P. J. van der Vliet, R. Almomani, K. R. Straasheijm, Y. D. Krom, R. Klooster, Y. Sun, J. T. den Dunnen, Q. Helmer, C. M. Donlin-Smith, G. W. Padberg, B. G. M. van Engelen, J. C. de Greef, A. M. Aartsma-Rus, R. R. Frants, M. de Visser, C. Desnuelle, S. Sacconi, G. N. Filippova, B. Bakker, M. J. Bamshad, S. J. Tapscott, D. G. Miller, S. M. van der Maarel, Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet*. **44**, 1370–1374 (2012).

3. S. Sacconi, R. J. L. F. Lemmers, J. Balog, P. J. van der Vliet, P. Lahaut, M. P. van Nieuwenhuizen, K. R. Straasheijm, R. D. Debipersad, M. Vos-Versteeg, L. Salviati, A. Casarin, E. Pegoraro, R. Tawil, E. Bakker, S. J. Tapscott, C. Desnuelle, S. M. van der Maarel, The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1. *Am J Hum Genet*. **93**, 744–751 (2013).

4. W. Zeng, Y. Y. Chen, D. A. Newkirk, B. Wu, J. Balog, X. Kong, A. R. Ball Jr., S. Zanotti, R. Tawil, N. Hashimoto, A. Mortazavi, S. M. van der Maarel, K. Yokomori, 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).

5. K. Hamanaka, D. Šikrová, S. Mitsuhashi, H. Masuda, Y. Sekiguchi, A. Sugiyama, K. Shibuya, R. J. L. F. Lemmers, R. Goossens, M. Ogawa, K. Nagao, C. Obuse, S. Noguchi, Y. K. Hayashi, S. Kuwabara, J. Balog, I. Nishino, S. M. van der Maarel, Homozygous nonsense variant in LRIF1 associated with facioscapulohumeral muscular dystrophy. *Neurology*. **94**, e2441–e2447 (2020).

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

7. M. Wohlgemuth, R. J. Lemmers, M. Jonker, E. van der Kooi, C. G. Horlings, B. G. van Engelen, S. M. van der Maarel, G. W. Padberg, N. C. Voermans, A family-based study into penetrance in facioscapulohumeral muscular dystrophy type 1. *Neurology*. **91**, e444 (2018).

8. E. Salort-Campana, K. Nguyen, R. Bernard, E. Jouve, G. Solé, A. Nadaj-Pakleza, J. Niederhauser, E. Charles, E. Ollagnon, F. Bouhour, S. Sacconi, A. Echaniz-Laguna, C. Desnuelle, C. Tranchant, C. Vial, F. Magdinier, M. Bartoli, M.-C. Arne-Bes, X. Ferrer, T. Kuntzer, N. Levy, J. Pouget, S. Attarian, Low penetrance in facioscapulohumeral muscular dystrophy type 1 with large pathological D4Z4 alleles: a cross-sectional multicenter study. *Orphanet J Rare Dis*. **10**, 2 (2015).

9. R. Tawil, J. Forrester, R. C. Griggs, J. Mendell, J. Kissel, M. McDermott, W. King, B. Weiffenbach, D. Figlewicz, T. F.-D. Group, Evidence for anticipation and association of deletion size with severity in facioscapulohumerd muscular dystrophy. *Ann Neurol*. **39**, 744–748 (1996).

10. X. Kong, N. V. Nguyen, Y. Li, J. S. Sakr, K. Williams, S. Sharifi, J. Chau, A. Bayrakci, S. Mizuno, S. Takahashi, T. Kiyono, R. Tawil, A. Mortazavi, K. Yokomori, *bioRxiv*, doi:10.1101/2022.10.14.512332.

11. C. T. Gordon, S. Xue, G. Yigit, H. Filali, K. Chen, N. Rosin, K. Yoshiura, M. Oufadem, T. J. Beck, R. McGowan, A. C. Magee, J. Altmüller, C. Dion, H. Thiele, A. D. Gurzau, P. Nürnberg, D. Meschede, W. Mühlbauer, N. Okamoto, V. Varghese, R. Irving, S. Sigaudy, D. Williams, S. F. Ahmed, C. Bonnard, M. K. Kong, I. Ratbi, N. Fejjal, M. Fikri, S. C. Elalaoui, H. Reigstad, C. Bole-Feysot, P. Nitschké, N. Ragge, N. Lévy, G. Tunçbilek, A. S. M. Teo, M. L. Cunningham, A. Sefiani, H. Kayserili, J. M. Murphy, C. Chatdokmaiprai, A. M. Hillmer, D. Wattanasirichaigoon, S. Lyonnet, F. Magdinier, A. Javed, M. E. Blewitt, J. Amiel, B. Wollnik, B. Reversade, De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nat Genet*. **49**, 249–255 (2017).

12. S. M. van der Maarel, G. Deidda, R. J. L. F. Lemmers, P. G. M. van Overveld, M. van der Wielen, J. E. Hewitt, L. Sandkuijl, B. Bakker, G.-J. B. van Ommen, G. W. Padberg, R. R. Frants, De Novo Facioscapulohumeral Muscular Dystrophy: Frequent Somatic Mosaicism, Sex-Dependent Phenotype, and the Role of Mitotic Transchromosomal Repeat Interaction between Chromosomes 4 and 10. *The American Journal of Human Genetics*. **66**, 26–35 (2000).

13. R. Resnick, C. J. Wong, D. C. Hamm, S. R. Bennett, P. J. Skene, S. B. Hake, S. Henikoff, S. M. van der Maarel, S. J. Tapscott, DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. *Cell Rep*. **29**, 1812-1820.e5 (2019).

14. S. Jiang, K. Williams, X. Kong, W. Zeng, N. V. Nguyen, X. Ma, R. Tawil, K. Yokomori, A. Mortazavi, Single-nucleus RNA-seq identifies divergent populations of FSHD2 myotube nuclei. *PLoS Genet*. **16**, 1–26 (2020).

15. D. Bosnakovski, M. D. Gearhart, E. A. Toso, S. H. Choi, M. Kyba, C. Biology, M. Sciences, S. Korea, M. Kyba, Low level DUX4 expression disrupts myogenesis. *Sci Rep* (2000), doi:10.1038/s41598-018-35150-8.

16. Z. Yao, L. Snider, J. Balog, R. J. L. F. Lemmers, S. M. van der Maarel, R. Tawil, S. J. Tapscott, DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum Mol Genet*. **23**, 5342–5352 (2014).

17. J. Chau, X. Kong, N. Viet Nguyen, K. Williams, M. Ball, R. Tawil, T. Kiyono, A. Mortazavi, K. Yokomori, Relationship of DUX4 and target gene expression in FSHD myocytes. *Hum Mutat*. **42**, 421–433 (2021).

18. L. N. Geng, Z. Yao, L. Snider, A. P. Fong, J. N. Cech, J. M. Young, S. M. van der Maarel, W. L. Ruzzo, R. C. Gentleman, R. Tawil, S. J. Tapscott, DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev Cell*. **22**, 38–51 (2012).

19. A. de Iaco, E. Planet, A. Coluccio, S. Verp, J. Duc, D. Trono, DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet*. **49**, 941–945 (2017).

20. A. Leidenroth, J. E. Hewitt, A family history of DUX4: Phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evol Biol*. **10** (2010), doi:10.1186/1471-2148-10-364.

21. C.-J. Wong, J. L. Whiddon, A. T. Langford, A. E. Belleville, S. J. Tapscott, Canine DUXC: implications for DUX4 retrotransposition and preclinical models of FSHD. *Hum Mol Genet*. **31**, 1694–1704 (2022).

22. P. G. Hendrickson, J. A. Doráis, E. J. Grow, J. L. Whiddon, J.-W. Lim, C. L. Wike, B. D. Weaver, C. Pflueger, B. R. Emery, A. L. Wilcox, D. A. Nix, C. M. Peterson, S. J. Tapscott, D. T. Carrell, B. R. Cairns, Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet*. **49**, 925–934 (2017).

23. V. Töhönen, S. Katayama, L. Vesterlund, E.-M. Jouhilahti, M. Sheikhi, E. Madissoon, G. Filippini-Cattaneo, M. Jaconi, A. Johnsson, T. R. Bürglin, S. Linnarsson, O. Hovatta, J. Kere, Novel PRD-like homeodomain transcription factors and retrotransposon elements in early human development. *Nat Commun*. **6**, 8207 (2015).

24. E. Madissoon, E.-M. Jouhilahti, L. Vesterlund, V. Töhönen, K. Krjutškov, S. Petropoulos, E. Einarsdottir, S. Linnarsson, F. Lanner, R. Månsson, O. Hovatta, T. R. Bürglin, S. Katayama, J. Kere, Characterization and target genes of nine human PRD-like homeobox domain genes expressed exclusively in early embryos. *Sci Rep*. **6**, 28995 (2016).

25. E.-M. Jouhilahti, E. Madissoon, L. Vesterlund, V. Töhönen, K. Krjutškov, A. Plaza Reyes, S. Petropoulos, R. Månsson, S. Linnarsson, T. Bürglin, F. Lanner, O. Hovatta, S. Katayama, J. Kere, The human PRD-like homeobox gene LEUTX has a central role in embryo genome activation. *Development*. **143**, 3459–3469 (2016).

26. S. Jagannathan, The evolution of DUX4 gene regulation and its implication for facioscapulohumeral muscular dystrophy. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*. **1868**, 166367 (2022).

Appendix I

DNA damage-activated promotes perinucleolar clustering of H3K9me3 domains of acrocentric chromosomes

1. Abstract

In response to genotoxic stress, cells activate DNA damage response/repair (DDR) which is critical to stabilize and maintain the integrity of a genome. Recent research has focused on the immediate responses following cellular damages, such as damage recognition, local chromatin structural changes and repair pathway choice at the damage sites. However, secondary damage/response has not been well understood. Genome-wide damage-induced epigenetic changes have been reported, such as nucleus-wide histone H3 hypoacetylation and chromosome territory rearrangements, but primary and secondary responses are not well separated, and details of mechanisms and functional significance remain obscure. In this study, we examined the nucleus-wide DDR in response to complex DNA damage induced by methyl methanesulfonate (MMS) and laser microirradiation. We found that chromatin domains harboring H3K9 trimethylation (H3K9me3) cluster to form a ring-like structure around nucleoli, which is spatially distinct from the initial DNA damage sites. To identify which genomic regions are involved in this ring formation, I performed HiChIP (HiC combined with immunoprecipitation) using antibody specific for H3K9m3 and found that the acrocentric chromosomes specifically increase intra-chromosomal H3K9me3 domain interactions. Since acrocentric chromosomes that encode

ribosomal RNA (rRNA) primarily localize in the perinucleolar area, the results strongly suggest that these acrocentric chromosomes specifically undergo structural changes involving H3K9me3 heterochromatin domains in response to DNA damage. The findings for the first time revealed H3K9me3-specific chromatin domain reorganization as a secondary response to DNA damage. Underlying regulatory mechanisms and their potential impact on cell survival are currently being investigated.

2. Introduction

Genome integrity is crucial for an organism to survive harmful mutations, protect organism from cancer progression and preserve genetic information(*1*). Human genome is constantly vulnerable to damage by exposure from various harmful sources such as chemical agents, UV irradiation or oxidative stress(*2–4*). Damages induced by these factors can cause DNA strand breaks and complex DNA damage(*5*). In order to deal with the damage, cells facilitate DNA damage responses (DDR) to repair the breaks and maintain genomic stability(*6*). DNA repair is tightly linked with cell cycle progression as unrepaired DNA can trigger signaling pathway to interfere with cell cycle before DNA replication (G1/S arrest) or cell division (G2/M arrest)(*7*). If DNA damage persists, programmed cell death will be activated to remove cells that are potentially becoming toxic to the organism.

Findings have indicated that DDR can induce epigenetic changes to promote cellular survival(*8*, *9*). Damage sensor such as PARP1 (Poly(ADP-ribose) polymerase 1) can rapidly recruits chromatin regulators at damage sites for repair. However, hyperactivation of PARP1 promote glycolysis pathway and eventually can lead to cell death(*10*, *11*). Histone

modifications are also involved in DDR(*12*). For instance, H3K9Ac and H3K56Ac were found to be globally reduced in response to DNA damage(*13*, *14*). These histone marks are specifically diminished at promoter regions of cell cycles and active genes in response to DNA damage. H3K9me3, a repressive histone modification, is deposited to nearby DNA double-strand breaks to support chromatin stabilization and repair activation(*15*). One of the early DDR is the phosphorylation of γH2AX which can lead to recruitment of other chromatin modifiers(*16*). In addition, γH2AX also controls cell cycle check point regulation, allowing sufficient time for DNA repair to take place.

Chromosome territory rearrangements were shown to occur in response to DNA damage(*17*, *18*). These studies used conventional DNA damaging agents that induce damage nuclear-wide. Therefore, whether these changes are happening at the damage sites or it is a secondary response distinct from damage site modification was unclear. Furthermore, it is not studied extensively the connection between chromatin architecture and DDR signaling. It is not clear whether and how genome-wide chromatin reorganization could promote DNA repair induced by complex DNA damage and/or damaged cell survival. In our current study, we investigated the genome-wide effect of DNA damage on H3K9 trimethylation (H3K9me3). We observed a ring formation of H3K9me3 clustered around nucleoli after 4 hours of recovery from DNA damage. This cluster of H3K9me3 forms in both G1 and S/G2 phase, but was associated with γH2AX only in S/G2, suggesting the secondary damage induced by DNA replication in these regions. Furthermore, we found that MMS-induced damage was associated with the gains of H3K9me3-specific chromatin interactions specifically within the acrocentric chromosomes. This study for the first time revealed the changes in H3K9me3 chromatin domain reorganization in response to

complex DNA damage. Underlying mechanisms and possible functional significance of this novel epigenetic response is currently underway.

3. Results

Laser-induced damage triggered H3K9me3 ring formation

We performed loci-specific damage by laser according to our previously published paper in HeLa cells(*19*). Damage was induced by laser for 1 min and allowed for 4 hours recovery before fixing and staining with H3K9me3. The yellow arrow indicates the damage site with distinct secondary DDR associated with a cluster of H3K9me3 around the peri-nucleolar area (Fig. 1A). Similar to laser, MMS (methyl methanesulfonate) induced H3K9me3 formation which also clustered with DAPI, indicating an increase in chromatin density around the peri-nucleolar area (Fig. 1B and D). In addition, HP1α, but not HP1γ, colocalizes to this ring structure (Fig. 1D and E). There was also no significant localization of H3K4me3 with the ring formation, implying that ring formation is H3K9me3-specific (Fig. 1D and E). Interestingly, there is a clustering of γH2AX with H3K9me3 around the peri-nucleolar area occurring at late S/G2 which suggests that the ring structure consists of double-strand breaks (DSBs) and possibly is associated with secondary damage induced in late S phase (Fig. 1C). Together, these results showed that H3K9me3-specific chromatin clustered to form a ring structure in response to DNA damage, which becomes a major site of replication-induced secondary DSB sites.

Global H3K9me3 remains unaltered after MMS treatment

To examine the status of global H3K9me3 in HeLa cells after MMS treatment, we performed CUT&RUN targeting H3K9me3 histone marks. HeLa cells were treated with MMS for 1 hour and allowed for 4 hours recovery before harvesting. Unlike conventional ChIP-seq method, CUT&RUN was carried out without cross-linking, hence preserving the native interaction of H3K9me3 and chromatin. CUT&RUN also allowed us to use less chromatin materials (around 0.5-1 million cells were used) and obtained high signal-to-noise ratio(*20*). Similar to our previous study, the alignment was done using SICER which is more suitable for histone marks with broad domains such as H3K9me3(*21*). Reads were normalized using RPKM and then proceeded to EdgeR for downstream analysis. To our surprise, global H3K9me3 was not different between control and MMS in HeLa cells (Figure 2A and B). There were only 2 differential peaks (data not shown). This implies that other changes such as H3K9me3 chromatin architecture might influence the H3K9me3 ring formation. This is consistent with the visible ring formation by DAPI staining, indicative of increased density of DNA (clustering of H3K9me3-positive chromatin) rather than increase of H3K9me3 modification.

H3K9me3-specific chromatin interactions significantly increased with MMS at acrocentric chromosomes

To determine H3K9me3-specific chromatin changes induced by DNA damage, we performed HiChIP (Highly Integrative Chromatin Immunoprecipitation) which is a sequencing method to capture specific chromatin interaction. We immunoprecipitated H3K9me3 on HeLa cells with the same treatment and recovery condition as described

above. To our knowledge, this is the first H3K9me3 HiChIP done to explore H3K9me3-specific chromatin interaction. Around 250M HiChIP reads per sample were filtered and mapped using HiC-Pro and differential interactions were calculated using Homer. A total of 509 H3K9me3-specific chromatin interactions were gained in HeLa cells with MMS treatment while 302 interactions were lost. Interestingly, the gains of interaction localized significantly on the acrocentric chromosomes (13, 14,15 and 22) (Figure 2C). Coincidently, NORs (Nucleolar organizer regions) which contain ribosomal DNA (rDNA) arrays are located on the p-arms of the acrocentric chromosomes. This observation implied that the increase of H3K9me3 chromatin interactions at the acrocentric chromosomes might serve to assist with DDR process. Particularly, chromosome 22 appeared to have the most increase in H3K9me3 chromatin interaction follows by chromosome 15 (Figure 2C and 3A, B). We also examined the inter-chromosomal H3K9me3 chromatin interactions and found no obvious difference between control and MMS treatment (Figure 3C). This further suggested that changes in H3K9me3 chromatin interaction occurred only within each chromosome. Together, the results from HiChIP indicated that secondary DNA damage induced by laser or MMS promoted increase in H3K9me3-specific chromatin interactions which could be associated with H3K9me3 ring formation from the immunofluorescent results.

4. Discussion

Chromatin remodeling participates in DDR by promoting recruitment of repairing complexes and regulating cell cycle check points. Past findings have primarily focused on

chromatin remodeling and movement at damage sites(*9*, *22–25*). In our current study, laser-induce damage triggered a distinct secondary response which was separate from the primary damage site (Figure 1A). We also observed that the secondary damage response induced by either laser or MSS was associated with H3K9me3 and γH2AX ring formation around the nucleoli which suggested sites of DNA damage repair. The mechanism underlying this response is not clearly understood but we hypothesize that the movement of chromatin could be involved with DNA repair and/or gene regulation that promotes cell survival.

From the HiChIP results, we found that H3K9me3 chromatin interactions increased with MMS treatment predominantly at the acrocentric chromosomes (Figure 2D). The p-arms of the acrocentric chromosomes contain ribosomal gene (rDNA) arrays which form the structure of the nucleolar organizer regions (NORs) and also are sites of rRNA transcription (*26*). It is not clear the association between H3K9me3 ring formation and DNA repair, but our results showed that the increase in H3K9me3 interactions of the acrocentric chromosomes could be related to the increase of H3K9me3 chromatin density around the nucleoli. More evidence and studies will be necessary to confirm this possibility. However, studying these rDNA regions can be challenging because the regions are highly repetitive and missing from human genome drafts(*27*). Therefore, H3K9me3-specific chromatin interactions at the p-arm of the acrocentric chromosomes were not able to be captured. Notably, chromosome 22 had the most increase in H3K9me3 chromatin interactions out of the five acrocentric chromosomes (Figure 2C and 3).

In conclusion, the study for the first time explored the H3K9me3-specific chromatin interactions in HeLa cells in the context of DNA damage response. Our preliminary results indicated that secondary DNA damage response which was distinct from the initial damage induced H3K9me3 ring formation around the nucleoli. The ring formation was likely to be involved in DNA repair due to the presence of γH2AX also around the nucleoli. Because the acrocentric chromosomes play an important role in nucleolar organizer regions, we anticipate that the increased signal of H3K9me3-specific chromatin interactions coincide with the ring formation. However, more studies will be required to investigate the connection between these two events.

5. Method

Cell culture

HeLa adenocarcinoma cells were cultured in DMEM (Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS), 2 mM l-glutamate, and 1% penicillin-streptomycin. HeLa cells were tested and shown to be free of mycoplasma contamination.

Immunofluorescent staining

Staining was performed as previously described(*28*). Briefly, cells grown on coverslips were fixed in 4% paraformaldehyde for 10 min at room temperature, permeabilized with 0.5% Triton X-100 in PBS, and blocked in blocking buffer (0.02% saponin, 0.05% NaN_3, 1% BSA, 4% horse serum and 0.1% gelatin in PBS) for 15 min at 37°C. The coverslips were

incubated overnight with primary antibodies at 4 °C followed by three PBS washes, then incubated with fluorescent secondary antibody for 30 min at 37°C, washed with PBS 3 times, counter-stained with DAPI, and mounted with Prolong Diamond Antifade Mountant. Images were acquired with a Zeiss LSM510 confocal laser microscope. The following antibodies were used for the study: H3K9me3 (Abcam, ab8898), H3K4me3 (Genetex, GTX128954), HP1α (Emdmillipore/Sigma, 05-689), HP1γ (Emdmillipore/Sigma, 05-690), γH2AX (Emdmillipore/Sigma, 05-636).

HiChIP procedure

The procedure was done using Arima-HiC kit with added modification for immunoprecipitation step to pull down H3K9me3-specific chromatin with an antibody. HeLa cells were grown in 15-cm plate and harvested after 4 hours of MMS recovery. Cross-linking and quality control were performed according to Arima-HiC kit instruction. Around 3ug of chromatin was used for DNA proximity ligation step. Immunoprecipitation step started after incubation of buffer C. Nuclei was spun down for 5 mins at 3000 rpm and resuspended with 150 ul of RIPA buffer (10 mM Tris pH 8.0, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS,0.1% sodium deoxycholate) supplemented with proteinase inhibitor (added right before use). Samples were incubated on ice for 10 mins before proceeding to sonication by bioruptor (20 cycles with 30 seconds on and off). Samples were centrifuged for 20 mins at 14,000 rpm and the supernatant was collected. Lysates were precleared with Protein G Dyna beads (Thermo Fisher Scientific) and nutate for 1 hour in 4°C. 1.5 ug of H3K9me3 antibody (abcam, cat# ab8898) was added to each sample

which was then incubated over night at 4°C. The next day, the samples were incubated with Protein G Dyna beads for 2 hours at 4°C. The samples were then washed with 1X Low Salt Immune Complex Wash Buffer, 1X High Salt Immune Complex Wash Buffer, 1X LiCl Immune Complex Wash Buffer, and 1X TE Buffer (0.1 mM Na2EDTA) before eluting with 1% SDS and 0.1 M $NaHCO_3$ buffer at 37°C. Revere cross-linking was done by adding proteinase K to the eluted samples and incubated at 55°C for 2.5 hours. DNA was purified using QIAGEN gel cleanup kit. Biotin enrichment was performed according to Arima protocol to capture biotinylated DNA.

Cut&Run procedure:

The procedure was done using CUTANATM ChIC/CUT&RUN kit from Epycypher (cat# 14-1048) according to the manufacture protocol. HeLa cells were grown in 10-cm plate and around 0.5M cells were used for each reaction. 0.5 ug of H3K9me3 antibody (Abcam, ab8898) was added to each sample.

DNA Library Preparation:

Swift library kit (Accel-NGS® 2S Plus DNA Library Kit, Swift Biosciences) was used to make DNA sequencing libraries for both HiChIP and Cut&Run samples. For Cut&Run samples, the library preparation procedure was done according to Swift Biosciences. However, for HiChIP experiment, the libraries were amplified while DNA bound to the enrichment beads using 2X HiFi HotStart Ready Mix with 11 cycles. Ampure XP beads were

used to purify the final DNA libraries with the fragment sizes between 100-500 bp (determine Ampure XP beads ratio to get the desired fragment sizes).

DNA Sequencing Data Analysis

Raw fastq files of H3K9me3 HiChIP of HeLa control and HeLa with MMS treatment were aligned and called in for valid interactions by HiC-Pro (version 3.0)(*29*). The hg38 genome reference was used for the alignment. HiC-Pro default parameters were used with the exception of allowing multi-mapping reads. HiC-Pro valid interactions were converted to hic files to visualize with Juicebox(*30*). Loop callings were done by Homer with the following parameters res 25000, window 50000, and norm(*31*). Differential loop analysis was also done by Homer to identify loops that were gained or lost in MMS treatment. Differential loops are shown as bed regions with karyotype density plot. Karyotype plot shows chromosomal regions with the gains of H3K9me3 loops in MMS treatment. H3K9me3 loops are highly increased in acrocentric chromosomes (13,14,15, and 22) as indicated by the red boxes. Chromosome 22 shows the most increase in H3K9me3 loops.

For Cut&Run, alignment was done using bowtie2 (version 2.4.1, default parameter) similarly to ChIP-seq analysis. Deeptools were used to make bigwig files for genome track viewer and also to make H3K9me3 signal heatmap. Peaking calling was done using SICER with the following parameters: -f 200 -s hg38 (the rest is default). H3K9me3 peaks were merged and each bam file was used to calculate reads for each sample using coverage command from bedtools. Downstream differential analysis was done using EdgeR.

6. Figures

Figure 1

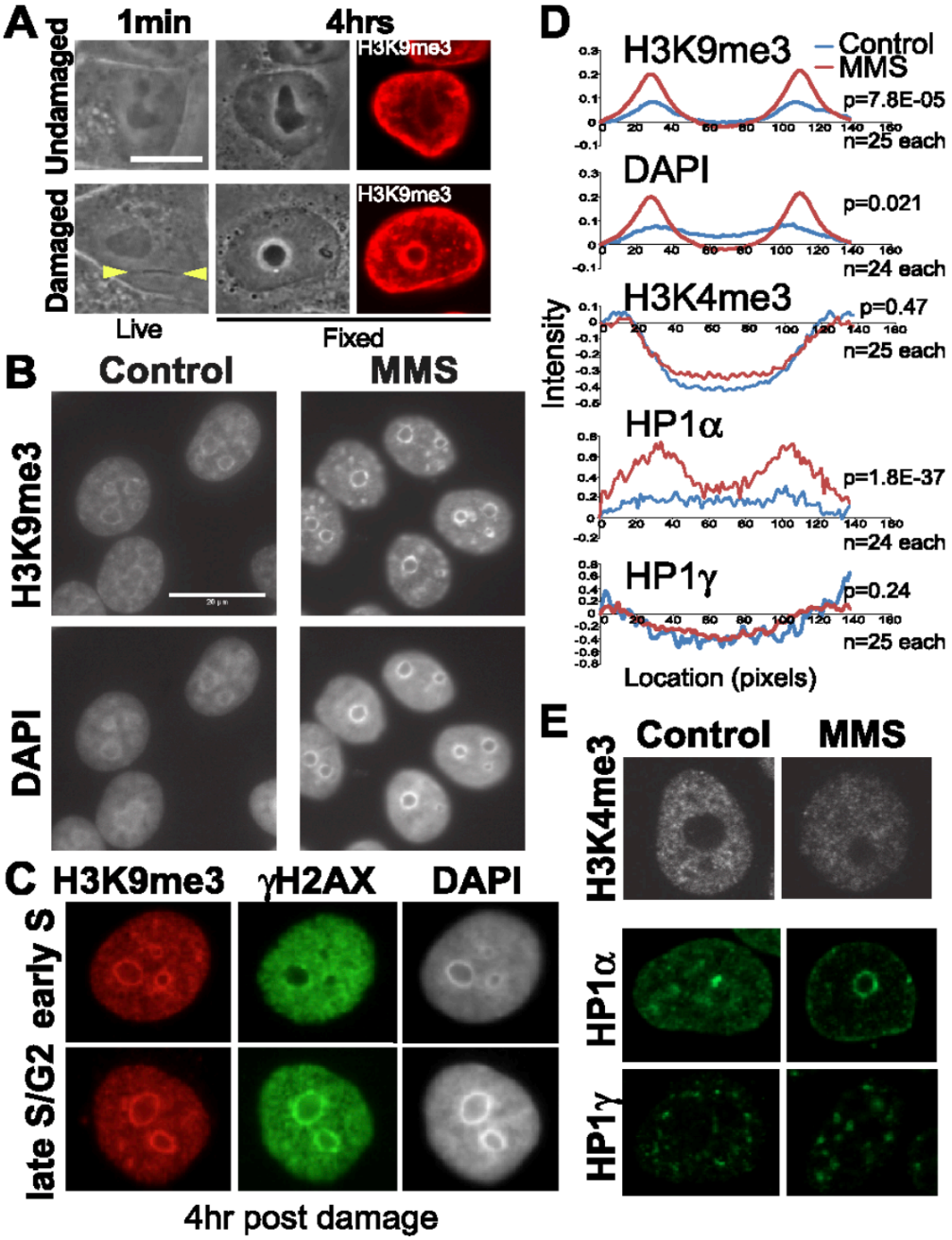


Figure 1: **Formation of H3K9me3 ring as secondary damage response induced by laser or MMS**

(A): H3K9me3 immunostaining of undamaged (no) and laser -damaged nuclei at indicated time points. Yellow arrowheads indicate the damage site.

(B): H3K9me3 and DAPI staining of control and MMS-treated cells at 4 hr post-damage.

(C): Cells were stained for H3K9me3, γH2AX and DAPI at 4 hr after MMS treatment at either early S or late S/G2 (synchronized by double-thymidine block).

(D): Fluorescence measurement across nucleolus using MicroSuite Biological Suite software. Individual nucleoli (n=24~25) were resized and fitted into a 1-inch square box, and an intensity measurement for each pixel along the diagonal line across the box with a resolution of 100 pixels/inch was normalized against the adjacent nucleoplasm and average intensity for control (blue) and MMS (red) were plotted for H3K9me3, DAPI, H3K4me3, HP1α and HP1γ. Sample numbers (n) and p-values are shown.

(E): Example immunofluorescent staining of H3K4me3 (no change), HP1α (change) and HP1γ (no change) with control and MMS treatment.

Figure 2

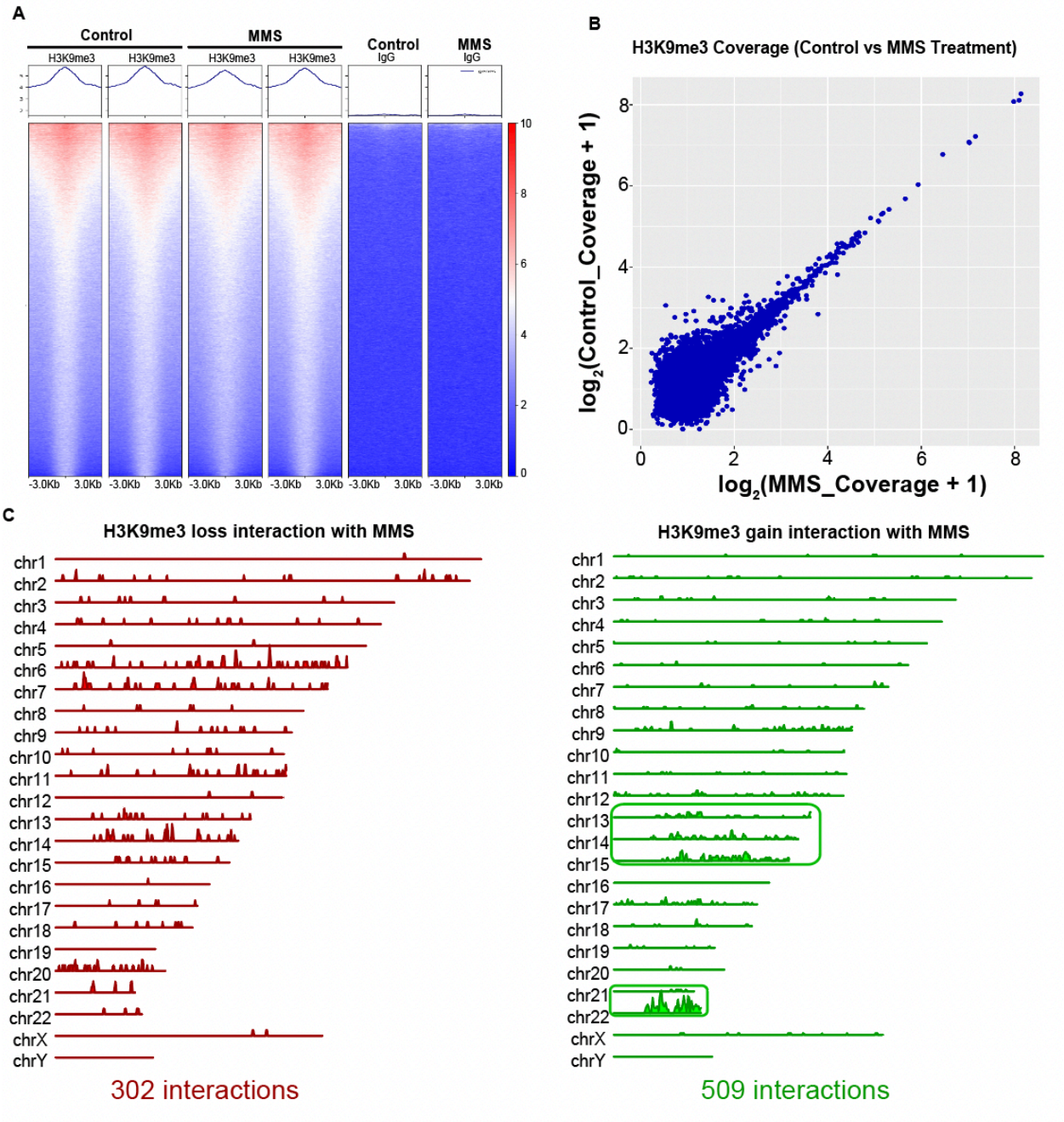

A
Control
MMS
Control
MMS
H3K9me3
H3K9me3
H3K9me3
H3K9me3
IgG
IgG
-3.0Kb
3.0Kb
10
8
6
4
2
0
B
H3K9me3 Coverage (Control vs MMS Treatment)
log2(Control_Coverage + 1)
log2(MMS_Coverage + 1)
C
H3K9me3 loss interaction with MMS
H3K9me3 gain interaction with MMS
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
chrX
chrY
302 interactions
509 interactions

Figure 2: **Specific changes of H3K9me3 chromatin interactions without changing in H3K9me3 ChIP-seq signal.**

(A): H3K9me3 and IgG heatmap signal generated by deeptools showing control and MMS treatment.

(B): Scatter plot of H3K9me3 coverage between control and MMS in $\log_2$ (RPKM)

(C): Karyotype plot of loss and gain of H3K9me3 interactions across the chromosomes. The boxes indicated acrocentric chromosomes with significant increase in H3K9me3 interaction

Figure 3

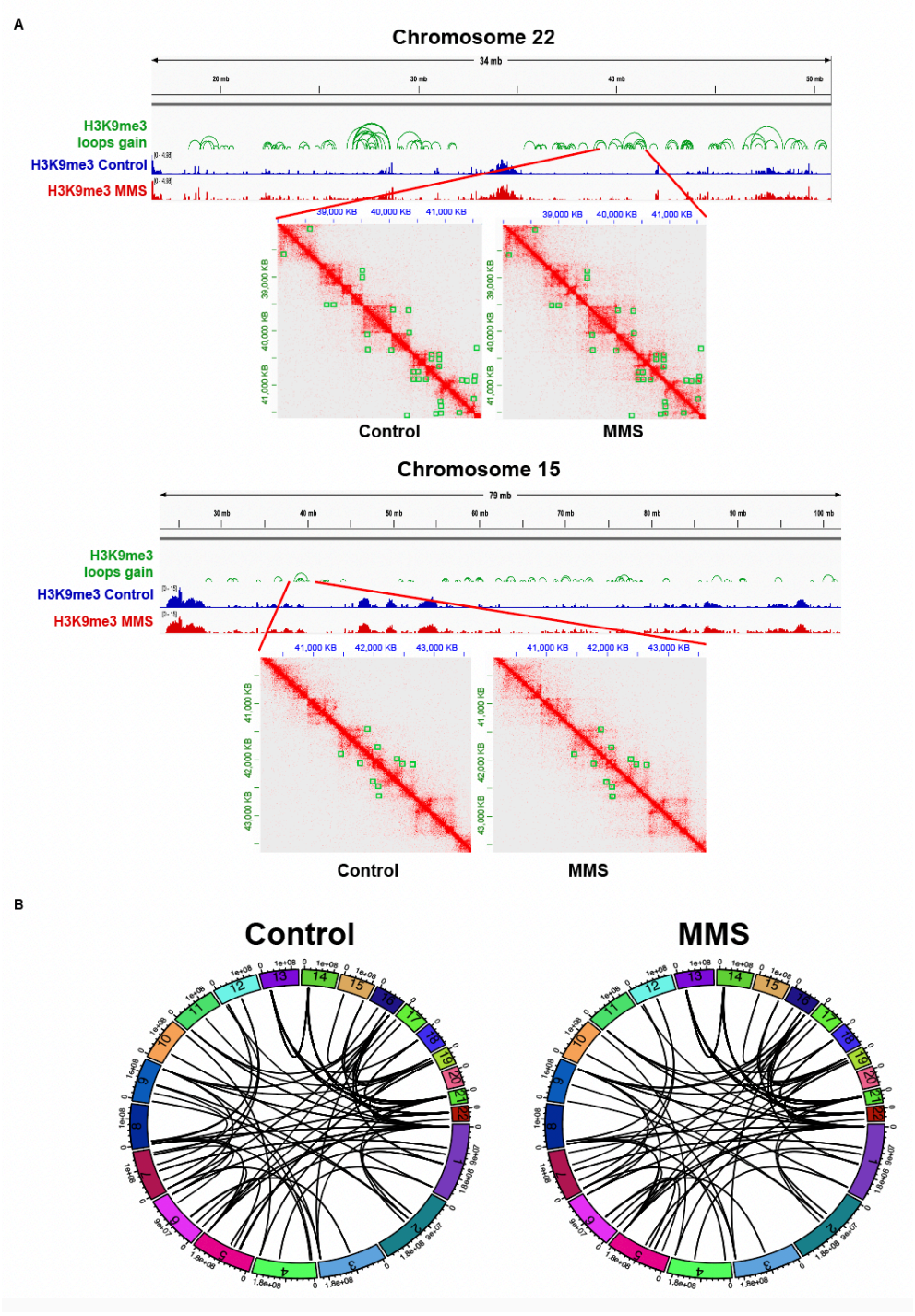

A
Chromosome 22
34 mb
H3K9me3 loops gain
H3K9me3 Control
H3K9me3 MMS
Control
MMS
Chromosome 15
79 mb
H3K9me3 loops gain
H3K9me3 Control
H3K9me3 MMS
Control
MMS
B
Control
MMS

Figure 3: Gains of H3K9me3 interactions are significant in acrocentric chromosomes without any changes in interchromosomal interactions

(A): IGV genome tracks and HiC heatmap (generated by Juicebox) showing the increase of H3K9me3 chromatin interactions at chromosome 22 and 15 with no obvious changes in H3K9me3 ChIP signal comparing between control and MMS treatment

(B): Circos plot indicates no changes in the global interchromosomal interactions between control and MMS treatment

References:

1. A. Carusillo, C. Mussolino, DNA Damage: From Threat to Treatment. *Cells*. **9** (2020), doi:10.3390/cells9071665.

2. R. C. Fry, T. J. Begley, L. D. Samson, GENOME-WIDE RESPONSES TO DNA-DAMAGING AGENTS. *Annu Rev Microbiol*. **59**, 357–377 (2005).

3. N. N. Khodarev, J. O. Park, J. Yu, N. Gupta, E. Nodzenski, B. Roizman, R. R. Weichselbaum, Dose-dependent and independent temporal patterns of gene responses to ionizing radiation in normal and tumor cells and tumor xenografts. *Proceedings of the National Academy of Sciences*. **98**, 12665–12670 (2001).

4. C. A. Koch-Paiz, S. A. Amundson, M. L. Bittner, P. S. Meltzer, A. J. Fornace, Functional genomics of UV radiation responses in human cells. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. **549**, 65–78 (2004).

5. I. v Mavragani, Z. Nikitaki, M. P. Souli, A. Aziz, S. Nowsheen, K. Aziz, E. Rogakou, A. G. Georgakilas, Complex DNA Damage: A Route to Radiation-Induced Genomic Instability and Carcinogenesis. *Cancers (Basel)*. **9** (2017), doi:10.3390/cancers9070091.

6. N. Chatterjee, G. C. Walker, Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen*. **58**, 235–263 (2017).

7. A. Campos, A. Clemente-Blanco, Cell Cycle and DNA Repair Regulation in the Damage Response: Protein Phosphatases Take Over the Reins. *Int J Mol Sci*. **21** (2020), doi:10.3390/ijms21020446.

8. N. P. Dantuma, H. van Attikum, Spatiotemporal regulation of posttranslational modifications in the DNA damage response. *EMBO J*. **35**, 6–23 (2016).

9. P. A. Jeggo, J. A. Downs, Roles of chromatin remodellers in DNA double strand break repair. *Exp Cell Res*. **329**, 69–77 (2014).

10. Z. Herceg, Z. Q. Wang, Functions of poly(ADP-ribose) polymerase (PARP) in DNA repair, genomic integrity and cell death. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. **477**, 97–110 (2001).

11. M. M. Murata, X. Kong, E. Moncada, Y. Chen, H. Imamura, P. Wang, M. W. Berns, K. Yokomori, M. A. Digman, NAD+ consumption by PARP1 in response to DNA damage triggers metabolic shift critical for damaged cell survival. *Mol Biol Cell*. **30**, 2584–2597 (2019).

12. C. R. Hunt, D. Ramnarain, N. Horikoshi, P. Iyengar, R. K. Pandita, J. W. Shay, T. K. Pandita, Histone Modifications and DNA Double-Strand Break Repair after Exposure to Ionizing Radiations. *Radiat Res*. **179**, 383–392 (2013).

13. A. Battu, A. Ray, A. A. Wani, ASF1A and ATM regulate H3K56-mediated cell-cycle checkpoint recovery in response to UV irradiation. *Nucleic Acids Res*. **39**, 7931–7945 (2011).

14. J. v. Tjeertes, K. M. Miller, S. P. Jackson, Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *EMBO Journal*. **28**, 1878–1889 (2009).

15. M. K. Ayrapetov, O. Gursoy-Yuzugullu, C. Xu, Y. Xu, B. D. Price, DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. *Proceedings of the National Academy of Sciences*. **111**, 9169–9174 (2014).

16. L.-J. Mah, A. El-Osta, T. C. Karagiannis, γH2AX: a sensitive molecular marker of DNA damage and repair. *Leukemia*. **24**, 679–686 (2010).

17. I. S. Mehta, M. Kulashreshtha, S. Chakraborty, U. Kolthur-seetharam, B. J. Rao, Chromosome territories reposition during DNA damage-repair response (2013).

18. S. N. Fatakia, M. Kulashreshtha, I. S. Mehta, B. J. Rao, Chromosome territory relocation paradigm during DNA damage response: Some insights from molecular biology to physics. *Nucleus*. **8**, 449–460 (2017).

19. X. Kong, A. R. Ball, K. Yokomori, "The Use of Laser Microirradiation to Investigate the Roles of Cohesins in DNA Repair" in *Cohesin and Condensin: Methods and Protocols*, K. Yokomori, K. Shirahige, Eds. (Springer New York, New York, NY, 2017; https://doi.org/10.1007/978-1-4939-6545-8_14), pp. 227–242.

20. P. J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife*. **6**, e21856 (2017).

21. C. Zang, D. E. Schones, C. Zeng, K. Cui, K. Zhao, W. Peng, A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*. **25**, 1952–1958 (2009).

22. V. Dion, V. Kalck, A. Seeber, T. Schleker, S. M. Gasser, Cohesin and the nucleolus constrain the mobility of spontaneous repair foci. *EMBO Rep*. **14**, 984–991 (2013).

23. V. Dion, S. M. Gasser, Chromatin Movement in the Maintenance of Genome Stability. *Cell*. **152**, 1355–1364 (2013).

24. B. R. Schrank, T. Aparicio, Y. Li, W. Chang, B. T. Chait, G. G. Gundersen, M. E. Gottesman, J. Gautier, Nuclear ARP2/3 drives DNA break clustering for homology-directed repair. *Nature*. **559**, 61–66 (2018).

25. C. P. Caridi, C. D'Agostino, T. Ryu, G. Zapotoczny, L. Delabaere, X. Li, V. Y. Khodaverdian, N. Amaral, E. Lin, A. R. Rau, I. Chiolo, Nuclear F-actin and myosins drive relocalization of heterochromatic breaks. *Nature*. **559**, 54–60 (2018).

26. M. van Sluis, M. Ó. Gailín, J. G. W. Mccarter, H. Mangan, A. Grob, B. Mcstay, Human NORs, comprising rDNA arrays and functionally conserved distal elements, are located within dynamic chromosomal regions (2019), doi:10.1101/gad.331892.

27. B. Mcstay, Nucleolar organizer regions: genomic "dark matter" requiring illumination (2016), doi:10.1101/gad.283838.

28. J. S. Kim, T. B. Krasieva, V. LaMorte, A. Malcolm, R. Taylor, K. Yokomori, Specific Recruitment of Human Cohesin to Laser-induced DNA Damage. *Journal of Biological Chemistry*. **277**, 45149–45153 (2002).

29. N. Servant, N. Varoquaux, B. R. Lajoie, E. Viara, C.-J. Chen, J.-P. Vert, E. Heard, J. Dekker, E. Barillot, HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*. **16**, 259 (2015).

30. N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, E. L. Aiden, Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst*. **3**, 99–101 (2016).

31. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell*. **38**, 576–589 (2010).