DUX4-induced histone variants H3.X and H3.Y mark DUX4 target genes for expression

Rebecca Resnick

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Stephen J Tapscott, Chair

Barbara Wakimoto

Steve Henikoff

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

DUX4-induced histone variants H3.X and H3.Y mark DUX4 target genes for expression

Rebecca Resnick

Chair of the Supervisory Committee:
Professor Stephen J. Tapscott
Department of Neurology

The DUX4 transcription factor is briefly expressed in the cleavage-stage embryo where it induces an early wave of zygotic gene transcription; whereas its mis-expression in skeletal muscle causes the muscular dystrophy facioscapulohumeral dystrophy (FSHD). Here, we show that DUX4 induces the expression of the histones *H3.X* and *H3.Y*. These histone variants are incorporated throughout the body of DUX4-induced genes. Following a brief pulse of DUX4, these histones contribute to greater perdurance and to enhanced re-activation of DUX4 target gene expression. These findings provide a model for a DUX4-induced chromatin memory that facilitates the expression of its target genes subsequent to a brief pulse of DUX4 expression.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Stephen Justice Tapscott for his wonderful mentorship and guidance. His support and encouragement were always there when I needed them. I am also grateful to the members of the Tapscott Lab, past and present, for making the lab a fun and exciting place to think and learn. Thanks also to my committee for invaluable advice, questions, and discussion throughout the evolution of my project.

Chapter 1. INTRODUCTION

This chapter should be considered in the context of the following publications:

1) Jagannathan S, Shadle SC, **Resnick R**, Snider L, Tawil RN, van der Maarel SM, Bradley RK, Tapscott SJ, *Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells.* Hum Mol Genet, 2016. 25(20): 4419-4431.

Contributions: conceived and performed experiments, writing—review and editing

2) Campbell AE, Shadle SC, Jagannathan S, Lim JW, **Resnick R**, Tawil R, van der Maarel SM, Tapscott SJ, *NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins.* Elife, 2018. Mar 13;7.

Contributions: funding acquisition, investigation, writing—review and editing

3) Campbell AE, Belleville AE, **Resnick R**, Shadle SC, Tapscott SJ, *Facioscapulohumeral dystrophy: activating an early embryonic transcriptional program in human skeletal muscle.* Hum Mol Genet, 2018. R2: R153-R162.

Contributions: funding acquisition, writing—review and editing

1.1 CLINICAL FEATURES OF FSHD

Facioscapulohumeral muscular dystrophy (FSHD) is an incompletely-penetrant disease that can be autosomal dominant (FSHD1) or recessive (FSHD2) and is characterized by progressive muscle weakening, typically starting in the face, upper arms, and shoulders. FSHD is the third most common form of muscular dystrophy after Duchenne's and Becker's, and estimates suggest that it affects between 1/10,000-1/20,000 individuals worldwide [1-3]. Some common clinical signs of FSHD include scapular winging, rounded shoulders, hollowed pectoral muscles, Beevor's sign (raising of the belly button when a supine patient lifts their head, engaging the lower abdominal muscles), and asymmetrical weakness. Unlike many other muscular dystrophies, cardiac muscle is unaffected and respiratory muscles are rarely involved. Muscle weakness can spread to nearly all skeletal muscle, though, ultimately leading to about 20% of patients being wheelchair-bound [4]. Onset of weakness is typically in the 2nd or 3rd decade of life, though more severe pediatric FSHD has been reported [5], and some patients remain undiagnosed until late in life as progression can be quite slow. Findings on muscle biopsy vary, and could include immune cell infiltration, central nucleation, fibrosis, variable fiber size, and necrosis [6]. Non-muscular symptoms are rare and only present in severely-affected patients and include high frequency hearing loss and a Coats-like retinopathy [7-9].

Severity of the disease can vary widely within a family despite the presence of the same causative mutation, suggesting there are unidentified modifiers of the disease. In fact, roughly 20% of patients with causative mutations experience no symptoms at all [4]. Additionally, women are often affected less severely than men in the same family [7].

1.2 MOLECULAR BASIS OF FSHD

FSHD was mapped to the subtelomere of chromosome 4q in 1990 [10], but finding the causative gene proved difficult despite a limited number of candidates. In the past decade, the field has reached a consensus that mis-expression of the double homeobox transcription factor *DUX4* in skeletal muscle is the cause of FSHD [4]. In cultured FSHD muscle cells, there is an apparently stochastic de-repression of the *DUX4* retrogene resulting in a burst of *DUX4* expression from a minority of myonuclei. DUX4 protein can be detected in ~0.1% of myoblasts cultured from an FSHD muscle biopsy, and RNA can only be reliably detected through many cycles of nested PCR [11], explaining the difficulty in initially identifying *DUX4* as the FSHD causative gene.

DUX4 has a somewhat unusual genetic structure; it is a retrogene embedded in the D4Z4 macrosatellite repeat and multiple copies of the repeat can be organized into tandem arrays. Each copy of D4Z4 contains one copy of *DUX4* which lacks a strong poly-adenylation signal. In unaffected individuals, the tandem D4Z4 array on chromosome 4q contains 11 to 100 copies [12]. This number of repeats helps to maintain the silencing of the *DUX4* gene. In addition to chromosome 4q, there is a D4Z4 array on chromosome 10q as well as scattered individual or low-copy repeats elsewhere in the genome. Due to polymorphisms between 4q and 10q, we know that only 4q is responsible for FSHD, though it is possible that *DUX4* may be expressed from other loci in other contexts.

There are two FSHD forms, FSHD1 and FSHD2, that are genetically distinct but clinically indistinguishable, except possibly by inheritance pattern. Both forms converge on de-repression of DUX4 but through different genetic mechanisms. FSHD1, which affects ~95% of people with FSHD, is caused by a contraction of the D4Z4 repeat array on chromosome 4q to between 1 and 10 repeats [13, 14]. This contraction reduces the repression of *DUX4*, leading to rare bursts of

expression. Additionally, a downstream allele with a polyadenylation signal is required for *DUX4* expression. This polyadenylation signal in the 4qA161 allele is spliced to the *DUX4* transcript in a different exon, thereby sufficiently stabilizing the transcript to permit some DUX4 protein synthesis [15]. Since the downstream allele is quite close to the shortened array, it is genetically linked and FSHD1 is inherited in an autosomal dominant disease. FSHD2 is caused by mutations in genes that repress *DUX4* expression and does not require a contraction of the 4q D4Z4 array. The only mutations known to cause FSHD2 are in the structural maintenance of chromosomes flexible hinge domain-containing 1 (*SMCHD1*) and DNA methyltransferase 3B (DNMT3B) genes [16, 17], though these account for only about 80% of FSHD2 cases, with the remaining 20% caused by unknown mutations. In both forms of the disease, the D4Z4 is hypomethylated [18, 19], leading to de-repression of *DUX4*.

There are currently no treatments for FSHD, but high-throughput drug screens have identified several classes of promising compounds. JQ1, a BET bromodomain inhibitor, shows anti-DUX4 activity. Beta adrenergic agonists also decrease *DUX4* and target gene expression [20]. In addition to small molecule inhibitors, RNA-based therapies have been shown to be safe and effective for other forms of muscular dystrophy, and there are drugs currently on the market for spinal muscular atrophy and Duchenne muscular dystrophy. A similar approach using anti-sense oligo therapy for FSHD could be quite effective and several groups have been working to identify an effective nucleic acid-based drug [21-24].

1.3 DUX4 AND DEVELOPMENT

Recent work from our lab and others revealed an important role for *DUX4* in development. *DUX4* is expressed in a short burst in pre-implantation human embryos at the 4-cell stage when

the zygotic genome is first activated, followed by rapid repression and loss of *DUX4* mRNA by the 8-cell stage. The gene expression profile that characterizes the 4-cell stage is notable in that the DUX4 binding motif is the strongest signature in this group of genes [25]. Indeed, much of the DUX4-induced transcriptional profile identified from studies in various FSHD models closely recapitulates this zygotic genome activation (ZGA) profile, further suggesting that DUX4 is critical for this initial activation of transcription. Interestingly, the mouse homologue of *DUX4*, *Dux*, is also expressed in a burst at ZGA, which happens at the 2-cell stage in mouse, and the transcriptome of Dux in mouse cells is greatly enriched in the ZGA totipotency signature [26]. Further studies in mice have shown that Dux is required for progression of embryogenesis beyond the morula stage , showing that Dux and likely DUX4 are required for early development and may direct the ZGA process, ensuring expression of some of the first embryonic genes.

1.4 DUX4 PATHOGENESIS

DUX4 activates a downstream program characterized by expression of a totipotency signature as described above, as well as many repetitive elements including HSATII repeats, endogenous retroviruses including the HERVL family, and repetitive gene families such as the PRAMEF cancer testis antigens which are normally expressed in testis and aberrantly expressed in some cancers [27, 28]. DUX4 also inactivates the nonsense-mediated decay (NMD) pathway through increased degradation of UPF1 [29]. This acts as a positive feedback loop to stabilize the *DUX4* mRNA, itself a target of NMD since the polyadenylation signal comes from an exon downstream of the coding sequence, a feature that triggers NMD. In addition to increasing *DUX4* levels, inactivation of NMD leads to accumulation of aberrant mRNAs which could be translated, leading to novel peptides. Additionally, *DUX4* expression leads to accumulation of double-stranded RNAs

that would normally be expected to activate the innate immune response [30]. However, *DUX4* also downregulates the immune response [28] through unknown mechanisms that are currently being studied. Knockdown of *RNASEL* and *PKR* were shown to rescue *DUX4* toxicity in an siRNA screen [30], suggesting that this immune pathway contributes to *DUX4*-mediated cell death.

The mechanism of *DUX4* toxicity has been studied extensively and it seems that there may be many pathways contributing rather than a single cause of death: immune response, RNA processing, proteotoxic stress, changes in extracellular matrix, and others. In the context of FSHD, cytotoxic T cells and other components of the patient's immune system could also be targeting these cells, as they are likely to be presenting novel antigens such as cancer testis antigens (many of which are *DUX4* targets) or protein products of the mRNAs that would have been targeted by NMD.

DUX4 is repressed through many different mechanisms. Recent work using locus-specific proteomics in control myoblast cells revealed two complexes in addition to SMCHD1 that repress *DUX4* expression: the Nucleosome Remodeling and Deacetylase (NuRD) complex and the CAF-1 complex [31]. In addition to protein-based repression, small RNAs are produced from the D4Z4 locus and may also serve to repress *DUX4* expression [22]. Additionally, the repeat has a very high GC content and many CpGs, making it a prime target for DNA methylation. The subtelomeric location would also be predicted to repress the region through telomere position effect. One study has even shown that telomere length can affect *DUX4* expression, with shorter telomeres leading to higher levels of *DUX4* in culture [32].

1.5 MODEL SYSTEMS OF FSHD

The repressive mechanisms described above, some intrinsic to D4Z4 or the *DUX4* sequence itself, combined with the rapid toxicity of *DUX4* expression and low levels of the protein in FSHD patient samples, have made it challenging to generate model systems to study DUX4 biology that allow the experimenter to control timing or levels of *DUX4* expression. A summary of model systems is diagramed in Figure 1.1. Delivering a *DUX4* transgene to wild-type myoblasts using lentivirus and studying cultured FSHD muscle cells has led to many advances in understanding this transcription factor and associated disease, but the toxicity has made longer-term studies difficult. Cultured cells derived from FSHD muscle have the disadvantage that only a small percentage, ~0.1%, of cells express *DUX4* at any given time [11]. This greatly dilutes the signal in assays that require a large populations of cells, though it is possible to sort out *DUX4*-expressing cells from these populations to assay [33]. The lentiviral system provides some temporal control of *DUX4* expression and the transcriptome of these cells is comparable to the sorted FSHD cells [34]. However, neither of these systems allows much control over level or precise timing of *DUX4* expression. Attempts to develop an inducible system were initially thwarted by lack of expression of the *DUX4* transgene upon induction, likely due to some of the silencing mechanisms described above. Codon-altering of the *DUX4* sequence allowed reduction of CpG dinucleotides and overall GC content, as well as interrupting potential seed sites for small RNAs, while maintaining the protein sequence of DUX4. Using this codon-altered *DUX4* (DUX4-CA) in a doxycycline-inducible lentiviral vector led to expression of the transgene and cell death upon induction, unlike the wild-type *DUX4* sequence which did not induce full-length *DUX4* mRNA [34]. However, it was necessary to isolate clonal populations with a single integration of this transgene and maintain the cells under constant selection to avoid silencing of the transgene or loss of the inducing cells

from the population. Induction of *DUX4* with doxycycline in this system leads to activation of the *DUX4* transcriptome, equivalent to both the lentiviral *DUX4* and sorted FSHD cell systems [34].

In addition to cell-based models, which are used in the work described in the following chapters, there is great interest in animal models of FSHD for learning more about the disease on a tissue or organism scale, and for testing potential therapeutics. Similar challenges have presented themselves here, with mice expressing very low levels of *DUX4*, perhaps because more robust expression was incompatible with life [35]. More recently-described mouse models with several layers of inducibility have proved more effective at titrating *DUX4* levels, as detailed in Figure 1.1 and described more fully in [36]. Even so, the small amounts of leaky *DUX4* lead to muscle and other phenotypes even without induction in some models. A more recent understanding of the differences between the DUX4 and mouse Dux transcriptomes [26] suggests that overexpression of mouse *Dux* in mice may better recapitulate the full transcriptional profile and more accurately model the disease process of FSHD than expression of the human *DUX4* in mouse models.

1.6 HISTONES AND HISTONE VARIANTS

Histones play a major role in the regulation of transcription in all eukaryotes. Every human cell contains nearly two meters of DNA, compacted and organized within the nucleus by histones and other proteins. The core histones H2A, H2B, H3, and H4 form nucleosomes by assembling octomers composed of two copies each of H2A and H2B, and two dimers of H3/H4. 147 base pairs of DNA wraps 1.7 times around each nucleosome [37], with inter-nucleosome spacing dependent on many factors, especially genomic context and transcriptional activity. The linker histone, H1, is not part of the core nucleosome but can further condense DNA by compacting nucleosomes.

Histones can serve to enhance or repress transcription at their site of incorporation, and different variants or post-translational modifications can alter a nucleosome's effects on gene expression.

Histone H3 has several variants including the replication-dependent H3.1 and H3.2 (together called canonical H3), and the replication-independent H3.3. These histones have strong selection bias, with perfect conservation of canonical H3 and H3.3 from *Drosophila* to *Homo sapiens* (Figure 1.2A and B, respectively). Additionally, these variants are quite similar to each other. H3.1 and H3.2 differ by only a single amino acid, while H3.3 has 4-5 amino acid differences from canonical H3. Canonical H3 is expressed only during S-phase and is incorporated into newly-synthesized DNA using the CAF1 chaperone complex. H3.3, in contrast, is expressed throughout the cell cycle and uses two different chaperones [38]. HIRA incorporates H3.3 predominantly into expressed regions of the genome [39, 40], whereas the ATRX/DAXX complex incorporates H3.3 into repressed regions such as telomeres [41-43].

Histone variants play important roles in different tissues and different stages of development. H3.3 has been shown to be critical at several stages of early mammalian development, such as silencing of some retroelements in mouse embryonic stem cells [44], and knockdown of H3.3 causes arrest of these pre-implantation embryos at the morula stage [45]. Knockout of H3.3 in mice causes heterochromatin dysfunction and karyotype abnormalities, providing further support for its role in pre-implantation development [44, 46-48].

In addition to these well-conserved variants, there are more divergent H3 variants (see Table 1). These tend to have specific, more limited roles in a cell or tissue type. CENPA, for example, is incorporated only into centromeres [49]. H3t is expressed in testis and forms unstable nucleosomes and is thought to facilitate protamine replacement of histones during spermiogenesis [50]. Several other H3 variants in humans have been identified but poorly characterized.

In addition to histone variants that can affect nucleosome stability and chromatin compaction, histones often have post-translational modifications that can also interact with or recruit other proteins. Many modifications are located in the N-terminal histone tail, an unstructured domain that protrudes from the nucleosome. Generally, acetylation of these residues is correlated with active chromatin and methylation with repression, but this is not always the case. Other modifications that have been reported include phosphorylation, sumoylation, and ubiquitination.

1.7 H3.X AND H3.Y

This work focuses on two recently-described histone variants, H3.X and H3.Y, which are ~80% identical to H3.3. Of the 4-5 amino acids that distinguish H3.3 from canonical H3, H3.X/Y share all but one and are thus considered H3.3-like variants. H3.X and H3.Y are present in primates and expressed in human testis and brain, as well as some tumors and cancer cell lines [51]. Both variants have been validated as histones and are stably incorporated into chromatin [51-53], but little is known about their *in vivo* distributions and roles.

Recently-published work [53] showed that H3.Y exclusively uses the HIRA chaperone complex and does not associate with the ATRX/DAXX complex. This would suggest that H3.Y and presumably, H3.X as well, would be incorporated into active regions of the genome. Two overexpression studies using tagged H3.Y [52, 53] have used chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) to map incorporation, and found that this was indeed the case—H3.Y was incorporated predominantly into highly-expressed genes and enriched at the transcriptional start sites (TSS). Additionally, Zink *et al*. (2017), showed that H3.Y incorporation overlapped with active histone mark H3K4me3 and was excluded in regions with the repressive H3K9me3 mark.

Other unique properties of H3.X/Y are suggested by their sequence. Substitution of individual H3.3 amino acids with some H3.X/Y residues has been shown to de-repress silenced repetitive elements in non-mammalian systems. For example, S10A and K14Q mutations in H3.3 show a loss of DNA methylation at a methylated transgene [54]. R53, K79, K115, and K122 substitutions, all residues where H3.X/Y differ from H3.3, show a loss of repetitive DNA silencing at telomeres and/or rDNA [55]. A glutamine at position 14 as in H3.X/Y can mimic K14 acetylation, a mark associated with gene activation and which recruits bromodomain proteins, suggesting that H3.X/Y may constitutively recruit these activators to genes where they are incorporated. Further supporting the idea that H3.X/Y facilitate gene expression, nucleosomes with H3.Y have been shown to have a looser, more open chromatin conformation than H3.3 nucleosomes though altered DNA contacts and exclusion of the H1 linker histone [52].

While these histone variants have been found only in primates thus far, many new H3 variants have been described in mice but not yet characterized [56]. Since the residues that give H3.X/Y unique properties and that mediate their function have not been fully elucidated, a simple comparison of sequences cannot reveal the presence or absence of H3.X/Y-like histones in other organisms and functional studies are necessary to determine whether these are truly primate-specific.

Chapter 2 of this work explores the role of H3.X/Y in the contexts of FSHD and zygotic genome activation. It addresses the contrast between a *DUX4* burst in the embryo leading to healthy development whereas sustained expression in culture or endogenous expression in FSHD are cytotoxic. The results in this chapter support a mechanism whereby a burst of *DUX4* establishes a chromatin memory through incorporation of histone variants H3.X and H3.Y, allowing this

program to persist after the transcription factor itself is gone. Chapter 3 examines the relationship between *DUX4* expression and telomere elongation through the Alternative Lengthening of Telomeres (ALT) pathway. We propose a negative feedback loop whereby shortened telomeres de-repress *DUX4,* activating ALT through expression of *ZSCAN4*, *H3.X/Y*, and degradation of UPF1, ultimately lengthening telomeres and repressing *DUX4* expression, but find no experimental support for this model. Chapter 4 summarizes these findings and their impact and proposes future studies to build on this work.

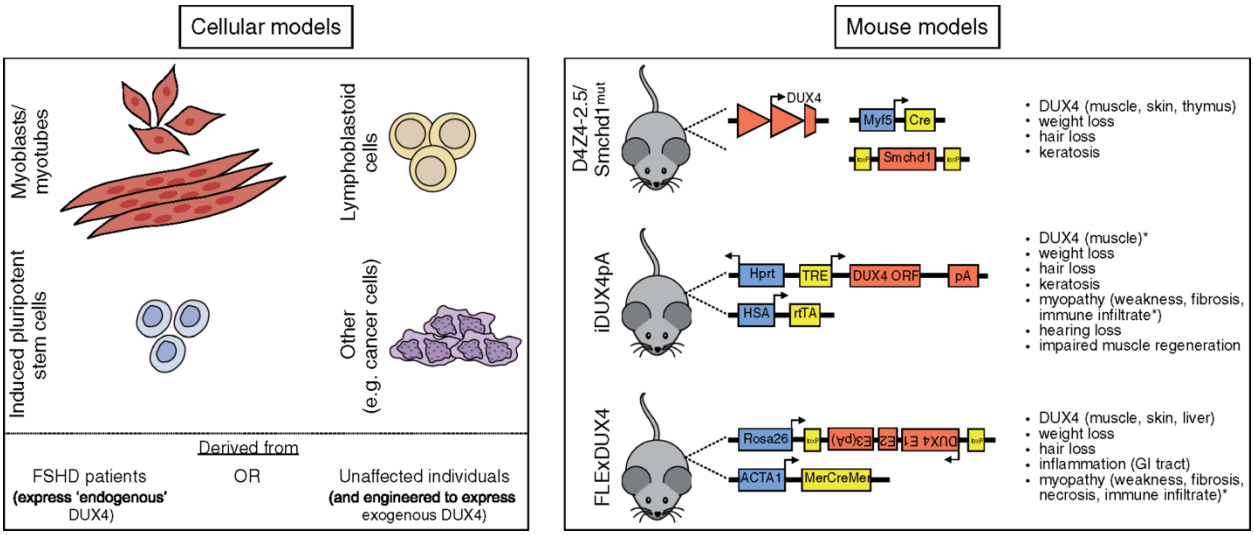


Figure 1.1: Cellular and murine models of FSHD. Figure from [36]

The normal and pathological functions of DUX4 are currently being investigated using the depicted cell (left) and murine (right) models. Note that DUX4 can be exogenously expressed in cell culture systems via transfection, transduction or stable integration of inducible transgenes. TRE, tetracycline response element; ORF, open reading frame; pA, polyadenylation signal; rtTA, reverse tetracycline-controlled transactivator; E1, exon 1; E2, exon 2; E3, exon 3; MerCreMer, Cre recombinase protein fused to two mutant estrogen-receptor ligand-binding domains; *, phenotype only apparent when DUX4 expression is induced.

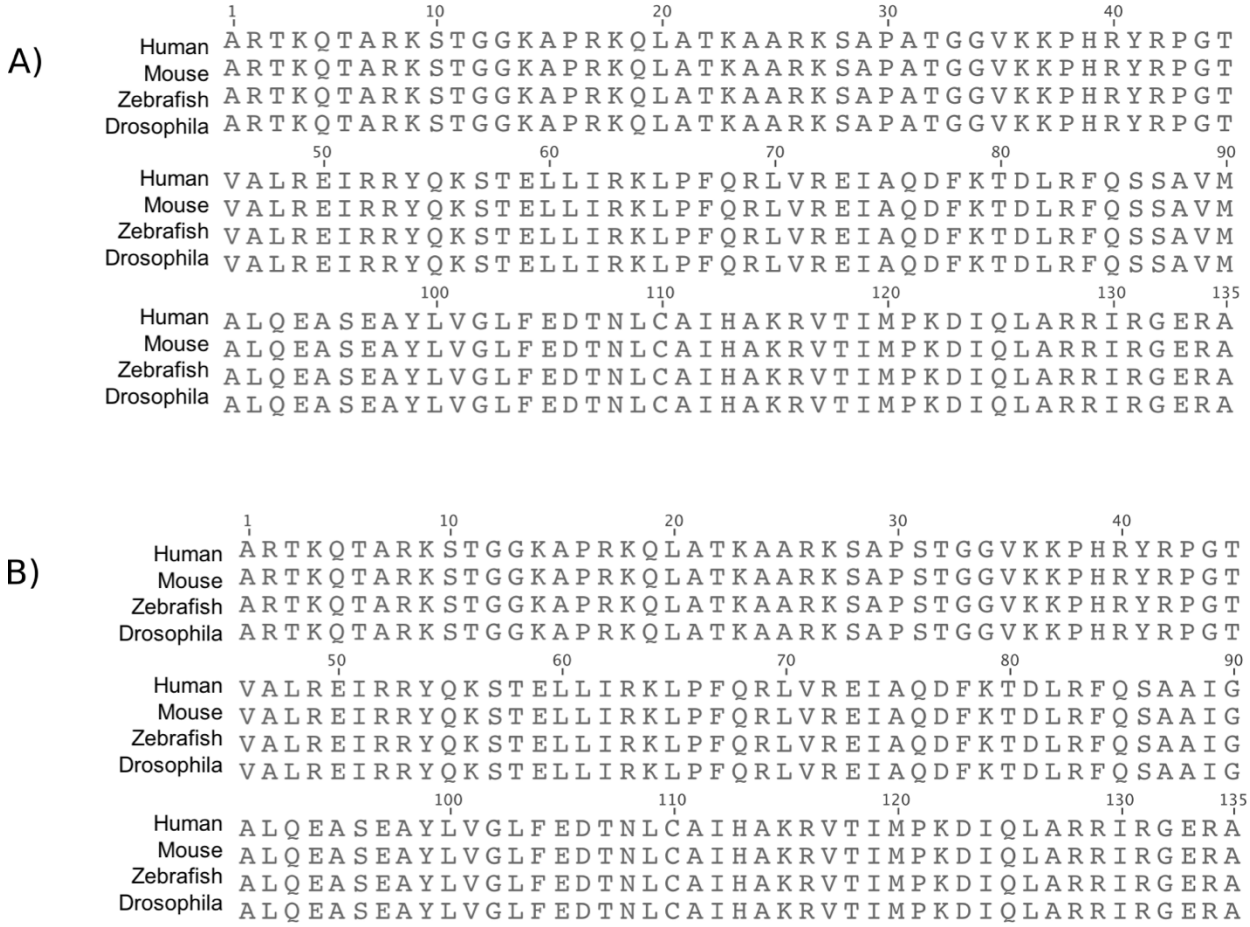


Figure 1.2: Canonical H3 and H3.3 are perfectly conserved from *Drosophila* to humans

Protein alignment of canonical H3.2 (A) and H3.3 (B).

Table 1: Human H3 variants

Name	**% amino acid identity to H3.1**
H3.1	100%
H3.2	99%
H3.3	96%
H3.t	97%
H3.5	93%
H3.X	79%
H3.Y	79%
CENPA	51%

Chapter 2. DUX4-INDUCED HISTONE VARIANTS *H3.X* AND *H3.Y* MARK *DUX4* TARGET GENES FOR EXPRESSION

A version of this chapter is currently under review at Developmental Cell as:

DUX4-induced histone variants H3.X and H3.Y mark DUX4 target genes for expression
Rebecca Resnick, Chao-Jen Wong, Peter J. Skene, Sandra B. Hake, Steven Henikoff, Silvère M. van der Maarel[7] and Stephen J. Tapscott

Contributions: conceived and performed experiments, wrote manuscript, and secured funding.

2.1 INTRODUCTION

Mis-expression of the double homeobox transcription factor *DUX4* in skeletal muscle is the cause of facioscapulohumeral muscular dystrophy (FSHD) [4]. In cultured FSHD muscle cells, there is an apparently stochastic de-repression of the *DUX4* retrogene resulting in a burst of *DUX4* expression from a minority of myonuclei. In contrast to the toxicity of *DUX4* expression in skeletal muscle cells, *DUX4* is normally expressed in the early cleavage-stage embryo where it regulates zygotic gene activation [25, 26, 57]. In both of these cases, the burst expression of *DUX4* results in a perdurant developmental or pathological phenotype. This could be due to the initiation of a transcription factor cascade or an induced chromatin memory, or both.

Histone variants play critical roles in early development, such as the recently-demonstrated requirement for *H3.3* in paternal genome activation in mouse preimplantation embryos [45] as well as for retroelement silencing in embryonic stem cells [44]. While canonical H3 is only produced during S phase and incorporated into newly synthesized DNA, H3.3 and H3.3 variants are made throughout the cell cycle [38] and use either the ATRX/DAXX complex to incorporate into repressed regions [41-43] or the HIRA chaperone to incorporate into transcriptionally active DNA [39, 40]. Canonical H3 and H3.3 are extremely well conserved among species and differ from each other by only 4 or 5 amino acids. More divergent histone variants, such as *CENP-A* and *H3t*, have more specialized roles in designating centromeres or facilitating the transition from histones to protamines during spermiogenesis, respectively [49, 50].

The histone variants *H3.X* and *H3.Y* were recently identified in the human genome as a multicopy gene family with copies interspersed with the *TAF11-Like* macrosatellite repeat [51]. Biochemical studies of H3.Y nucleosomes showed that they resulted in a more relaxed chromatin configuration than H3.3 nucleosomes, excluded linker histone H1, were incorporated at

transcriptional start sites (TSSs) of actively-transcribed genes, and that H3.Y used the HIRA chaperone [52, 53]. These features suggests that H3.Y, and possibly H3.X as well, might be incorporated at active genes and contribute to the maintenance of an open chromatin conformation.

Here, we show that DUX4 induces the expression of *H3.X* and *H3.Y* and that these histone variants are incorporated in highly transcribed *DUX4* target genes. A short pulse of *DUX4* in wild-type myoblasts that mimics the developmental expression pattern was not cytotoxic but also induced *H3.X/Y* expression. We also show that a second pulse of *DUX4* resulted in greater perdurance of *DUX4* target gene expression and enhanced activation, dependent on the presence of H3.X/Y. Together, these results indicate that incorporation of these histone variants imposes a chromatin memory at *DUX4*-induced genes that contributes to prolonged expression and sensitization to subsequent induction. During development, this might facilitate the perdurance of *DUX4*-regulated genes after the initial cleavage-stage pulse of *DUX4*, whereas in FSHD muscle, the increased sensitivity of *DUX4* targets to a subsequent pulse of *DUX4* might amplify *DUX4*-target gene expression throughout the post-mitotic multinucleated muscle fiber and contribute to toxicity.

2.2 RESULTS

2.2.1 *DUX4 induces the expression of histone variants H3.X and H3.Y*

To study the transcriptional network activated by DUX4, we have used a well-characterized cellular model system of human myoblasts with a doxycycline-inducible *DUX4* transgene (MB135iDUX4 cells) [34]. Induction of *DUX4* in these cells has been shown to induce many of the *DUX4*-regulated genes belonging to the transcriptional program characteristic of the 4-cell cleavage stage embryo [25, 26] and to recapitulate the transcriptional consequences of the

endogenous *DUX4* in FSHD cells [34, 58]. As such, it is a validated model system for the identification of *DUX4*-regulated genes and the biological consequences of *DUX4* expression.

Further analysis of our previous RNA-seq datasets [34] revealed high expression of unannotated transcripts that were not detected in the absence of *DUX4* and were located in the region of the *TAF11-Like* macrosatellite repeat array on chromosome 5. Some of these sequences corresponded to histone variants *H3.X* and *H3.Y* [51], as well as a previously unreported related sequence we designated *H3.Z*. Compared to *H3.X* or *H3.Y*, *H3.Z* has a frameshift mutation that is predicted to disrupt the histone fold and results in a longer protein (Figure S2.1A,B). DUX4 ChIP-seq [28] showed DUX4 binding near the TSS for the *H3.X*, *H3.Y*, and *H3.Z* transcripts (Figure 2.1A and Figure S2.1C), suggesting they are direct targets of DUX4.

Western analysis with an antibody to an epitope shared by H3.X, H3.Y, and H3.Z [51] detected both H3.X and H3.Y protein between 8 and 16 hours after DUX4 induction in the MB135iDUX4 cells, with levels increasing up to 24 hours (Figure 2.1B). H3.X protein is slightly larger than H3.Y as shown by *in vitro* translated protein, generating a closely-spaced doublet. In addition to H3.X and H3.Y, an additional band migrating at the size of H3.Z was also detected. ZSCAN4, a well-characterized DUX4 target, was detected slightly earlier than H3.X/Y/Z, at 8 hours after *DUX4* induction (Figure 2.1B).

To determine whether endogenous DUX4 also regulated *H3.X/Y*, we used myoblast cell lines derived from individuals with FSHD1 and FSHD2, the two forms of the disease, that show sporadic de-repression of *DUX4* in ~0.1% of cells [11], with increasing frequency and amount of *DUX4* expression upon differentiation to myotubes [11, 59, 60]. RT-qPCR detected elevated levels of *DUX4*, *H3.X*, *H3.Y*, and *H3.Z* in both FSHD1 and FSHD2 myoblast cultures but not in controls, with increased expression in the FSHD myotubes (Figure 2.1C, Figure S2.1D).

Immunofluorescence showed strong nuclear H3.X/Y staining in FSHD myoblasts and myotubes, which also co-localized with DUX4 staining. No control myoblasts or myotubes stained positively for either H3.X/Y or DUX4 (Figure 2.1D, E).

During development, a burst of *DUX4* expression can be detected at the cleavage stage [25]. Re-analysis of this human embryo RNA-sequencing [25] demonstrated that expression of *H3.X* and *H3.Y* coincided with the period of *DUX4* expression (Figure 2.1F). Together, these data identify *H3.X* and *H3.Y* as genes regulated by *DUX4* and show that they are co-expressed with the endogenous *DUX4* in biologically relevant contexts, i.e., FSHD muscle cells and the cleavage-stage human embryo. In addition, we have previously shown *DUX4* expression in the testis [11] where *H3.X/Y* have also been reported to be expressed [51].

2.2.2 *H3.X/Y are incorporated in expressed regions of the genome*

Previous studies using exogenously-expressed *H3.Y* showed that it was incorporated into nucleosomal DNA by the HIRA chaperone complex [53] at the TSSs of highly-expressed genes [52, 53]. To determine the incorporation pattern of the endogenous H3.X/Y in *DUX4*-expressing cells, we induced *DUX4* in MB135iDUX4 myoblasts and used the antibody-targeted MNase CUT&RUN assay [61, 62] to map H3.X/Y incorporation genome-wide (schematic in Figure 2.2A). H3.X/Y localized in domains ranging from 500bp to nearly 100kb (Figure S2.2), with ~75% of domains overlapping genic regions and the remaining 25% intergenic (Figure 2.2B), similar to what was seen using ChIP-seq in a previous over-expression study [53].

H3.X/Y were preferentially incorporated at more highly-expressed genes (Figure 2.2C) and robustly-induced *DUX4* target genes (Figure 2.2D). Similarly, 80% of intergenic *DUX4*-induced lncRNAs (n = 380) showed H3.X/Y incorporation, and presence of an intergenic H3.X/Y domain was correlated with induction by *DUX4* as compared to random intergenic bins of comparable

length (Figure 2.2E). Specific examples of *DUX4* targets and constitutively-expressed genes are shown in Figure 2.2F and G, respectively. However, when averaging over larger sets of genes, it is evident that in constitutively-expressed genes the H3.X/Y domains showed enrichment at the TSS (Figure 2.2H, top), whereas at *DUX4* target genes the H3.X/Y enrichment was throughout the full transcribed region of the gene (Figure 2.2H, bottom).

Together, these data show that endogenous H3.X/Y were incorporated in transcribed regions of the genome, both throughout gene bodies and in intergenic regions. This included incorporation of these histones at *DUX4* target genes when they were highly induced upon *DUX4* expression. As has been previously reported, incorporation was enriched at the TSS of constitutively-expressed genes [52, 53]; surprisingly, however, *DUX4*-induced genes showed incorporation of H3.X/Y throughout the gene body.

2.2.3 *A pulse of DUX4 activates target gene expression with little cell toxicity*

Mis-expression of *DUX4* induces apoptotic cell death in nearly every cell tested [33, 63, 64], yet it is expressed in the germline and early embryo [11, 25], and the Dux mouse ortholog was shown to be necessary for early embryo development [57]. Unlike in cell culture models where toxicity is seen with continuous *DUX4* expression, *DUX4* is only briefly expressed in the early embryo. To test whether cellular toxicity required constitutive expression of *DUX4*, we treated MB135iDUX4 myoblasts with a 4-6 hour "pulse" of doxycycline to induce transient *DUX4* expression. In contrast to the continuous expression of *DUX4* that resulted in the death of nearly the entire cell population by 48 hours, a 4 hour pulse of doxycycline resulted in very little cell death or toxicity at the 48 hour time point (Figure 2.3A). Immunodetection showed that the short pulse of doxycycline induced DUX4 expression in nearly all of the nuclei (Figure 2.3B) and Western analysis showed that the proteins for DUX4 and its targets ZSCAN4 and H3.X/Y/Z were

each detectable at similar timepoints as for continuous DUX4 expression and persisted for at least 24 hours (Figure 2.1B and Figure 2.3C). Remarkably, the H3.X and H3.Y proteins remained detectable even four days after the initial pulse (Figure 2.3D). To further verify this method, we compared the transcriptome after a pulse of *DUX4* using RNA-seq and found that it mostly recapitulated the transcriptional program changes characterized by the continuous expression of DUX4 seen in [34] ($R^2 = 0.63$) (Figure 2.3E). These results suggest that the duration of DUX4 expression might be the major determinant of toxicity rather than the cell type, and that a brief pulse of DUX4 results in a robust activation of its transcriptional program and prolonged presence of H3.X/Y proteins.

2.2.4 *H3.X/Y incorporation increases the perdurance and re-expression of DUX4-target genes*

The persistence of H3.X/Y protein for several days after a pulse of *DUX4* suggested that these histone variants could provide a chromatin memory at the *DUX4* target genes. To determine whether H3.X/Y incorporation at *DUX4* target genes could increase perdurance of gene expression and/or facilitate subsequent gene expression, we induced a pulse of *DUX4* followed by a second pulse two days later when H3.X/Y would be present in chromatin. We compared gene expression with or without H3.X/Y by using siRNAs that targeted both *H3.X* and *H3.Y* (siXY) or control siRNA (siControl) (Figure 2.4A-C). RNA for the *DUX4* targets *ZSCAN4* and *TRIM43*, both of which overlap with H3.X/Y domains (see Figure 2.2F), were robustly induced 24 hours after the first pulse of *DUX4* expression with decreased levels by 48 hours after the pulse. *H3.X/Y* knockdown did not alter the level of induction but did result in decreased perdurance of the RNA expression for both target genes (Figure 2.4A-C). A second pulse of *DUX4* on day 2 showed a super-induction of *ZSCAN4* and *TRIM43* 24 hours later (day 3) with RNA levels roughly 6-fold

greater than after the single pulse. *H3.X/Y* knockdown prevented this super-induction after the second pulse despite similar, or slightly higher, levels of *DUX4*. Like the first pulse, the expression of *ZSCAN4* and *TRIM43* showed greater perdurance of expression in the siControl cells compared to the siXY cells. This suggests that incorporation of H3.X/Y facilitated the increased inducibility of these genes and the persistence of their RNA expression. In contrast to these *DUX4* target genes, constitutively-expressed genes that incorporated H3.X/Y, e.g., *XRCC5* and *TPT1* were unaffected by H3.X/Y knockdown (Figure S2.3A,B).

Expanding these experiments genome-wide with RNA-seq 24 hours after pulsing with *DUX4* showed that H3.X/Y were necessary to sensitize nearly all *DUX4*-induced genes for subsequent super-induction: a single pulse of *DUX4* showed induction of targets (Figure 2.4Di) that was not affected by *H3.X/Y* KD (Figure 2.4Dii); whereas the super-induction of *DUX4* target genes with the second pulse (Figure 2.4Diii) was prevented by *H3.X/Y* knockdown (Figure 2.4Div). Globally, non-*DUX4* targets with H3.X/Y incorporation were unaffected by the pulses and KD treatments (Figure 2.4E). These results demonstrate that H3.X and H3.Y were incorporated into DUX4-induced genes and that this enhanced subsequent expression of these genes.

2.3 DISCUSSION

Together, these data show that *H3.X* and *H3.Y* are direct targets of *DUX4*; are induced co-incident with *DUX4* expression in human embryos and in FSHD muscle cells; are incorporated at genes induced by *DUX4*; and that their incorporation both promotes the perdurance of *DUX4*-target gene expression and facilitates their subsequent induction. Previously, other groups have described important biochemical properties of H3.X and H3.Y, mostly through *in vitro* and mis-expression studies [51-53, 65]. In particular, the *in vitro* biochemical studies predicting a more

relaxed chromatin with less efficient H1 binding [52] are consistent with our *in vivo* studies showing enhanced transcription of genes incorporating H3.X/Y. In this way our study builds on these important advances and characterizes the biological consequences of endogenous H3.X/Y expression with implications for roles in both development and disease. Our work also uncovered an additional expressed sequence from the *H3.X/Y* repetitive region, *H3.Z*, which will require further characterization to determine whether it represents a *bona fide* nucleosomal histone or has the capacity to interact with H4 or HIRA. With this in mind, we caution that although we attribute the findings in our current study predominantly to the more abundant H3.X and H3.Y, H3.Z is also targeted by the siRNAs against H3.X and H3.Y. Hence, it is possible that H3.Z might also contribute to either the perdurance or enhanced target activation if it is similarly incorporated into nucleosomes (Figure 2.3D).

In contrast to DUX4 target genes, constitutively-expressed genes were mostly unaffected by incorporation of H3.X/Y. This was also associated with a different pattern of incorporation. Whereas H3.X/Y were incorporated throughout the full-length of the *DUX4* target gene, incorporation at constitutively-expressed genes was largely flanking the TSS, as was previously reported in studies that mis-expressed H3.Y in HeLa cells [52, 53]. It is possible that this different pattern of incorporation represents the difference between a constitutively-expressed gene and an induced but previously silent gene. Because the majority of genes robustly regulated by *DUX4* are not expressed in myoblasts, their nucleosomes would likely have the canonical H3 histones that would be replaced with H3.X/Y and H3.3 when actively transcribed. In contrast, constitutively-expressed genes would already have replaced the canonical H3 histones with H3.3 and there might be less turnover in the gene body compared to newly-induced genes, restricting the H3.X/Y incorporation to the nucleosomes flanking the TSS that undergo more rapid turnover. Although

speculative, this difference might account for the specificity of the perdurance and hyper-induction to the *DUX4*-induced genes that incorporate H3.X/Y.

Another important finding of our study is that a transient burst of *DUX4* expression is not toxic to skeletal muscle cells. Using a well-characterized cell culture model of FSHD [34] with inducible *DUX4*, it was possible to mimic the kinetics of embryonic *DUX4* expression with short pulses, leading to expression of targets including *H3.X/*Y without inducing cell death. This made it feasible to study the effects of *DUX4* expression over several days, which has not been previously possible. This finding has interesting implications for both stem cell biology and FSHD muscular dystrophy. In stem cell biology, a brief expression of *DUX4* similar to that in the early embryo occurs in a subset of embryonic stem (ES) cells or induced pluripotent (iPS) cells, where it also induces the totipotent transcriptional program associated with the naïve state. Therefore, our findings also suggest that the difference between the early embryo or ES/iPS cells that survive *DUX4* expression and FSHD cells or other somatic cells that die when *DUX4* is constitutively expressed might be the duration of expression rather than another protective factor in the early stem cells.

In this regard, skeletal muscle might be particularly susceptible to repeated bursts of *DUX4* expression. As our study has demonstrated, incorporation of H3.X/Y in *DUX4* target genes increased the perdurance of their expression and enhanced subsequent activation (Supplemental Figure S2.4A). As skeletal muscle is postmitotic and multinucleated, the H3.X/Y incorporated following a pulse of *DUX4* could create a prolonged sensitivity for *DUX4* target expression, both in the nucleus that initially expressed *DUX4* and in the adjacent nuclei that received DUX4 and H3.X/Y from their shared cytoplasm. In this model, stochastic bursts of *DUX4* in different nuclei in a myotube might result in progressive accumulation of H3.X/Y in an expanding nuclear domain,

and progressive enhancement of expression of *DUX4* target genes (Figure S2.4B, right), resulting in the toxicity seen with the constitutive expression of *DUX4*. In contrast, incorporation of H3.X/Y following the embryonic burst of *DUX4* would be diluted by subsequent cell divisions (Figure S2.4B, left). Initially, this could result in greater perdurance of the totipotentency program but ultimately not reach toxic levels.

2.4 MATERIALS AND METHODS

Accession codes

MB135iDUX4 pulsed RNA-seq data and CUT&RUN data have been submitted to the Gene Expression Omnibus (GEO) and are awaiting accession numbers. Other datasets used in this study include DUX4 ChIP-seq (GEO accession number GSE33838), continuous DUX4 RNA-seq (GEO accession number GSE85461), and early embryo RNA-seq (GEO accession number GSE72379). The H3.Z sequence has been submitted to GenBank and is awaiting an accession number.

Cell culture

FSHD, control, and MB135iDUX4 myoblasts (described in Jagannathan *et al*., 2016) were grown in F10 growth medium (Gibco Life Technologies) supplemented with 10% FBS (HyClone), 1% pen/strep (Life Technologies), 10pg/mL fgf (Life Technologies), 1μM dexamethasone (Sigma), and 2μg/mL puromycin or 10μg/mL blasticidin as appropriate to maintain inducibility of the DUX4 transgene. FSHD and control cells were differentiated by growing to confluence and changing to differentiation medium: DMEM (Gibco) with 1% horse serum (Life Technologies), 1% pen/strep, and 10 μM each transferrin and insulin for 72 hours. Pulsed MB135iDUX4 cells were treated with 1μg/mL doxycycline in growth medium for 6 hours for RNA-seq experiments

or 4 hours for all other experiments, rinsed with PBS, and fresh growth medium added. Cells with continuous induction were treated with 1μg/mL doxycycline in growth medium overnight or as specified. For Figure 1C, ControlA = 54-1 [66], ControlB = MB135 [28], FSHD1 = 54-2 [66], and FSHD2 = MB200 [11].

siRNA treatment

siRNAs for H3.X/Y were designed using the Dharmacon siDESIGN Center. Two siRNAs that gave >90% knockdown were pooled for all experiments to minimize off-target effects with a non-targeting siRNA (Dharmacon D-001210-02-05) used as a control. For pulse experiments, cells were transfected with 50 pg siRNA in OPTIMEM with 7.5μL Lipofectamine RNAiMax 16 hours before each doxycycline treatment.

CUT&RUN

MB135iDUX4 cells were treated with or without doxycycline for 18 hours before harvesting. Protocol was followed as in [62] with modifications to scale up for 5 million cells per sample. 100μL beads were used per sample and wash/incubation volumes were increased to 300-500μL. 0.05% digitonin was used in the wash buffer. 15μL H3.X/Y primary antibody (clone 8H6-2111, Active Motif 61161) was used per H3.X/Y sample and incubated for 2 hrs. 25μL rabbit anti-rat secondary (ab6703) was added to each sample for 1 hr. After MNase digestion, fragments were liberated for 20 minutes and DNA was then purified using the phenol/chloroform method. Three biological replicates were performed for H3.X/Y samples, both with and without dox, and one replicate for each IgG condition.

CUT&RUN domain calling

Our CUT&RUN data consists of 25 bps long, paired-end reads with average fragment length of 180 bps. We aligned reads to both human genome hg38 and spike-in genome dm6 using Bowtie2-2.2.6 with the following comment: bowtie2 --local --very-sensitive-local --no-unal --no-mixed --no-discordant -q --phred33 -I 10 -X 700.

Since H3.X/Y were incorporated in many large regions within gene bodies, conventional peak calling algorithms such as MACS2 were not applicable. We thus developed an in-house R/Bioconductor package *domainCalling* (https://github.com/TapscottLab/domainCalling). The major functionalities of *domainCalling* include spike-in factor normalization and domain (broad peak) detection. To call the domains, the algorithm starts by counting reads overlapping with a sliding window throughout the genome for non-background (DOX+) and background (IgG or DOX-) samples. This window-based counting scheme is implemented by the csaw R/Bioconductor package [67]. After the counts are normalized by spike-in normalization, it filters out uninteresting windows if the average abundance of the non-background samples is (1) less than three-fold above background or (2) does not exceed the threshold, which is three reads per window. Finally, the retained windows are merged with neighbors within 2kbp distance. H3.X/Y domains were called for merged regions longer than 500 bps.

RNA-seq

MB135iDUX4 cells were treated with siRNA knockdown as described above followed by DUX4 pulsing, either once or twice, in triplicate, and harvested 24 hours after the start of a pulse. Untreated cells were also harvested from triplicate wells as negative controls. The NucleoSpin RNA kit (Macherey-Nagel) was used to extract RNA from whole cells, following the manufacturer's instructions. RNA-seq libraries were prepared using the Illumina TruSeq RNA

Sample Prep v2 Kit and a PerkinElmer Sciclone NGSx Workstation. All 15 libraries were pooled and sequenced on two flow lanes.

RNA-seq data analysis for doxycycline-pulsed samples

To preprocess the RNA-seq reads, we filtered out unqualified reads and aligned the reads to human genome hg38 using Tophat-2.1.0. We then profiled the gene expression using features collected from Ensembl v88 and the hit-counting function *summarizeOverlaps()* from Bioconductor's GenomicAlignments package. To identify robust DUX4 target genes, we used DESeq2 comparing pulse1 siControl samples to negative controls. The alternative hypothesis is set to $|\beta| > 4$, where β denotes log2 fold change. 170 genes with adjusted p-value < 0.05 were determined as robust DUX4 targets.

Western blotting

Protein was directly lysed from tissue culture plates using 2X Laemlli Buffer with 4% beta mercaptoethanol, sonicated, and boiled for 10 minutes. Samples were loaded on 4-12% polyacrylamide gel (Novex) and run with MES buffer, then transferred to a PVDF membrane. Membranes were blocked in 5% milk in PBST for 1 hour and incubated with primary antibody in 5% milk overnight at 4° (see below for details on antibodies). After washing, membranes were incubated with appropriate secondary antibodies for 1 hour, washed, and detected with chemiluminescence on film.

Reverse Transcription and qPCR

RNA was extracted as described for RNA-seq, treated with DNAse I (ThermoFisher), and heat inactivated. 500ng-1µg of RNA was reverse transcribed using SuperScript III First Strand

cDNA Synthesis (ThermoFisher) according to manufacturer's instructions, using oligo-dT priming. A no-enzyme sample was also run with a mix of all RNA samples as a control. qPCR was performed using 1x iTaq SYBR Green Master Mix (Bio-Rad) and primers at 1μM each and analyzed using the standard curve method and *RPL-27* as a normalizer. For Figure 1C, samples were also normalized to the average of all control replicates. Any samples from the control lines that had undetectable signal were set to 0 and not used for normalizing. Primers are listed below. *H3.X* and *H3.*Y primers are from [51].

Immunofluorescence

Briefly, cells were fixed in 1% formaldehyde for 15 minutes, permeabilized in PBST, and incubated overnight with primary antibodies. Plates were then washed with PBS, incubated 1 hour with fluorescent secondary antibody, counter-stained with DAPI, and imaged using an immersion lens.

Cloning H3.Z

A putative *H3.Z* sequence was identified from RNA-seq reads and used to design primers slightly outside this region. Amplicons from cDNA of *DUX4*-expressing cells were individually subcloned using the TOPO system, miniprepped using the NucleoSpin Plasmid kit (Macherey-Nagel) according to manufacturer's instructions, and Sanger sequenced to generate the final *H3.Z* sequence.

Software

The bioinformatics analysis was mostly performed on R-3.4.3/Bioconductor-3.5. The major infrastructure packages used include *csaw*, *edgeR*, *GenomicAlignments*, *ChIPseeker* and *ggplot2*.

siRNA sequences

siXY1: UCAAGAAGCCUCACCGCUAUU
siXY2: GCGGGAAAUCAGAAAGUACUU

qPCR primer sequences

RPL-27F: GCAAGAAGAAGATCGCCAAG
RPL-27R: TCCAAGGGGATATCCACAGA
X+YF: GGACCTGCGCTTCCAGAG
X+YR: CATGTCTCGGGGCATAATTG
XF: ACCAAAGCCGCCAGAAAA
XR: GAGCAGCTGCGTGGACTT
YF: CCACCAAAGCAGCCGGAAA
YR: GCTGCGTGGACTTCTGGTACT
ZF: AGGCGTGTCACAGTTATGCC
ZR: CCGTCTAGAGTGCAGCGTTT
DUX4 transgene F: TAGGGGAAGAGGTAGACGGC
DUX4 transgene R: CGGTTCCGGGATTCCGATAG
DUX4 endogenous F: CGGAGAACTGCCATTCTTTC
DUX4 endogenous R: CAGCCAGAATTTCACGGAAG
Zscan4F: TGGAAATCAAGTGGCAAAAA
Zscan4R: CTGCATGTGGACGTGGAC
Trim43F: ACCCATCACTGGACTGGTGT
Trim43R: CACATCCTCAAAGAGCCTGA
XRCC5F: GCTGCAGTTGCACTTTCCTC
XRCC5R: CCACGCCGACTTGAGGATTA
TPT1F: CACGATGAGATGTTCTCCGAC
TPT1R: CACCAATGAGCGAGTCATCA

H3.Z cloning primers

5'UTR F: AGGTTCAGCCCAGTGACAAG
3'UTR R: ATCAGTGGGTCTGAGGTCCC
ZF: AGAAGTCCACACAGCTGCTTC
ZR: CTGGGCATAACTGTGACACG

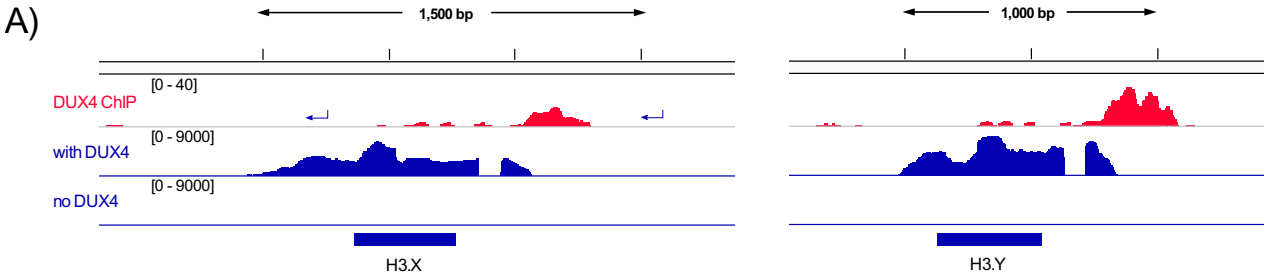
Antibodies for Western blot

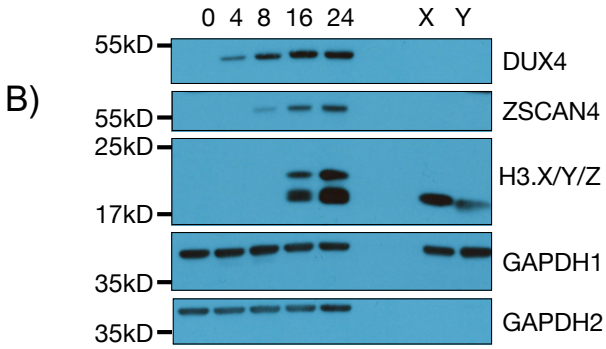
GAPDH: GTX28245 (GeneTex) lot#23184 **RRID:AB_370675**
DUX4: E14-3 1:500
Zscan4: PA5-32106 1:1000 (ThermoFisher) lot#TE25706711
H3.X/Y: clone 8H6-2111 supernatant 1:5
Goat anti-rat secondary: 1:5000 (Jackson 112-035-068)
Goat anti-rabbit secondary: 1:5000

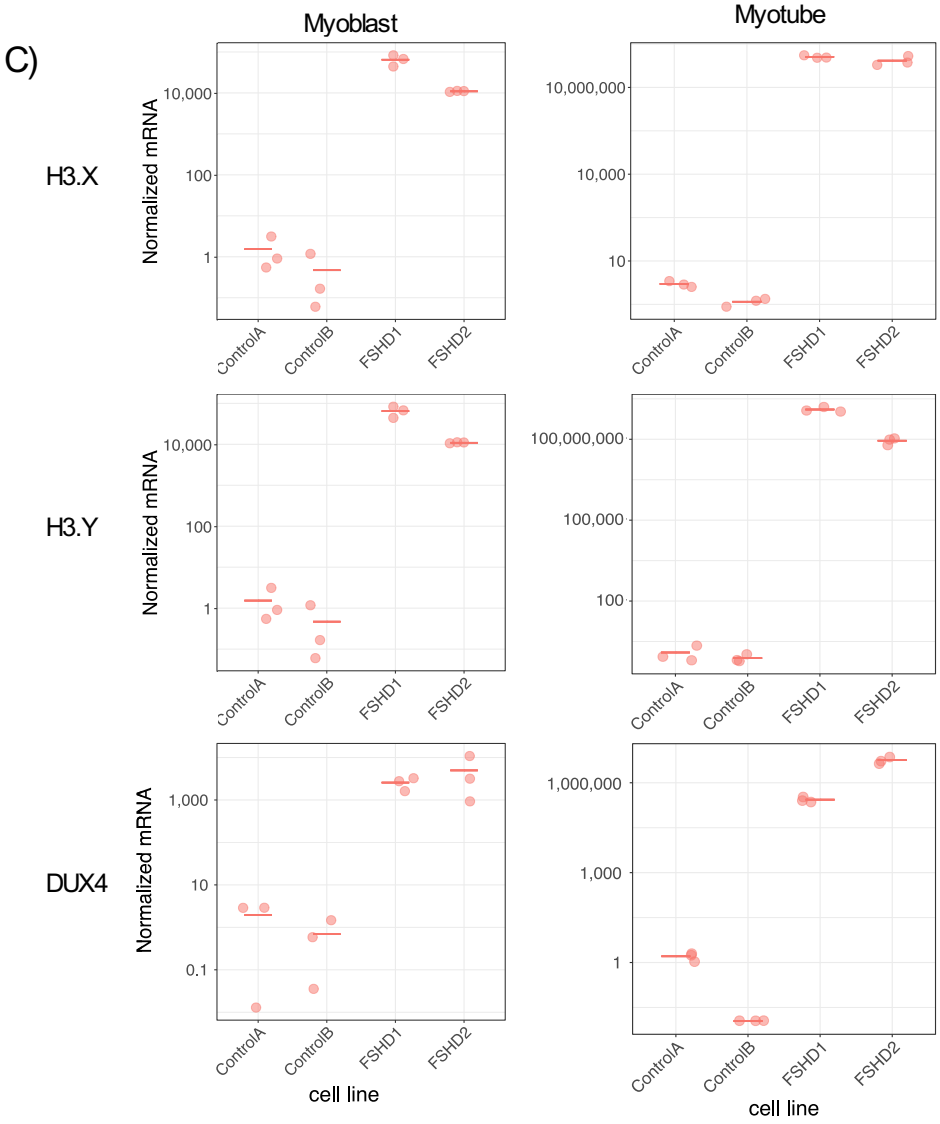
Goat anti-mouse secondary: 1:5000

Antibodies for IF

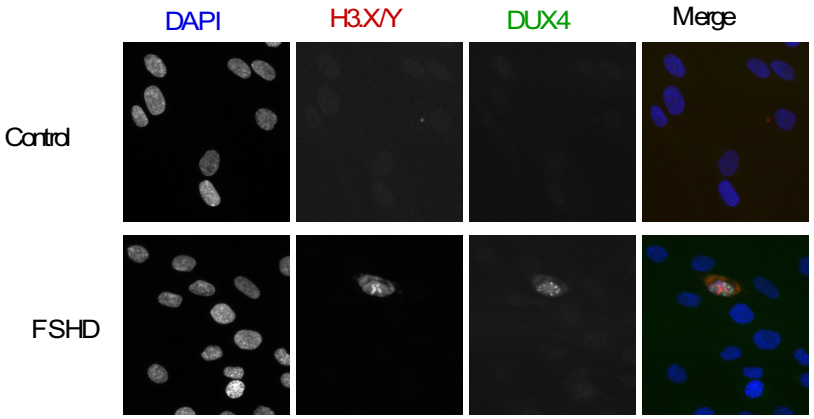
DUX4: E14-3 1:250
H3.X/Y: clone 8H6-2111 supernatant 1:20
TRITC anti-rat secondary: 1:250
FITC anti-rabbit secondary: 1:250


A)
1,500 bp
1,000 bp
[0 - 40]
DUX4 ChIP
[0 - 9000]
with DUX4
[0 - 9000]
no DUX4
H3.X
H3.Y

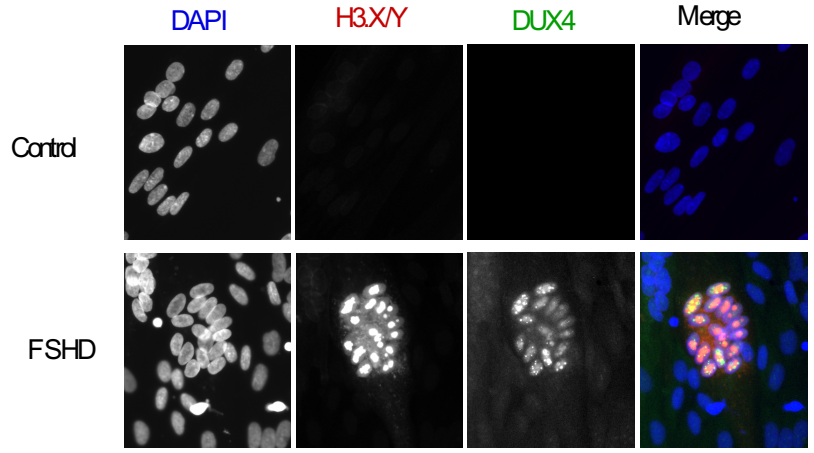

B)
0 4 8 16 24 X Y
55kD
DUX4
55kD
ZSCAN4
25kD
17kD
H3.X/Y/Z
35kD
GAPDH1
35kD
GAPDH2


C)
Myoblast
Myotube
H3.X
H3.Y
DUX4
Normalized mRNA
10,000
100
1
10,000,000
10,000
10
100,000,000
100,000
100
1,000
0.1
1,000,000
ControlA
ControlB
FSHD1
FSHD2
cell line

D)


DAPI
H3.X/Y
DUX4
Merge
Control
FSHD

E)


DAPI
H3.X/Y
DUX4
Merge
Control
FSHD

F)

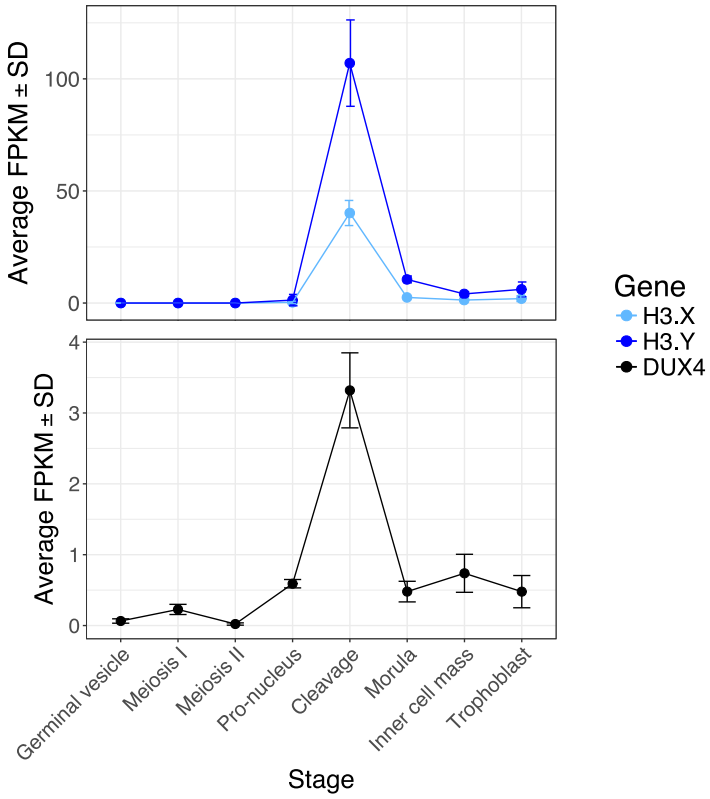
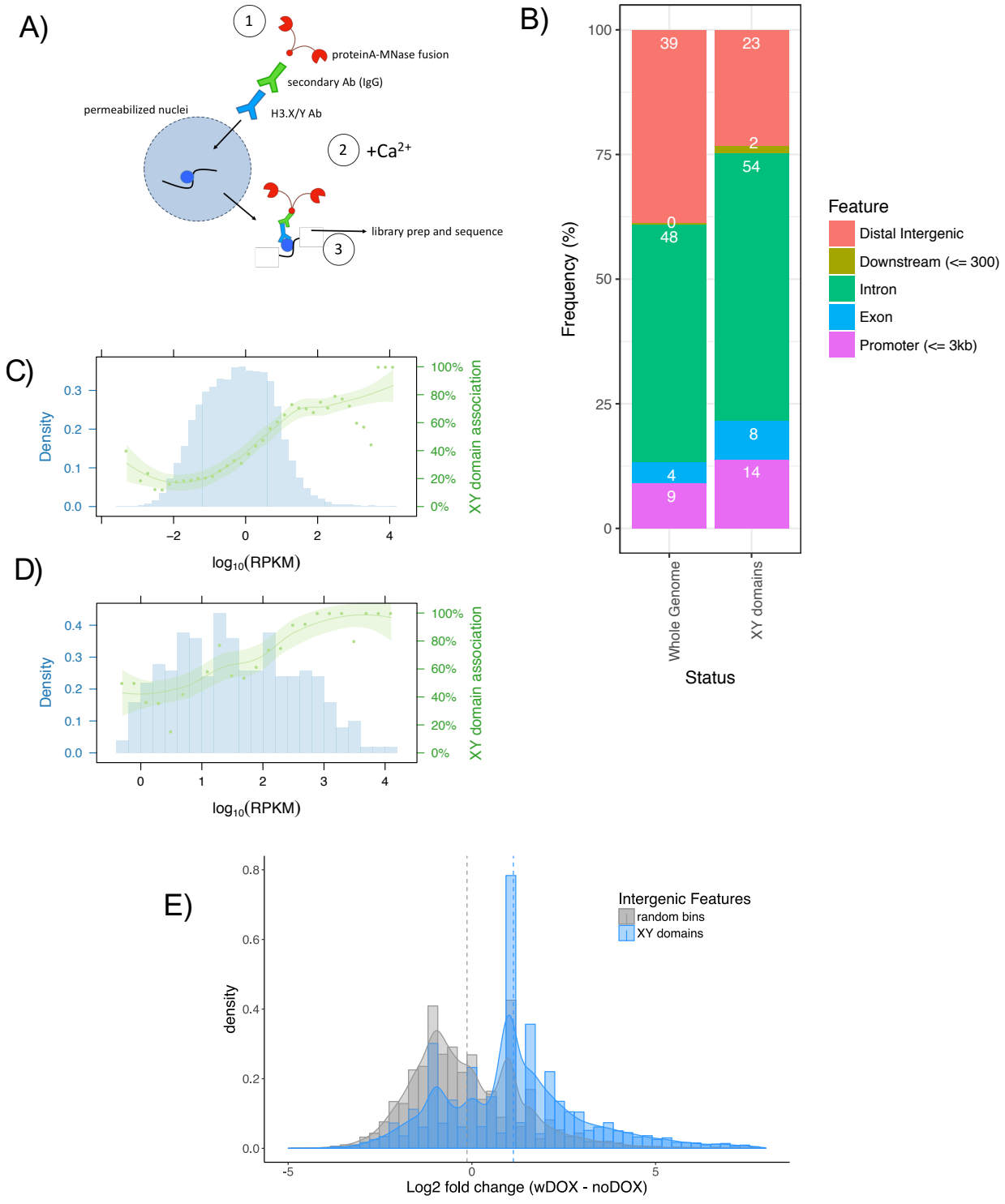
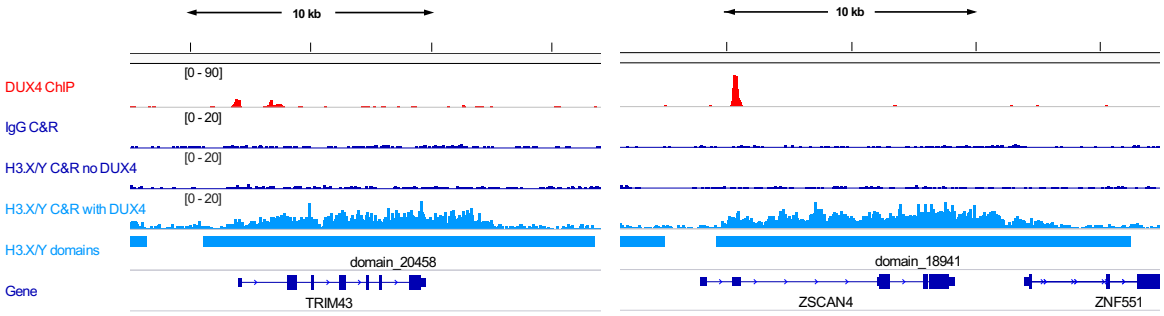

Average FPKM ± SD
100
50
0
4
3
2
1
0
Gene
H3.X
H3.Y
DUX4
Germinal vesicle
Meiosis I
Meiosis II
Pro-nucleus
Cleavage
Morula
Inner cell mass
Trophoblast
Stage

Figure 2.1: DUX4 induces the expression of histone variants *H3.X* and *H3.Y*

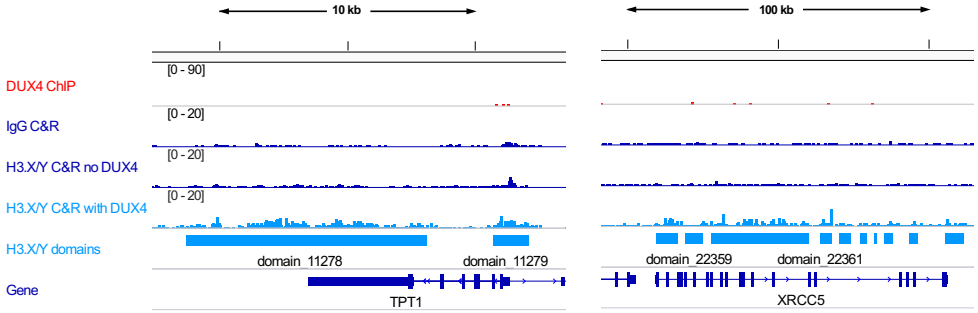
(A) RNA-seq (blue) and ChIP-seq (red) tracks from Jagannathan *et al*., 2016 and Geng *et al*., 2012, respectively, showing that *H3.X* and *H3.Y* were robustly induced with *DUX4* induction and that DUX4 was bound upstream of both *H3.X* and *H3.Y*. (B) Western blot analysis of MB135iDUX4 time course after *DUX4* induction. H3.X and H3.Y migrate at slightly different sizes as shown by *in vitro* translated protein (lanes X and Y), forming a closely-spaced doublet in the induced cells. The larger band corresponds to the predicted size for H3.Z. GAPDH1 and GAPDH2 represent loading controls for DUX4 or ZSCAN4 and H3.X/Y/Z, respectively. (C) RT-qPCR analysis of *DUX4*, *H3.X*, and *H3.Y* in control, FSHD1, and FSHD2 myoblasts and myotubes performed in biological triplicates. Values were normalized to the average of the control samples. (D,E) Immunofluorescence of H3.X/Y (red) and DUX4 (green) in FSHD1, FSHD2, and control myoblasts (FSHD2 not shown) (D) and myotubes (FSHD1 not shown). (E). DUX4 and H3.X/Y co-stain rare FSHD myoblast cells and many nuclei within myotubes, with no staining seen in control cells. DAPI channel in myoblasts and H3.X/Y and DUX4 channels in myotubes were brightened equally for all samples. (F) Average FPKM (Fragments Per Kilobase of transcript per Million reads) of reads from indicated stages of early human embryos [25] that map to *H3.X*, *H3.Y*, and *DUX4*.


A)
1
proteinA-MNase fusion
secondary Ab (IgG)
permeabilized nuclei
H3.X/Y Ab
2
+Ca2+
library prep and sequence
3
B)
Frequency (%)
100
75
50
25
0
39
0
48
4
9
23
2
54
8
14
Whole Genome
XY domains
Status
Feature
Distal Intergenic
Downstream (<= 300)
Intron
Exon
Promoter (<= 3kb)
C)
Density
0.3
0.2
0.1
0.0
−2
0
2
4
log10(RPKM)
100%
80%
60%
40%
20%
0%
XY domain association
D)
Density
0.4
0.3
0.2
0.1
0.0
0
1
2
3
4
log10(RPKM)
100%
80%
60%
40%
20%
0%
XY domain association
E)
density
0.8
0.6
0.4
0.2
0.0
-5
0
5
Log2 fold change (wDOX - noDOX)
Intergenic Features
random bins
XY domains

F)


10 kb
10 kb
DUX4 ChIP
IgG C&R
H3.X/Y C&R no DUX4
H3.X/Y C&R with DUX4
H3.X/Y domains
Gene
[0 - 90]
[0 - 20]
[0 - 20]
[0 - 20]
domain_20458
domain_18941
TRIM43
ZSCAN4
ZNF551

G)


10 kb
100 kb
DUX4 ChIP
IgG C&R
H3.X/Y C&R no DUX4
H3.X/Y C&R with DUX4
H3.X/Y domains
Gene
[0 - 90]
[0 - 20]
[0 - 20]
[0 - 20]
domain_11278
domain_11279
domain_22359
domain_22361
TPT1
XRCC5

H)

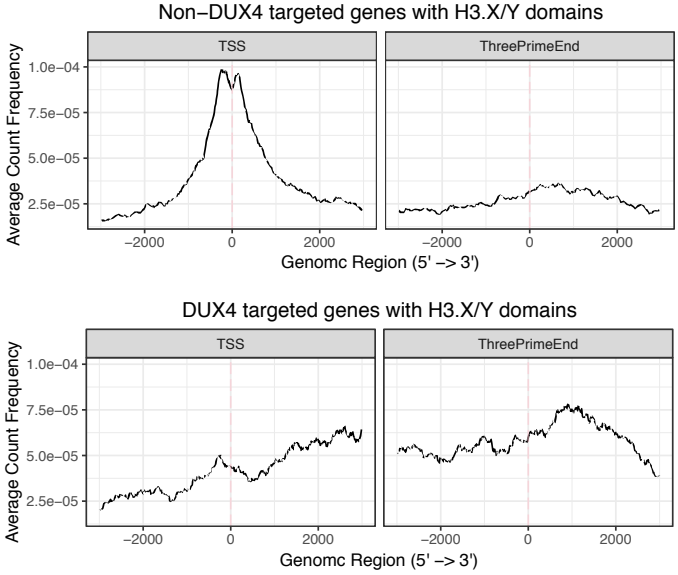
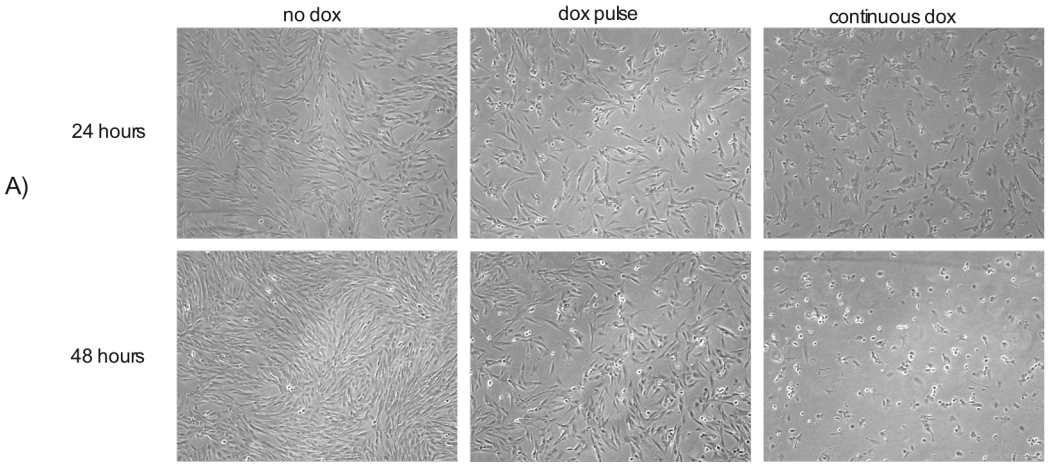
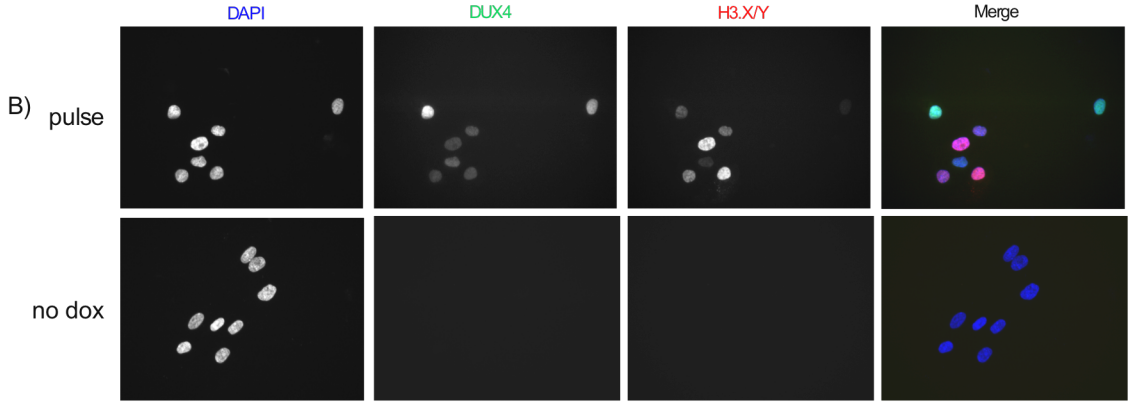
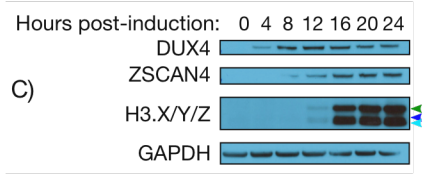

Non−DUX4 targeted genes with H3.X/Y domains
TSS
ThreePrimeEnd
Average Count Frequency
1.0e−04
7.5e−05
5.0e−05
2.5e−05
−2000
0
2000
Genomc Region (5' −> 3')
DUX4 targeted genes with H3.X/Y domains
TSS
ThreePrimeEnd
Average Count Frequency
1.0e−04
7.5e−05
5.0e−05
2.5e−05
−2000
0
2000
Genomc Region (5' −> 3')

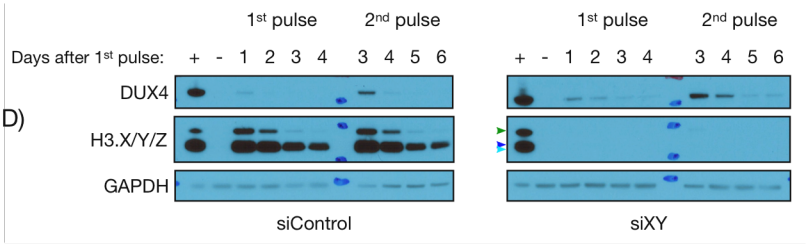
Figure 2.2: H3.X/Y are incorporated in expressed regions of the genome

(A) Schematic of CUT&RUN protocol. H3.X/Y nucleosome is represented by a blue circle. (B) H3.X/Y domains were predominantly located in genes (~75%). (C) Histogram of all gene expression (RPKM) from RNA-seq in induced MB135iDUX4 cells [34] (blue, left axis) and percentage of genes in each bin with one or more H3.X/Y domains (green, right axis). (D) Same as C) for robust DUX4 targets (n = 251). (E) Histogram of intergenic region expression change with *DUX4* induction. Comparing intergenic H3.X/Y domains with random intergenic bins of comparable size shows an association between H3.X/Y domains and DUX4-induced transcription. Vertical dashed lines mark mean log2 fold change with DUX4 for each group. (F) DUX4 ChIP-seq [28], IgG or H3.X/Y CUT&RUN (C&R), and called H3.X/Y domains at DUX4 target genes *TRIM43* and *ZSCAN4*. (G) Same as F but for genes constitutively expressed at similar levels to *TRIM43* (*TPT1*) and *ZSCAN4* (*XRCC5*). (H) Analysis of the distribution of H3.X/Y CUT&RUN reads over genes. Top panel shows highly-expressed, non-*DUX4* target genes with H3.X/Y domains (n = 679), bottom panel shows *DUX4* target genes only (n = 191).


A)
no dox
dox pulse
continuous dox
24 hours
48 hours


B)
DAPI
DUX4
H3.X/Y
Merge
pulse
no dox


C)
Hours post-induction: 0 4 8 12 16 20 24
DUX4
ZSCAN4
H3.X/Y/Z
GAPDH


D)
1st pulse
2nd pulse
Days after 1st pulse: + - 1 2 3 4 3 4 5 6
DUX4
H3.X/Y/Z
GAPDH
siControl
siXY

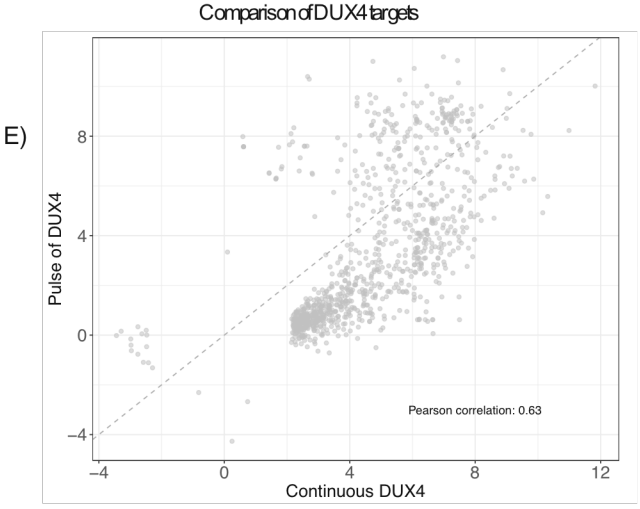

E)
Comparison of DUX4 targets
Pulse of DUX4
Continuous DUX4
Pearson correlation: 0.63
8
4
0
−4
−4
0
4
8
12

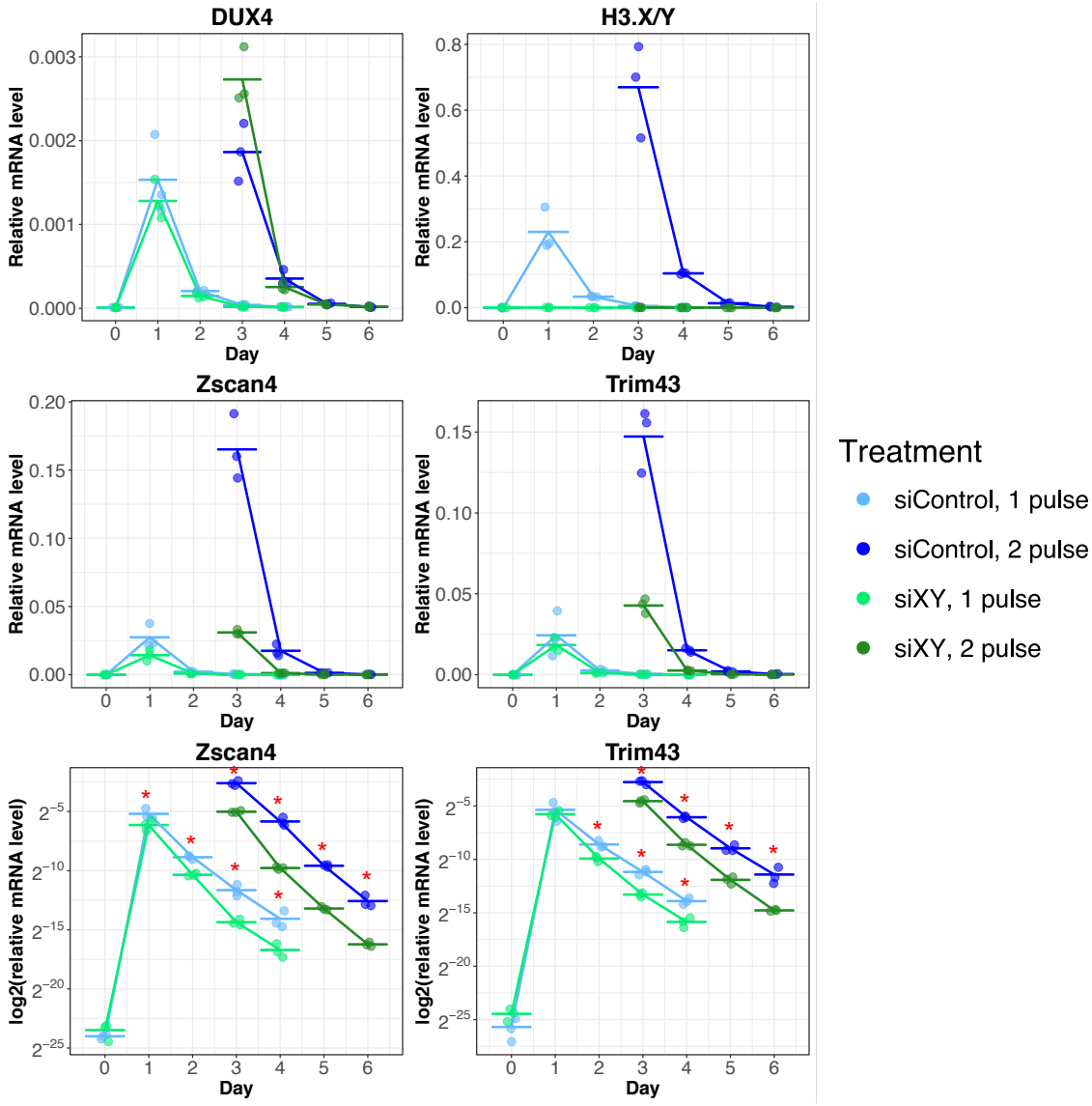
Figure 2.3: A pulse of *DUX4* activates target gene expression with little cell toxicity

(A) MB135iDUX4 cells 24 and 48 hours after no induction (no dox), a 4-hour pulse of doxycycline (dox pulse), or continuous doxycycline exposure (continuous dox), show toxicity with continuous but not pulsed *DUX4*. 48 hour images were brightened equally for all samples. (B) DUX4 and H3.X/Y immunofluorescence in MB135iDUX4 cells 8 hours after the start of a 4 hour pulse of doxycycline (pulse) shows induction in nearly all cells, with no staining in uninduced (no dox) cells. (C) Western blot analysis of MB135iDUX4 cells up to 24 hours after *DUX4* pulse. Cells were induced from hours 0-4. H3.X/Y/Z are identified with colored arrowheads—H3.Z is green, H3.X is dark blue, and H3.Y is light blue. (D) MB135iDUX4 cells with control or *H3.X/Y* knockdown with 1 or 2 pulses of *DUX4* on days 0 and 2, respectively, and harvested 1-4 days after each pulse (days 1-4 for 1st pulse, days 3-6 for 2nd pulse). + = continuous dox overnight, - = uninduced day 0 cells. H3.X/Y/Z are identified as in C. (E) Comparison of *DUX4*-induced genes from RNA-seq datasets in MB135iDUX4 cells after continuous (from Jagannathan *et al*., 2016) and pulsed *DUX4* expression. Axes show degree of gene induction (log2-fold change over no *DUX4*) with a cutoff of 2 (4-fold induction).

A)

B)

C)


DUX4
H3.X/Y
Zscan4
Trim43
Zscan4
Trim43
Relative mRNA level
log2(relative mRNA level)
Day
Treatment
siControl, 1 pulse
siControl, 2 pulse
siXY, 1 pulse
siXY, 2 pulse

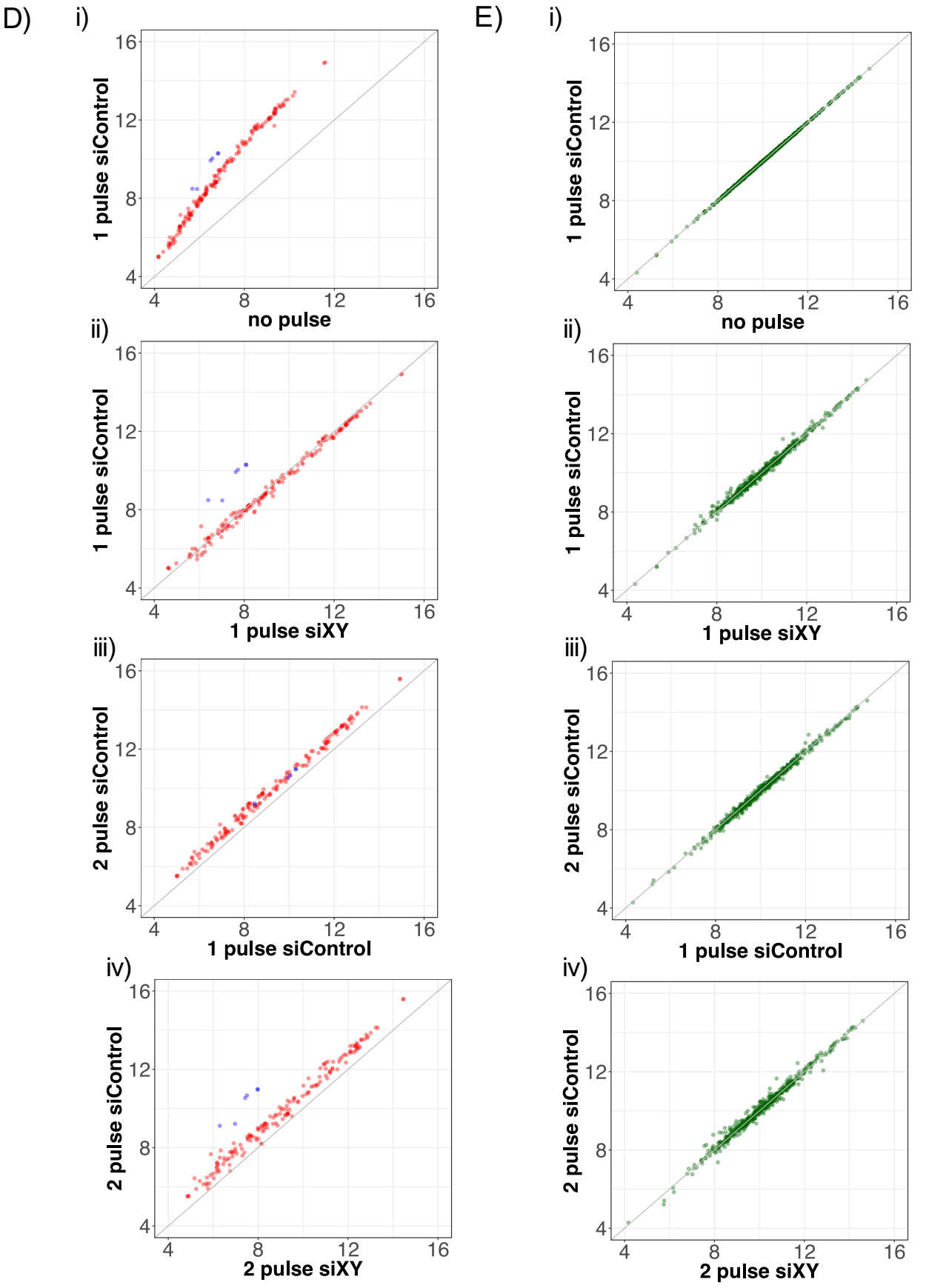

D)
i)
1 pulse siControl
no pulse
ii)
1 pulse siControl
1 pulse siXY
iii)
2 pulse siControl
1 pulse siControl
iv)
2 pulse siControl
2 pulse siXY
E)
i)
1 pulse siControl
no pulse
ii)
1 pulse siControl
1 pulse siXY
iii)
2 pulse siControl
1 pulse siControl
iv)
2 pulse siControl
2 pulse siXY
4
8
12
16

Figure 2.4: H3.X/Y incorporation increases the perdurance and re-expression of DUX4-target genes.

(A-C) RT-qPCR in MB135iDUX4 cells with 1 or 2 pulses of *DUX4* and treatment with siH3.X/Y (green) or siControl (blue). Cells were harvested before induction and 1-4 days after each pulse, with 3 biological replicates for each sample shown, relative to *RPL-27*. qPCR targets include controls for induction and knockdown (A) and *DUX4* targets with H3.X/Y incorporation (B,C). Data for C are the same as B but plotted on a log scale to more clearly demonstrate perdurance of *DUX4* target gene expression when H3.X/Y are incorporated. * = p-value <0.05, one-tailed Wilcoxon rank sum test comparing knockdowns within each pulse. (D) Expression of *DUX4* targets measured by RNA-seq in MB135iDUX4 cells, shown as average log2 normalized read counts of biological triplicates. Sequences in the *H3.X/Y* family targeted by siXY are shown in blue. Null model (no difference between conditions) shown in grey. (i) induction of *DUX4* targets (ii) comparison of *H3.X/Y* and control knockdown with a single pulse (iii) super-induction of *DUX4* targets with a second pulse in control knockdown samples (iv) knockdown of *H3.X/Y* eliminates super-induction. (E) Expression of genes with *H3.X/Y* domains that are unaffected by *DUX4*, plotted as in D.

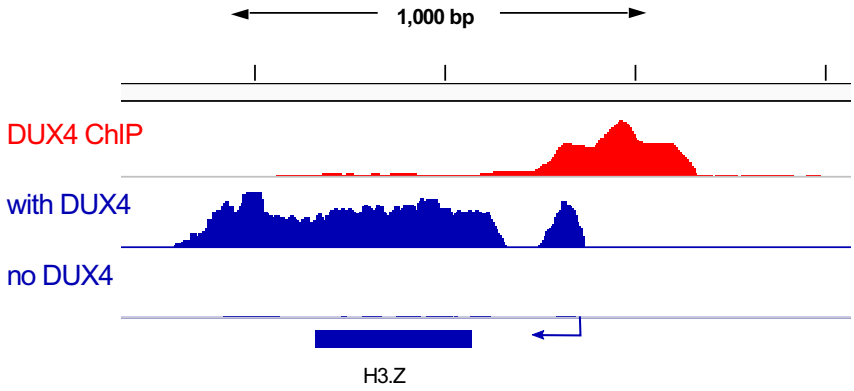
A)


H3.X
H3.Y
H3.Z

B)

H3.1
H3.3
H3.X
H3.Y
H3.Z

C)


1,000 bp
DUX4 ChIP
with DUX4
no DUX4
H3.Z

D)

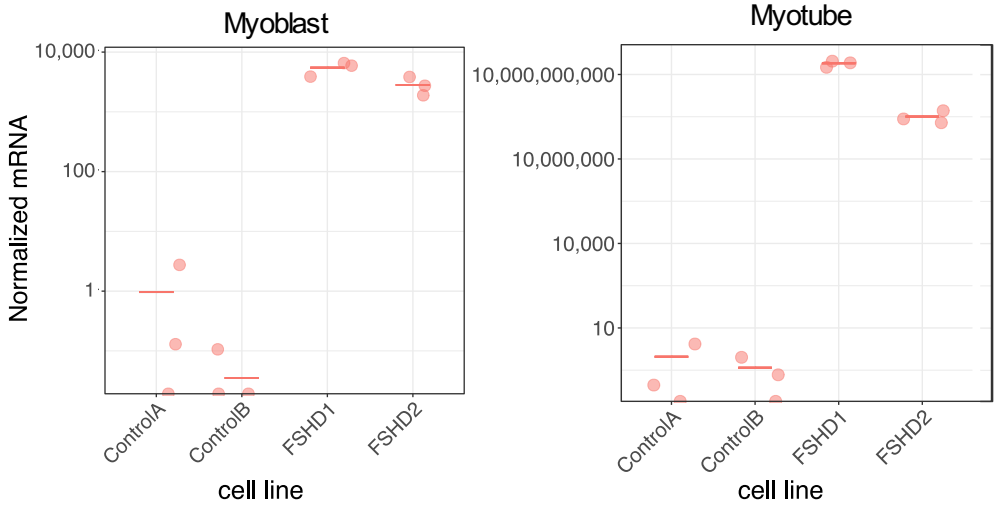

Myoblast
Myotube
Normalized mRNA
10,000
100
1
10,000,000,000
10,000,000
10,000
10
ControlA
ControlB
FSHD1
FSHD2
cell line

Figure S 2.1: The *H3.Z* variant was directly induced by DUX4

(A) DNA sequence alignment of *H3.X*, *H3.Y*, and *H3.Z*. *H3.Z* frameshift is highlighted in blue. (B) Protein sequence alignment of H3.1, H3.3, H3.X, H3.Y, and H3.Z with variant residues highlighted in grey. Epitope for H3.X/Y/Z antibody is noted with a blue line. (C) RNA-seq reads of MB135iDUX4 cells with and without DUX4 induction [34] that align to the *H3.Z* locus, as well as DUX4 ChIP-seq [28] showing a peak upstream of *H3.Z*. (D) RT-qPCR analysis of *H3.Z* as in Figure 2.1C.

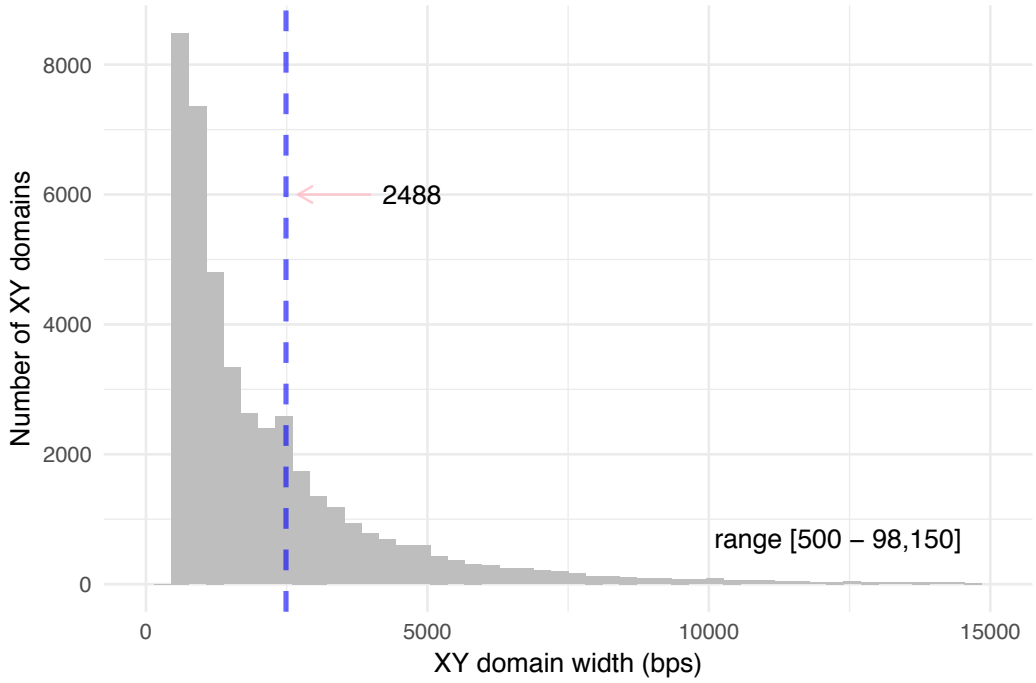


Figure S 2.2: Histogram of H3.X/Y CUT&RUN domain sizes

Mean is marked with a dashed blue line.

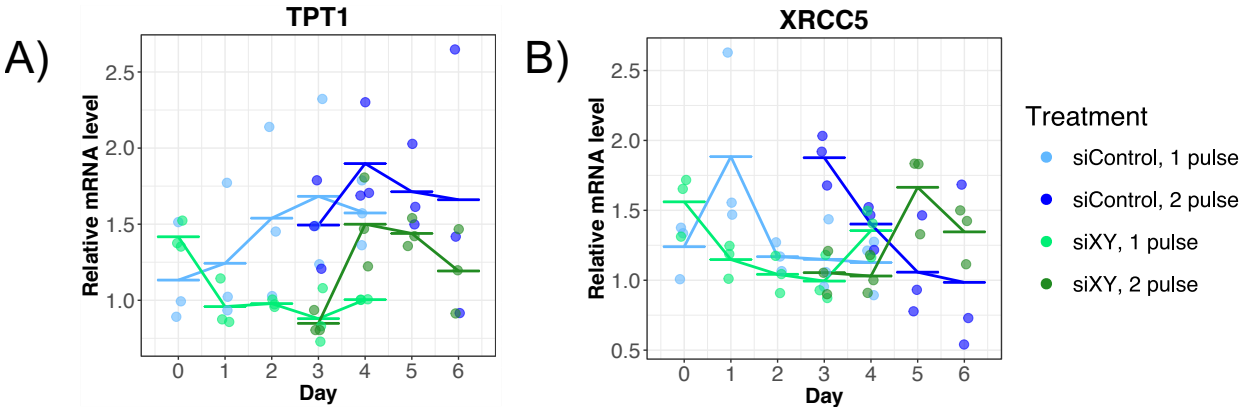


Figure S 2.3: siH3.X/Y does not affect expression of constitutively-expressed genes

(A,B) RT-qPCR in MB135iDUX4 cells with 1 or 2 pulses of *DUX4* and treatment with siH3.X/Y (green) or siControl (blue). Cells were harvested before induction and 1-4 days after each pulse, with 3 biological replicates for each sample shown, relative to *RPL-27*.

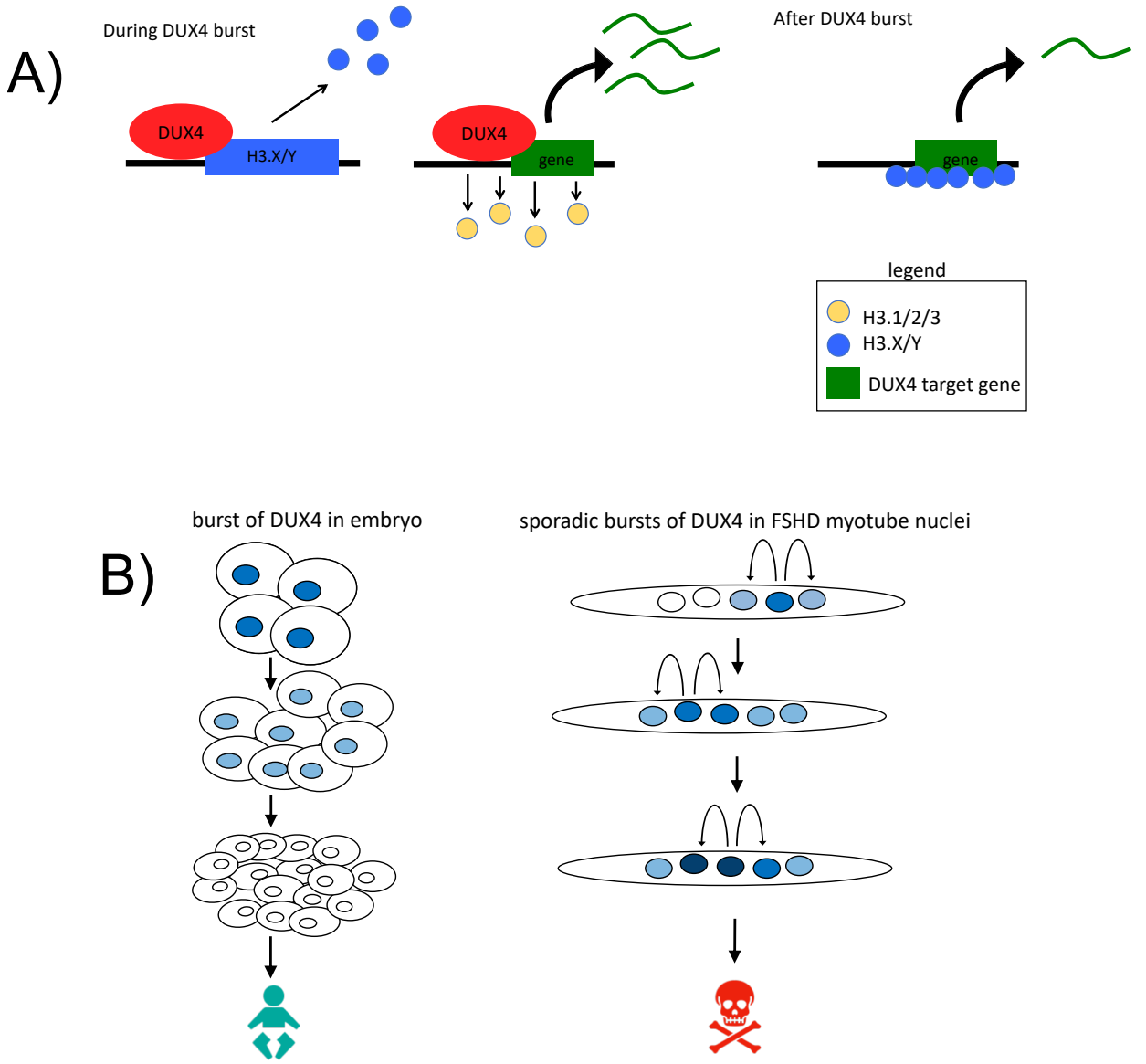


Figure S 2.4: Models of H3.X/Y function

(A) Model of H3.X/Y induction during a *DUX4* burst and incorporation into DUX4-regulated genes. (B) Incorporation of H3.X/Y (deeper blue denotes more H3.X/Y incorporation) at DUX4-induced genes might contribute to more perdurant DUX4-target expression in the early embryo and to an amplification of DUX4-target expression with successive bursts of DUX4 in different nuclei of a multinucleated muscle (left and right, respectively).

Chapter 3. DUX4 IS NOT SUFFICIENT TO INDUCE TELOMERE LENGTHENING

This chapter contains unpublished work.

3.1 INTRODUCTION

Among other places in the genome, the D4Z4 macrosatellite repeat containing the *DUX4* retrogene is located in a tandem array on the subtelomere of chromosome 4, and has been proposed to have a telomere position effect. Mouse *Dux* is also located in a subtelomeric array despite arising from a completely independent retrotransposition event, suggesting that this localization may serve an important, shared purpose. The only published study to look at the effect of telomere length on human *DUX4* expression showed an inverse correlation in cells derived from an FSHD patient [32], with shorter telomeres in isogenic cell lines leading to greater *DUX4* expression in culture.

DUX4 and mouse *Dux* activate species-specific totipotency networks but share some overlapping targets that are important in pre-implantation embryos, including *ZSCAN4* [26], suggesting a shared pathway. *ZSCAN4* has been shown to be an iPS reprogramming factor and is critical in mouse for progression from the 2-cell to the 4-cell stage [68]. ZSCAN4 protein binds to telomeres and facilitates telomere lengthening even in the absence of telomerase [69], signifying its involvement in the alternative lengthening of telomeres (ALT) pathway. Mouse *Tcstv1* and *Tcstv3*, which do not have human homologues, work with *Zscan4* to elongate telomeres in mice and increase telomere sister chromatid exchange, a proposed mechanism of ALT [70]. They are also targets of mouse Dux [26, 71], providing a further link between ALT and Dux.

The early embryonic expression of *Zscan4* in mice occurs during a dramatic telomere lengthening that is not dependent on telomerase and has evidence of frequent sister chromatid telomere exchange [72]. Studies examining telomeres in human sperm found highly heterogeneous telomere length in older men, which is inconsistent with the telomerase pathway but is characteristic of ALT [73, 74], suggesting that ALT may be playing a role in telomere lengthening during spermatogenesis as well, where DUX4 is also thought to be expressed. Approximately 15%

of cancers use ALT to maintain telomeres [75, 76], making it an attractive pathway to target for therapy. Additionally, ALT has a very high correlation with loss of ATRX or DAXX, components of the H3.3 chaperone complex, as well as with H3.3 driver mutations in brain tumors [77], suggesting that H3.3 or H3.3-like histones play an important role in ALT.

An additional link to *DUX4* biology is through UPF1, a protein critical in the nonsense-mediated decay (NMD) mRNA quality control pathway, which is degraded upon *DUX4* expression [29]. The *DUX4* transcript is itself a target of NMD, so this creates a positive-feedback loop through stabilization of *DUX4* mRNA. UPF1 also plays a role in telomere maintenance, binding directly to telomeres, telomerase, and the shelterin protein TPP1, and acts to help stabilize telomeres [78]. ALT is characterized by telomere instability, with unequal telomere exchange and elongation, suggesting that UPF1 depletion could be yet another mechanism whereby DUX4 contributes to activation of ALT.

The strong correlation between *DUX4* and *H3.X/Y/Z* expression and telomere lengthening via the telomerase-negative ALT pathway makes this relationship of great interest. H3.X/Y were first reported in the U2OS osteosarcoma cell line [51] which is often used as a model for studying ALT. *DUX4* is expressed in rare iPS cells [11], and there is evidence that iPS cells use both ALT and telomerase [79, 80]. The timing and location of endogenous DUX4 and H3.X/Y/Z expression is consistent with tissues where there is evidence for ALT, as described above [72, 74]. Additionally, H3.X/Y were not detected in HeLa cells, which do not use ALT [51], while IF for H3.X/Y/Z in U2OS cells showed speckles of perinuclear staining [51], where silenced chromatin such as telomeres tends to reside. Taken together, this expression pattern suggests a potential role for H3.X/Y/Z in ALT, both during development and cancer pathogenesis, as *H3.X/Y* have also been detected in some tumors [51]. If these histones prove critical for this process, they are attractive

targets for cancer therapy as ALT positivity is common in particularly difficult-to-treat tumor types such as glioblastoma multiforme and osteosarcoma, and the specificity of these histone variants to tumors but not normal tissue [51] would allow for a targeted therapy with less toxic side effects.

This chapter focuses on determining whether the ALT pathway is activated by DUX4 and H3.X/Y to contribute to disease pathology in FSHD and cancer. If DUX4 can in fact activate the ALT pathway, this could explain the short burst of *DUX4* expression seen in early embryos through creation of a negative feedback loop: shortened telomeres in the zygote lead to de-repression of *DUX4* and thus *ZSCAN4* and *H3.X/*Y expression and degradation of UPF1 (Figure 3.1A). H3.X/Y could then be incorporated at telomeres by the ATRX/DAXX complex, maintaining a more open chromatin conformation and facilitating ALT, and ZSCAN4 could bind (Figure 3.1B). The subsequent rapid telomere elongation would lead to formation of heterochromatin at the D4Z4 array, thereby repressing *DUX4* and ending the burst of expression (Figure 3.1C). While positive-feedback loops that enhance *DUX4* expression have been described [29], the mechanism for rapid repression of *DUX4* in the embryo is unknown, as is the initiating event that leads to the de-repression of *DUX4* in the first place. This model would explain both of these events and provide a mechanism for the *DUX4* bursting pattern as well as initiation of telomere lengthening in the early embryo.

Of note, this work was performed before the chaperones for H3.X/Y were determined by Zink *et al.* [53], when both ATRX/DAXX and HIRA were thought to recognize H3.3 exclusively by amino acids 87-90. These are shared by H3.X/Y and H3.3, suggesting that both chaperone complexes would recognize these variants and that ATRX/DAXX could incorporate H3.X/Y at telomeres. It is now clear that only HIRA incorporates H3.X/Y (and this was used to map the

ATRX/DAXX and HIRA recognition sites more thoroughly), making it highly unlikely that H3.X/Y would be incorporated at telomeres and affect their biology [53].

3.2 RESULTS

3.2.1 *Telomeres are unaffected by FSHD status or differentiation*

To test whether cells expressing *DUX4* have different telomere lengths or ALT activation than cells without *DUX4*, I compared primary FSHD and wild-type muscle cell lines, as cells with an active telomerase pathway (eg immortalized with telomerase) can suppress ALT [81, 82]. One assay for ALT activation measures the amount of "C-circle", circular DNAs composed of the C-rich telomere strand, which are quite specific to cells using the ALT pathway [83, 84]. These C-circles are partly single-stranded and partly double-stranded, with the portion of G-rich strand making them self-priming. They can thus be amplified using rolling circle amplification with phi polymerase as described [85]. Rolling circle amplification reactions are performed with or without the addition of polymerase, followed by qPCR for telomeres and a control single-copy gene to normalize across cell lines. In the polymerase-minus sample, the qPCR measures mean telomere content (see schematic of C-circle assay, Figure S3.1A). In the polymerase-plus sample, qPCR measures telomere content plus C-circles. Subtracting the telomere content from this total gives a measure of C-circle content, and thus ALT activity, in the cells (Figure S3.1B).

Control, FSHD1, and FSHD2 cells were assayed as described above, both as myoblasts and differentiated myotubes. *DUX4* levels increase upon differentiation so telomere content and C-circle levels would be expected to be higher in FSHD than control lines and in myotubes as compared to myoblasts if *DUX4* is sufficient to induce ALT. The myoblasts and myotubes tested had much lower telomere content than U2OS cells (~15-20%) and showed no substantial

difference between cell lines or with differentiation (Figure 3.2A). C-circles were not detected in any of the muscle lines, though there was strong signal from the U2OS positive control (Figure 3.2B). These data suggest that FSHD and differentiation status do not affect telomere content or show ALT activation, which is inconsistent with what would be predicted by our model. However, it is possible that levels of *DUX4* expression, averaged across the population, are too low to detect much difference in this context.

3.2.2 *ALT may increase in U2OS cells with starvation/overgrowth treatment through a* DUX4-*independent mechanism*

Since *DUX4* expression in FSHD cells seems to be insufficient to induce ALT, next we wanted to determine whether it could enhance the pathway. To do this, we used the U2OS osteosarcoma cell line. A GFP DUX4-reporter construct is visible in rare U2OS cells which also express H3.X/Y, presumably driven by *DUX4* expression as well (Figure 3.3A). A previous report showed that starvation/overgrowth (SO) treatment increased *H3.X/Y* levels in these cells [51], suggesting that it should also increase *DUX4* levels. To test whether this treatment could increase ALT, I assayed three independent U2OS cell lines that had been cultured in our lab (ST), the Tsukiyama lab (TT), or ordered from ATCC. Assaying for telomere content using the polymerase-minus C-circle assay showed differences in telomere content between the three cell lines, but no consistent increase with SO treatment (Figure 3.3B). The C-circle assay with polymerase, however, showed an increase in ALT activation for all three cell lines with SO treatment, consistent with my hypothesis (Figure 3.3C). To determine whether this was in fact due to an increase in *DUX4* expression, I performed RT-qPCR on RNA from these cells, using FSHD2 MB200 and Susa cells as positive controls that have different levels of endogenous *DUX4* expression. While *DUX4* was detectable in the U2OS lines, levels were quite low and did not

consistently increase with SO treatment; in the TT line, in fact, they dropped considerably (Figure 3.3D). Since endogenous *DUX4* levels can be low and difficult to measure accurately, I also looked at several *DUX4* targets by RT-qPCR including *H3.X/Y* to confirm this finding. Again, I saw no increase in expression with SO treatment (Figure 3.3E). Together, these data suggest that SO treatment can increase C-circles in cells with an active ALT pathway, but that this does not substantially increase telomere length (at least in the timespan used for this experiment) and is independent of *DUX4*, *H3.X/Y*, or *ZSCAN4* levels.

3.2.3 *Pulsing DUX4 does not lengthen telomeres or induce/increase ALT in U2OS or wild-type myoblasts*

Studies using culture conditions intended to increase *DUX4* expression were insufficient to alter telomere content or affect the ALT pathway. To determine whether this was due to insufficient induction of *DUX4*, we wanted to increase levels with exogenous expression. To do this, I transduced primary MB135 wild-type myoblasts and U2OS cells with the inducible *DUX4* lentiviral system described in Chapter 2 to make doxycycline-inducible cell lines, as well as using the immortalized MB135iDUX4 clone described in Chapter 2. These cells were treated with or without 24 hours of doxycycline before harvesting, and their DNA was analyzed using the telomere restriction fragment length Southern blotting method. U2OS cells had longer and more heterogenous telomeres than the primary or immortalized MB135 myoblasts (Figure 3.4A, lanes 1-3). Primary myoblasts (Figure 3.4A lanes 4-6) seemed to have longer telomeres than the immortalized cells (Figure 3.4A lanes 7-9), but there were no obvious differences between untransduced, uninduced DUX4i, and *DUX4*-induced cells for any of the cell lines.

To rule out the possibility that 24 hours was not enough time to induce sufficient ALT activity to see an increase in telomere length, we next pulsed cells with *DUX4.* This would induce *DUX4*

and its targets without toxicity to allow us to assay the cells after several days. Primary MB135 myoblasts and U2OS cells were either left untreated or treated with 3-5 8-hour pulses, and harvested after one week. Telomere content was measured using qPCR and showed no significant increase in either cell line after pulsing with *DUX4* (Figure 3.4B). As seen previously, the untreated myoblasts had no detectable C-circles while the U2OS cells had easily detectable C-circles, but *DUX4* treatments did not affect these levels (Figure 3.4C), suggesting that previous negative results were not explained by the short timescale of the experiments.

Occasionally, pulsing cells several times with *DUX4* in other contexts had led to silencing of the transgene, with remaining sensitive cells dying and leaving a doxycycline-unresponsive population. To ensure that this was not confounding results, DUXiU2OS cells were treated with 5 daily pulses of *DUX4* and then left untreated for 48 hours as in the above experiment. They were then treated with doxycycline again for 48 hours and harvested. RT-qPCR showed that these cells were still able to activate the *DUX4* transgene (DS2 primers) and targets (Figure 3.4D). These data further suggest that *DUX4* is not sufficient to activate ALT, and do not support our proposed model.

3.3 DISCUSSION

In this chapter I have proposed a model whereby DUX4 activates the ALT telomere elongation pathway through its targets ZSCAN4, H3.X/Y, and the degradation of UPF1, providing a model for the short burst of *DUX4* expression seen in early development. I tested this model by assaying telomere length, content, and the presence of C-circles, a marker of ALT activity, in several different contexts with and without *DUX4*: FSHD and control myoblasts and myotubes, SO-treated U2OS cells, and myoblast and U2OS cells with exogenous *DUX4* expression. These

samples represented a wide range of cell states, including those with and without active ALT pathway, with different lengths of telomeres, primary and cancer cells, different levels of *DUX4* expression, and different amounts of time to allow for pathway activation.

Only one of these experiments showed an effect on the ALT pathway: SO-treating U2OS cells increased C-circle content in three independent U2OS lines. However, this did not lengthen telomeres or affect *DUX4* or target levels. This suggests that the ALT pathway may have been enhanced through a *DUX4*-independent mechanism, but the lack of telomere elongation coupled to the increase in C-circles suggests that it is likely to be an unrelated increase rather than a marker of ALT activation in this context. An orthogonal approach using a different assay such as staining for telomere-associated PML bodies, Southern blot, or quantification of sister telomere chromatid exchange would clarify whether ALT was actually affected in this context. Taken all together, however, these data do not support the proposed model.

While further work would be needed to truly disprove the model proposed here, these data strongly suggest that *DUX4* is not sufficient to induce or enhance ALT. Knowing as we do now that H3.X/Y exclusively use the HIRA chaperone complex and not ATRX/DAXX [53], there is a clear explanation for why H3.X/Y are not involved in the ALT pathway as they would not be incorporated into telomeres. However, many strong connections between the ALT pathway, DUX4, ZSCAN4, and UPF1 remain, suggesting that further inquiry into a potential causal relationship may still be warranted. One simple next step would be to determine whether *DUX4* and *ZSCAN4* are expressed in other ALT-positive cell lines. Work from Dr. Chew in Robert Bradley's Lab has identified many *DUX4*-positive tumors, which could be a good place to start this analysis. Another simple experiment would be to knock down *DUX4* in U2OS and other ALT-positive cell lines using shRNA to determine whether this transcription factor is necessary for ALT

activation. Since this would presumably not be toxic and there is little risk of construct silencing the way we see with *DUX4* overexpression, these cells could be cultured for even longer than the experiments described in this chapter, giving sufficient time for telomeres to shorten enough to be detectable.

3.4 METHODS

Southern blot

Primary MB135 myoblasts and U2OS cells were transduced with 150µL high-titer pCW57.1 DUX4-CA lentivirus using polybrene and selected with puromycin. Both cell lines, as well as the immortalized MB135iDUX4 clone, were treated with or without doxycycline for 24 hours, and untransduced parent cells were also treated with doxycycline to control for effects of the drug. Pellets were snap frozen in liquid N_2. DNA was prepped according to [86]. Briefly, cells were thawed, lysed, and treated with RNAse and ProteinaseK, then DNA was phenol/chloroform extracted, salt precipitated, and resuspended in water. 15µg of DNA were digested using RsaI and HinfI for 16.5 hours, treated with RNAseA a second time, and re-extracted using phenol/chloroform/isoamyl alcohol. 3µg of DNA were run per lane in a 0.5% agarose, 0.5X TBE gel, then and denatured and neutralized. DNA was transferred to a Nytron SPC 0.45µm nylon membrane with 20X SSC overnight, crosslinked, and pre-hybridized in FBI at 65°. The membrane was then hybridized with 70µL probe (see probe description below) for 18 hours, washed 4 times, and imaged on film.

Southern blot probe

1μL of 10μM telomere probe sequence was labeled using 10U T4 PNK and 40 μC γ32ATP at 37° for 40 minutes, then cleaned up using the GE G-50 ProbeQuant Micro Column according to the manufacturer's directions. This gave a signal of ~$15*10^6$ DPM/μL.

Telomere probe sequence: CCC TAA CCC TAA CCC TAA CCC TAA

C-circle assay

Genomic DNA was extracted with the GeneJET kit (Thermo Fisher) from freshly-harvested cells according to the manufacturer's directions, with an extended proteinaseK treatment of 1 hour. Rolling circle amplification of C-circles was performed as in [83], with and without addition of phi polymerase, using approximately 15ng of DNA per reaction. Samples were then diluted 2.75x in TE.

qPCR for telomeres

5 μL of diluted product from C-circle reactions with and without phi polymerase were run in 20 μL reactions in 2-3 replicate qPCR wells as in [87] with 1x iTaq SYBR Green Master Mix (Bio-Rad) for 40 cycles. Primers both for telomeres and VAV2, a single-copy gene used for normalization, were run for all samples. The standard curve method was used for analysis, with serially diluted U2OS C-circle reaction products (with and without phi polymerase) run as the standards.

Pulsing DUXi cells

MB135 cells were treated with 3 8-hour pulses of doxycycline, every other day, or daily pulses for 5 days, and harvested 2 days after the last pulse. U2OS cells were pulsed for either 8 hours for

4 consecutive days and harvested after the last pulse or for 5 consecutive days and harvested 2 days after the last pulse.

Reverse Transcription and qPCR

The NucleoSpin RNA kit (Macherey-Nagel) was used to extract RNA from whole cells, following the manufacturer's instructions. 500ng-1µg of RNA was treated with DNAseI (ThermoFisher) and heat inactivated, then reverse transcribed using SuperScript III First Strand cDNA Synthesis (ThermoFisher) according to manufacturer's instructions, using oligo-dT priming. A no-enzyme sample was also run with a mix of all RNA samples as a control. qPCR was performed in triplicate wells using 1x iTaq SYBR Green Master Mix (Bio-Rad) and primers at 1µM each and analyzed using the ΔCT method and *RPL-27* as a normalizer. Primers are listed below. *H3.X* and *H3.Y* primers are from [51].

Immunofluorescence

Briefly, cells were fixed in 1% formaldehyde for 15 minutes, permeabilized in PBST, and incubated overnight with primary antibodies. Plates were then washed with PBS, incubated 1 hour with fluorescent secondary antibody, counter-stained with DAPI, and imaged using an immersion lens.

Primers for qPCR telomere assay (from [87])

TelomereF: GGTTTTTGAGGGTGAGGGTGAG GGTGAGGGTGAGGGT
TelomereR: TCCCGACTAT CCCTATCCCTATCCCTATCCCTATCCCTA
VAV2F: TGGGCATGACTGAAGATGAC
VAV2R: ATCTGCCCTCACCTTCTCAA

Primers for RT-qPCR

RPL-27F: GCAAGAAGAAGATCGCCAAG
RPL-27R: TCCAAGGGGATATCCACAGA
H3.X+YF: GGACCTGCGCTTCCAGAG
H3.X+YR: CATGTCTCGGGGCATAATTG
H3.ZF: AGGTTCAGCCCAGTGACAAG
H3.ZR: GGTGGCTTTGTGGTCAGTCT
Zscan4F: TGGAAATCAAGTGGCAAAAA
Zscan4R: CTGCATGTGGACGTGGAC
MBD3L2F: CACCTCTTTTCCAAGCCTGCC
MBD3L2R: ACCTGGTTGTCAGGATGAGAC
DUX4 endogenous F: CGGAGAACTGCCATTCTTTC
DUX4 endogenous R: CAGCCAGAATTTCACGGAAG
DS2F: TGACTGGATATGTTGTGTTTTACAG
DS2R: CAACCCCGGATCCTTAGTG
KHDC1LF: CACCAATGGCAAAGCAGTGG
KHDC1LR: TCAGTCTCCGGTGTACGGTG

Antibodies for IF

H3.X/Y: clone 8H6-2111 supernatant 1:5
TRITC anti-rat secondary: 1:250

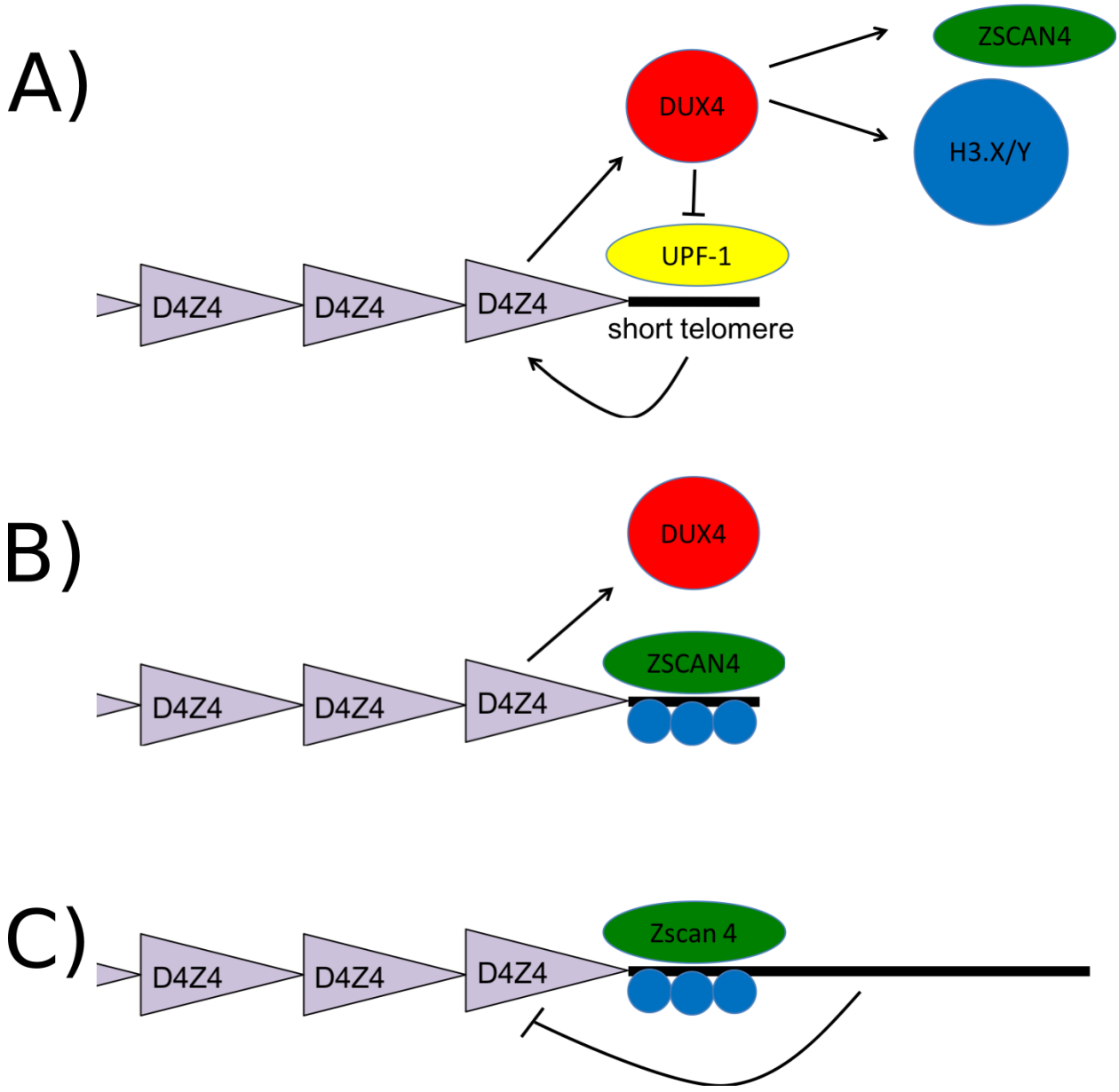


Figure 3.1: Proposed model of DUX4 activation of ALT through induction of ZSCAN4 and H3.X/Y and degradation of UPF1

(A) Short telomeres lead to de-repression of DUX4 which induces ZSCAN4 and H3.X/Y and the degradation of telomere-bound UPF1. (B) H3.X/Y are incorporated into telomeres by ATRX/DAXX and ZSCAN4 binds and activates the ALT pathway. (C) Telomeres are lengthened through ALT and feed back to shut down *DUX4* expression

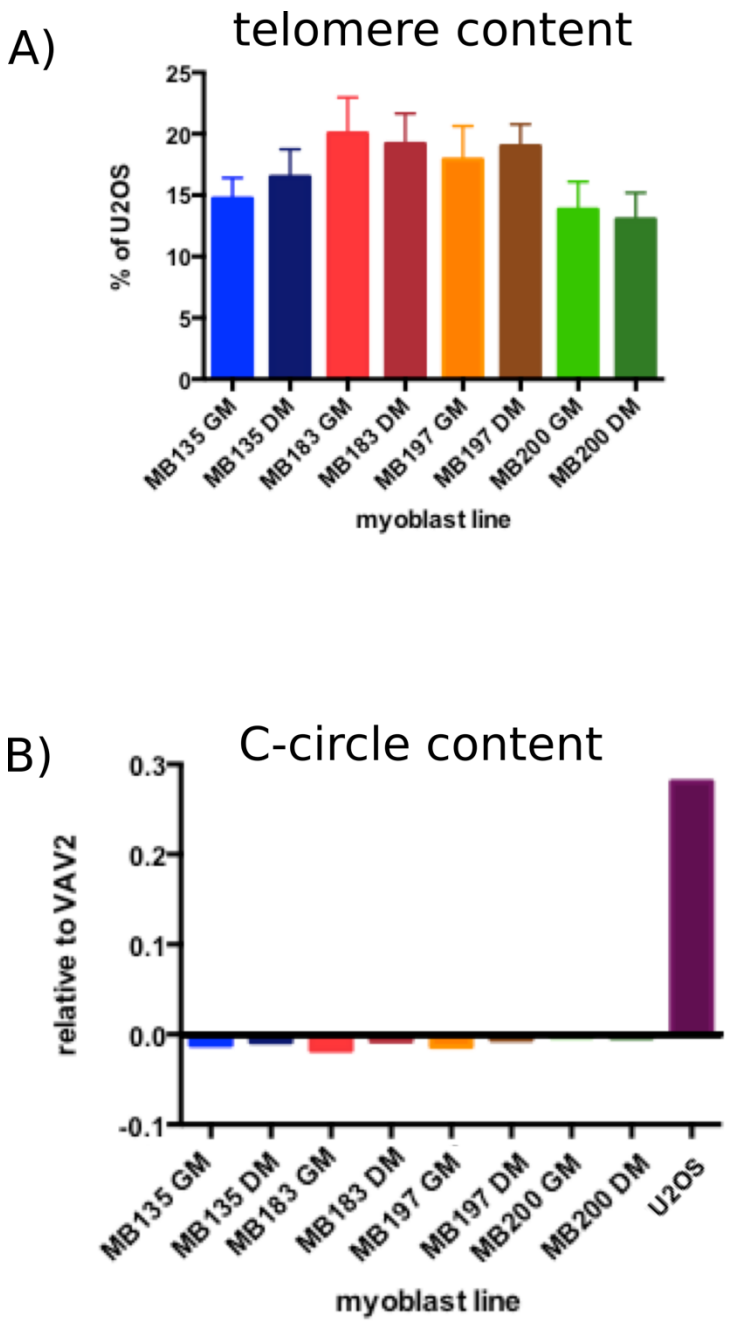


Figure 3.2: Telomeres are unaffected by FSHD status or differentiation

(A) Mean telomere content as measured by qPCR (polymerase-minus C-circle assay) normalized to U2OS cells in primary wild-type (MB135), FSHD1 (MB183, MB197), and FSHD2 (MB200) myoblasts (GM) and myotubes (DM) shows no difference between control and FSHD cells or increase upon differentiation. Error bars show standard deviation of triplicate qPCR reactions. (B) Mean C-circle content measured by qPCR (polymerase-plus C-circle assay) and normalized to single-copy gene VAV2 shows no C-circles in muscle cells regardless of FSHD status or differentiation, with U2OS cells as a positive control.

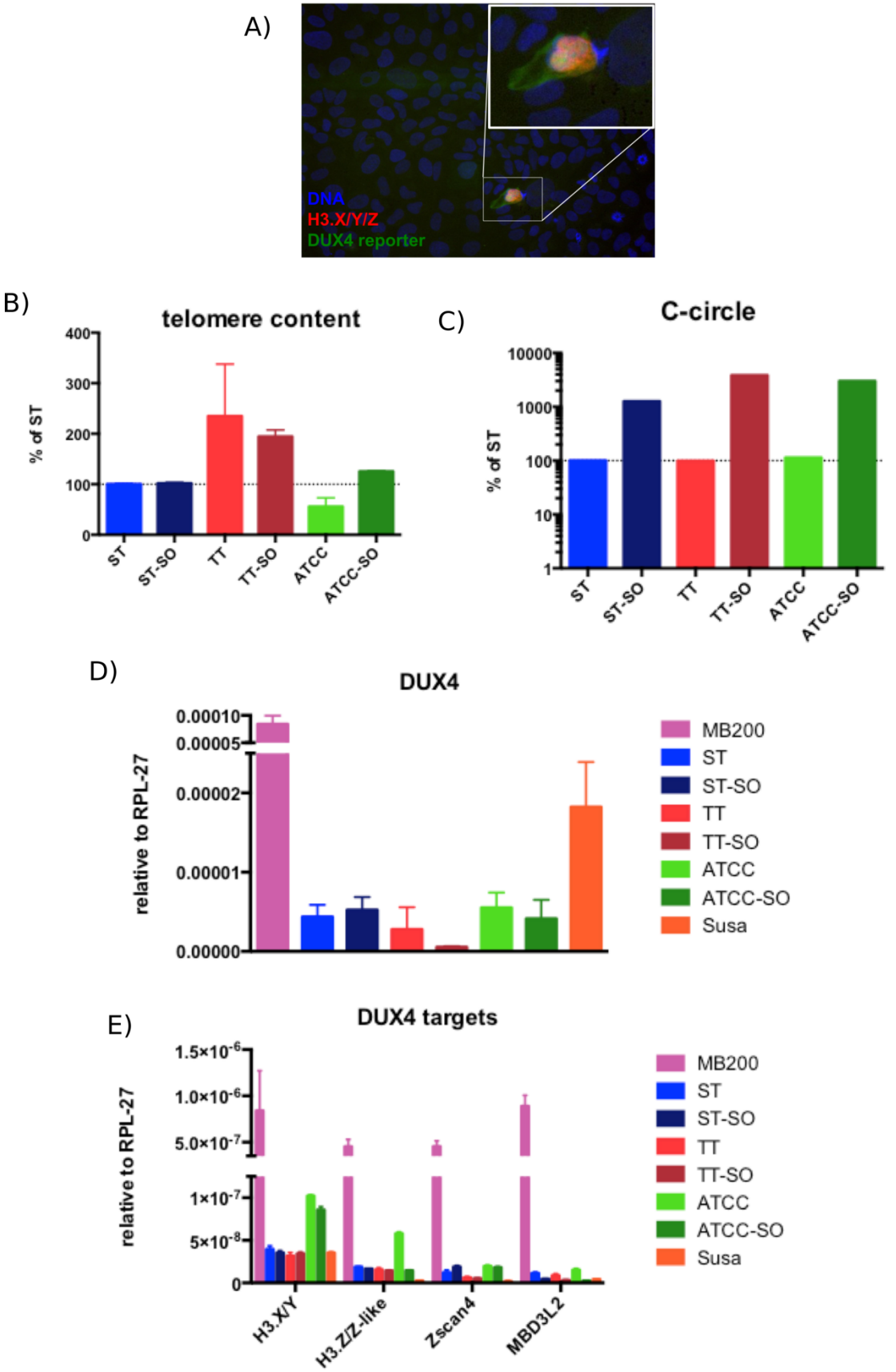

A)
DNA
H3.X/Y/Z
DUX4 reporter
B)
telomere content
% of ST
400
300
200
100
0
ST
ST-SO
TT
TT-SO
ATCC
ATCC-SO
C)
C-circle
% of ST
10000
1000
100
10
1
ST
ST-SO
TT
TT-SO
ATCC
ATCC-SO
D)
DUX4
relative to RPL-27
0.00010
0.00005
0.00002
0.00001
0.00000
MB200
ST
ST-SO
TT
TT-SO
ATCC
ATCC-SO
Susa
E)
DUX4 targets
relative to RPL-27
1.5×10-6
1.0×10-6
5.0×10-7
1×10-7
5×10-8
0
H3.X/Y
H3.Z/Z-like
Zscan4
MBD3L2
MB200
ST
ST-SO
TT
TT-SO
ATCC
ATCC-SO
Susa

Figure 3.3: ALT may increase in U2OS cells with starvation/overgrowth treatment through a *DUX4*-independent mechanism

(A) Immunofluorescence in U2OS cells (ST line) for H3.X/Y and a GFP DUX4-reporter construct show co-localization in rare nuclei. (B) Mean telomere content as measured by qPCR (polymerase-minus C-circle assay) in three independent U2OS cell lines (ST—from Tapscott Lab and used in other figures, TT—from Tsukiyama lab, ATCC—from American Type Culture Collection), with and without starvation/overgrowth (SO) treatment, normalized to untreated ST cells, shows no consistent increase with SO treatment. Error bars are standard deviation of duplicate qPCR reactions. (C) Mean C-circle content measured by qPCR (polymerase-plus C-circle assay) and normalized to untreated ST cells shows an increase in ALT for all three U2OS lines with SO treatment. (D) *DUX4* levels as measured by RT-qPCR show no increase with SO treatment. MB200 FSHD2 and Susa cells are positive controls with higher levels of *DUX4* than U2OS cells. (E) *DUX4* target genes as measured by RT-qPCR show no increase in U2OS cells with SO treatment. MB200 FSHD2 and Susa cells are positive controls that also express *DUX4*.

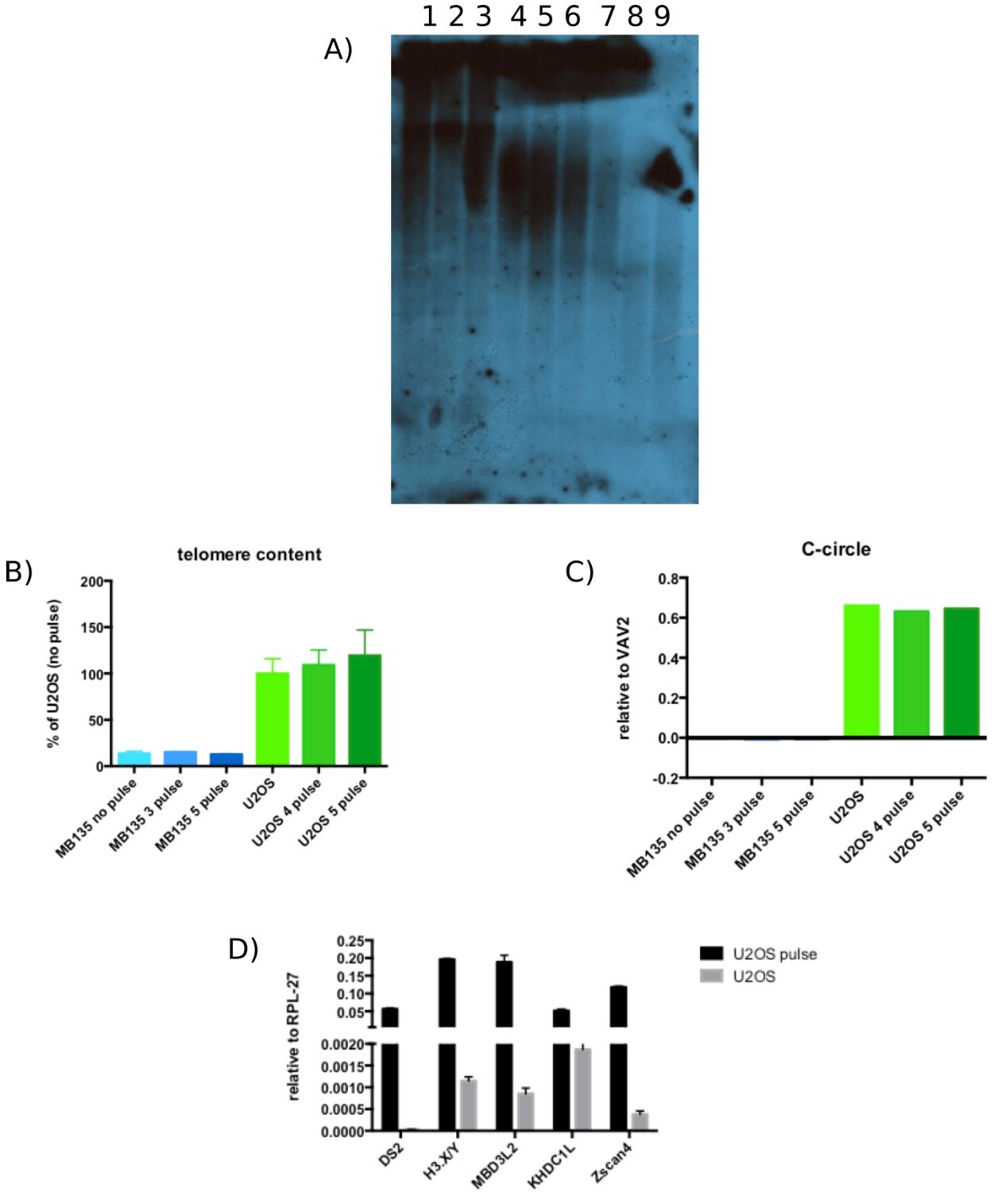

1 2 3 4 5 6 7 8 9
A)
B)
telomere content
% of U2OS (no pulse)
200
150
100
50
0
MB135 no pulse
MB135 3 pulse
MB135 5 pulse
U2OS
U2OS 4 pulse
U2OS 5 pulse
C)
C-circle
relative to VAV2
0.8
0.6
0.4
0.2
0.0
-0.2
MB135 no pulse
MB135 3 pulse
MB135 5 pulse
U2OS
U2OS 4 pulse
U2OS 5 pulse
D)
relative to RPL-27
0.25
0.20
0.15
0.10
0.05
0.0020
0.0015
0.0010
0.0005
0.0000
DS2
H3.X/Y
MBD3L2
KHDC1L
Zscan4
U2OS pulse
U2OS

Figure 3.4: Pulsing DUX4 does not lengthen telomeres or induce/increase ALT in U2OS or wild-type myoblasts

(A) Telomere restriction fragment length analysis by Southern blot. Lanes 1-3 are U2OS cells, 4-6 are primary wild-type MB135 myoblasts, and 7-9 are immortalized wild-type MB135 myoblasts. Lanes 1, 4, and 7 are untreated. Lanes 2, 5, and 8 are DUX4-inducible cells without doxycycline. Lanes 3, 6, and 9 are DUX4-inducible cells with 24 of doxycycline treatment. (B) Mean telomere content of DUX4i wild-type MB135 myoblasts and DUX4i U2OS cells after no treatment, 3 pulses of DUX4, or 5 pulses of DUX4, as measured by qPCR (polymerase-minus C-circle assay). Error bars are standard deviation of biological duplicates. (C) Mean C-circle content measured by qPCR (polymerase-plus C-circle assay) and normalized to single-copy gene VAV2 shows no effect of DUX4 pulses on ALT. (D) *DUX4* transgene (DS2) and target expression in DUXi U2OS cells, untreated or pulsed as in B,C, as measured by RT-qPCR. Error bars are standard deviation of triplicate qPCR wells.

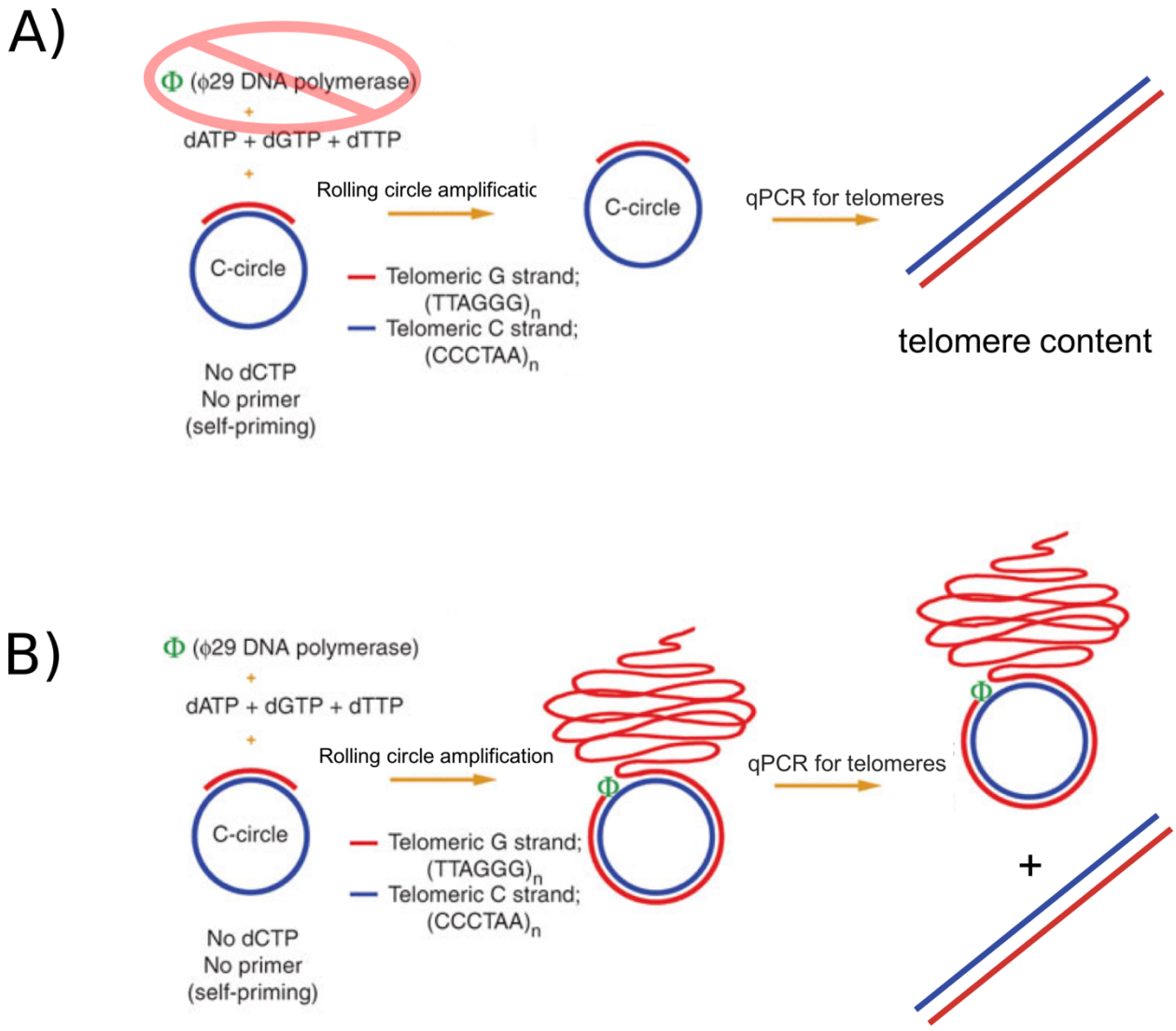


Figure S 3.1: Schematic of C-circle assay

(A) If no phi polymerase is added to the initial reaction, no rolling circle amplification occurs and qPCR using primers for telomeres measures telomere content. (B) When phi polymerase is added to the initial reaction, rolling circle amplification of C-circles occurs and qPCR using primers for telomeres measures the sum of telomere content and C-circles. Subtracting the measurement derived with a polymerase-plus reaction from the polymerase-minus reaction gives C-circle content, a measurement of ALT activation. Figure lightly adapted from [83]

Chapter 4. DISCUSSION

In my thesis work, I set out to characterize the biological roles of histone variants H3.X and H3.Y in development and disease. These variants had been validated as histones but little was known about their biological function when I began this project. The finding that they were robustly induced by *DUX4* gave me a system where I could study the role of these histone variants in a variety of endogenous developmental and disease contexts, including pre-implantation embryos, iPS cells, FSHD, and cancer.

In one component of my project, I showed that DUX4 induction of these histone variants leads to incorporation at both coding and non-coding regions of the genome, including DUX4 targets, and that this incorporation sensitizes DUX4 targets to re-activation. These findings build off of published overexpression studies to show the incorporation pattern of endogenous H3.X/Y and provide evidence of an important functional role for these histones in FSHD pathology and early embryonic development. One exciting aspect of this model is that it could explain the muscle-specific pathology seen in FSHD. The mutations that cause FSHD would presumably lead to *DUX4* de-repression in many tissues throughout the body, yet muscle seems to be uniquely sensitive. The multinucleated, post-mitotic nature of skeletal muscle, however, is somewhat unique among tissues. Sporadic bursts of *DUX4* that can spread between nuclei within a syncytium, combined with an H3.X/Y-based epigenetic memory that sensitizes the cells to subsequent bursts within the same cell, can account for the effects we see in patients.

To further extend this work, I am currently working to determine the mechanism whereby H3.X/Y can affect gene expression at DUX4 target genes, many of which are repressed in the absence of *DUX4*, and why this effect is not seen at constitutively-active genes. *In vitro* studies of H3.Y [52] suggest that it maintains a more open chromatin conformation through looser DNA contacts and exclusion of H1 binding, and this model is consistent with our data as well.

Constitutively-active genes have already established a permissive chromatin context so incorporation of H3.X/Y would have little effect on expression, whereas many *DUX4* targets are in heterochromatin or other repressed regions of the genome and need to be remodeled to become accessible. Though DUX4 is a pioneer factor, it may be able to induce more easily and to a greater degree, especially during a short burst, if the chromatin is already somewhat accessible, which would account for the superinduction seen with a second pulse of *DUX4*. To test this model, I am currently using ATAC-seq to assay chromatin accessibility after a pulse of *DUX4*, with and without *H3.X/Y* knockdown. We are currently analyzing sequencing results from the first round of experiments, looking at DUX4iMB135 cells harvested 48 hours after a pulse of *DUX4.*

In addition to determining a potential role for H3.X/Y in FSHD pathology, this work also identified expression of these histones during the developmental *DUX4* burst. Incorporation of H3.X/Y as a form of chromatin memory of *DUX4* expression in the embryo would serve to prolong the effects of this bursting expression pattern. This provides a mechanism for enhancing the important developmental effects of *DUX4* while modulating its downstream toxicity. Reactivation of this pathway in an inappropriate context such as muscle, however, could lead to enhancement of the toxicity. Further support for this mechanism could come from using iPS cells as a model of the early embryo to determine whether H3.X/Y have a similar incorporation pattern and superinductive effect to that seen in myoblasts. I attempted some of these experiments but found the *DUX4* induction to be inconsistent in these cells, making these experiments technically challenging. Isolation of a more consistent DUX4i clone of iPS cells would make these experiments feasible.

Work from our lab and others showed that a burst of *DUX4* during embryogenesis is not toxic, in contrast with our FSHD models where *DUX4* is potently cytotoxic. We hypothesized that either

there was a protective factor in the embryo that prevented *DUX4* toxicity or that the dosage/kinetics of expression did not reach a toxic threshold. Either of these possibilities has implications for FSHD therapy, as an embryonic protective factor could guide development of a therapeutic. The work described here provides support for the idea that it is the dosage of *DUX4* that is important, since inducible myoblast cells can be pulsed and activate the *DUX4* program with very little or no toxicity. This is an important finding that gives us a new understanding of the kinetics of endogenous *DUX4* expression in FSHD. It is also intriguing to speculate on a beneficial role for *DUX4* in cancer cells since it could potentially help reprogram them to a more stem-like state without killing them. This finding will also enable more long-term studies to examine the downstream effects of *DUX4* on RNA processing, cell secretion, innate immunity, cell cycle, and other important aspects of *DUX4* biology that have previously been difficult to assay, given the short time window before cells succumb to apoptosis with continuous *DUX4* expression.

The RNA-seq experiment that originally identified *H3.X* and *H3.Y* as DUX4 targets also led to the discovery of a related sequence that we have called *H3.Z*. While *H3.Z* is quite similar to *H3.X/Y* in the N-terminal portion of the protein, a frameshift in the histone fold leads to a much longer protein. While H3.Z seems to be expressed at high levels after *DUX4* expression as shown in Chapter 2, we do not know whether it can function as a histone and be incorporated into chromatin, though this seems unlikely given the interruption of the histone fold; bind to chaperones or H4; or whether it is modified. As mentioned in Chapter 2, we presume the effects we see in myoblasts are due to H3.X/Y but this has not formally been tested. Treatment of DUXiMB135 cells with siRNAs that specifically target *H3.Z* would help distinguish the effects of this variant and determine whether the superinduction with a second pulse of *DUX4* is indeed mediated by H3.X/Y.

Work described in Chapter 3 focused on the ALT pathway of telomere elongation and a potential role for *DUX4* or *H3.X/Y* in mediating this pathway, given the many correlations between expression of these genes and activation of ALT. This also provided an attractive model for the bursting expression of *DUX4* during development, as it is currently unknown how *DUX4* is activated at the 4-cell stage or why it is rapidly repressed. An understanding of this biology could be applied towards FSHD treatments, potentially inhibiting the bursting of *DUX4* in muscle or repressing that burst more quickly, before damage to the tissue can occur. My work uncovered no connection between *DUX4* and ALT and suggests that *DUX4* expression is not sufficient to induce or enhance ALT in the cancer or muscle contexts tested.

Further studies to determine whether *DUX4* is necessary for ALT may still prove fruitful, as knocking down *DUX4* would presumably remove *ZSCAN4* from the cells, which is an important component of the ALT pathway. The subtelomeric location of the independent retrotransposition of both human *DUX4* and mouse *Dux*, as well as the conservation of *ZSCAN4* as a target of both transcription factors. suggests that this could be an important, conserved pathway.

In conclusion, my thesis work has furthered our understanding of *DUX4* biology in both development and disease through identification of a role for the histone variants *H3.X* and *H3.Y*. This work has implications for both FSHD pathology as well as pre-implantation human development and zygotic genome activation and provides an additional connection between this earliest developmental pathway and adult disease.

REFERENCES

1. Flanigan, K.M., et al., *Genetic characterization of a large, historically significant Utah kindred with facioscapulohumeral dystrophy.* Neuromuscul Disord, 2001. **11**(6-7): p. 525-9.
2. Mostacciuolo, M.L., et al., *Facioscapulohumeral muscular dystrophy: epidemiological and molecular study in a north-east Italian population sample.* Clin Genet, 2009. **75**(6): p. 550-5.
3. Deenen, J.C., et al., *Population-based incidence and prevalence of facioscapulohumeral dystrophy.* Neurology, 2014. **83**(12): p. 1056-9.
4. Tawil, R., S.M. van der Maarel, and S.J. Tapscott, *Facioscapulohumeral dystrophy: the path to consensus on pathophysiology.* Skelet Muscle, 2014. **4**: p. 12.
5. Goselink, R.J.M., et al., *Early onset facioscapulohumeral dystrophy - a systematic review using individual patient data.* Neuromuscul Disord, 2017. **27**(12): p. 1077-1083.
6. Statland, J.M., et al., *Muscle Pathology Grade for Facioscapulohumeral Muscular Dystrophy Biopsies.* Muscle & nerve, 2015. **52**(4): p. 521-526.
7. Statland, J.M., et al., *Reevaluating measures of disease progression in facioscapulohumeral muscular dystrophy.* Neuromuscul Disord, 2013. **23**(4): p. 306-12.
8. Lutz, K.L., et al., *Clinical and genetic features of hearing loss in facioscapulohumeral muscular dystrophy.* Neurology, 2013. **81**(16): p. 1374-7.
9. Brouwer, O.F., et al., *Hearing loss in facioscapulohumeral muscular dystrophy.* Neurology, 1991. **41**(12): p. 1878-81.
10. Wijmenga, C., et al., *Location of facioscapulohumeral muscular dystrophy gene on chromosome 4.* Lancet, 1990. **336**(8716): p. 651-3.
11. Snider, L., et al., *Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene.* PLoS Genet, 2010. **6**(10): p. e1001181.
12. van Deutekom, J.C., et al., *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, 1996. **5**(12): p. 1997-2003.
13. van Deutekom, J.C., et al., *FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit.* Hum Mol Genet, 1993. **2**(12): p. 2037-42.
14. Wijmenga, C., et al., *Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy.* Nat Genet, 1992. **2**(1): p. 26-30.
15. Lemmers, R.J., et al., *A unifying genetic model for facioscapulohumeral muscular dystrophy.* Science, 2010. **329**(5999): p. 1650-3.
16. Lemmers, R.J., et al., *Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2.* Nat Genet, 2012. **44**(12): p. 1370-4.
17. van den Boogaard, M.L., et al., *Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy.* Am J Hum Genet, 2016. **98**(5): p. 1020-1029.
18. van Overveld, P.G., et al., *Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy.* Nat Genet, 2003. **35**(4): p. 315-7.
19. Hartweck, L.M., et al., *A focal domain of extreme demethylation within D4Z4 in FSHD2.* Neurology, 2013. **80**(4): p. 392-9.

20. Campbell, A.E., et al., *BET bromodomain inhibitors and agonists of the beta-2 adrenergic receptor identified in screens for compounds that inhibit DUX4 expression in FSHD muscle cells.* Skelet Muscle, 2017. **7**(1): p. 16.
21. Chen, J.C., et al., *Morpholino-mediated Knockdown of DUX4 Toward Facioscapulohumeral Muscular Dystrophy Therapeutics.* Mol Ther, 2016. **24**(8): p. 1405-11.
22. Lim, J.W., et al., *DICER/AGO-dependent epigenetic silencing of D4Z4 repeats enhanced by exogenous siRNA suggests mechanisms and therapies for FSHD.* Hum Mol Genet, 2015. **24**(17): p. 4817-28.
23. Marsollier, A.C., et al., *Antisense targeting of 3' end elements involved in DUX4 mRNA processing is an efficient therapeutic strategy for facioscapulohumeral dystrophy: a new gene-silencing approach.* Hum Mol Genet, 2016. **25**(8): p. 1468-78.
24. Ansseau, E., et al., *Antisense Oligonucleotides Used to Target the DUX4 mRNA as Therapeutic Approaches in FaciosScapuloHumeral Muscular Dystrophy (FSHD).* Genes, 2017. **8**(3).
25. Hendrickson, P.G., et al., *Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons.* Nat Genet, 2017. **49**(6): p. 925-934.
26. Whiddon, J.L., et al., *Conservation and innovation in the DUX4-family gene network.* Nat Genet, 2017. **49**(6): p. 935-940.
27. Young, J.M., et al., *DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis.* PLoS Genet, 2013. **9**(11): p. e1003947.
28. Geng, L.N., et al., *DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy.* Dev Cell, 2012. **22**(1): p. 38-51.
29. Feng, Q., et al., *A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy.* Elife, 2015. **4**.
30. Shadle, S.C., et al., *DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy.* PLoS Genet, 2017. **13**(3): p. e1006658.
31. Campbell, A.E., et al., *NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins.* Elife, 2018. **7**.
32. Stadler, G., et al., *Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy.* Nat Struct Mol Biol, 2013. **20**(6): p. 671-8.
33. Rickard, A.M., L.M. Petek, and 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, 2015. **24**(20): p. 5901-14.
34. Jagannathan, S., et al., *Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells.* Hum Mol Genet, 2016.
35. Krom, Y.D., et al., *Intrinsic epigenetic regulation of the D4Z4 macrosatellite repeat in a transgenic mouse model for FSHD.* PLoS Genet, 2013. **9**(4): p. e1003415.
36. Campbell, A.E., et al., *Facioscapulohumeral dystrophy: activating an early embryonic transcriptional program in human skeletal muscle.* Human Molecular Genetics, 2018. **27**(R2): p. R153-R162.
37. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8 A resolution.* Nature, 1997. **389**(6648): p. 251-60.

38. Ahmad, K. and S. Henikoff, *Histone H3 variants specify modes of chromatin assembly.* Proc Natl Acad Sci U S A, 2002. **99 Suppl 4**: p. 16477-84.
39. Tagami, H., et al., *Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis.* Cell, 2004. **116**(1): p. 51-61.
40. Ahmad, K. and S. Henikoff, *The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly.* Mol Cell, 2002. **9**(6): p. 1191-200.
41. Drane, P., et al., *The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3.* Genes Dev, 2010. **24**(12): p. 1253-65.
42. Goldberg, A.D., et al., *Distinct factors control histone variant H3.3 localization at specific genomic regions.* Cell, 2010. **140**(5): p. 678-91.
43. Lewis, P.W., et al., *Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres.* Proc Natl Acad Sci U S A, 2010. **107**(32): p. 14075-80.
44. Elsasser, S.J., et al., *Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells.* Nature, 2015. **522**(7555): p. 240-4.
45. Kong, Q., et al., *Histone variant H3.3-mediated chromatin remodeling is essential for paternal genome activation in mouse preimplantation embryos.* J Biol Chem, 2018. **293**(10): p. 3829-3838.
46. Yuen, B.T., et al., *Histone H3.3 regulates dynamic chromatin states during spermatogenesis.* Development, 2014. **141**(18): p. 3483-94.
47. Jang, C.W., et al., *Histone H3.3 maintains genome integrity during mammalian development.* Genes Dev, 2015. **29**(13): p. 1377-92.
48. Lin, C.J., M. Conti, and M. Ramalho-Santos, *Histone variant H3.3 maintains a decondensed chromatin state essential for mouse preimplantation development.* Development, 2013. **140**(17): p. 3624-34.
49. Howman, E.V., et al., *Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice.* Proc Natl Acad Sci U S A, 2000. **97**(3): p. 1148-53.
50. Tachiwana, H., et al., *Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T.* Proc Natl Acad Sci U S A, 2010. **107**(23): p. 10454-9.
51. Wiedemann, S.M., et al., *Identification and characterization of two novel primate-specific histone H3 variants, H3.X and H3.Y.* J Cell Biol, 2010. **190**(5): p. 777-91.
52. Kujirai, T., et al., *Structure and function of human histone H3.Y nucleosome.* Nucleic Acids Res, 2016. **44**(13): p. 6127-41.
53. Zink, L.M., et al., *H3.Y discriminates between HIRA and DAXX chaperone complexes and reveals unexpected insights into human DAXX-H3.3-H4 binding and deposition requirements.* Nucleic Acids Res, 2017. **45**(10): p. 5691-5706.
54. Adhvaryu, K.K., et al., *Substitutions in the amino-terminal tail of neurospora histone H3 have varied effects on DNA methylation.* PLoS Genet, 2011. **7**(12): p. e1002423.
55. Hyland, E.M., et al., *Insights into the role of histone H3 and histone H4 core modifiable residues in Saccharomyces cerevisiae.* Mol Cell Biol, 2005. **25**(22): p. 10060-70.
56. Maehara, K., et al., *Tissue-specific expression of histone H3 variants diversified after species separation.* Epigenetics Chromatin, 2015. **8**: p. 35.

57. De Iaco, A., et al., *DUX-family transcription factors regulate zygotic genome activation in placental mammals.* Nat Genet, 2017. **49**(6): p. 941-945.
58. Yao, Z., et al., *DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle.* Hum Mol Genet, 2014. **23**(20): p. 5342-52.
59. Krom, Y.D., et al., *Generation of Isogenic D4Z4 Contracted and Noncontracted Immortal Muscle Cell Clones from a Mosaic Patient: A Cellular Model for FSHD.* Am J Pathol, 2012. **181**(4): p. 1387-401.
60. Jones, T.I., et al., *Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: evidence for disease modifiers and a quantitative model of pathogenesis.* Hum Mol Genet, 2012. **21**(20): p. 4419-30.
61. Skene, P.J. and S. Henikoff, *An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites.* Elife, 2017. **6**.
62. Skene, P.J., J.G. Henikoff, and S. Henikoff, *Targeted in situ genome-wide profiling with high efficiency for low cell numbers.* Nat Protoc, 2018. **13**(5): p. 1006-1019.
63. Kowaljow, V., et al., *The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein.* Neuromuscul Disord, 2007. **17**(8): p. 611-23.
64. Wallace, L.M., et al., *DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo.* Ann Neurol, 2011. **69**(3): p. 540-52.
65. Kujirai, T., et al., *Identification of the amino acid residues responsible for stable nucleosome formation by histone H3.Y.* Nucleus, 2017. **8**(3): p. 239-248.
66. Krom, Y.D., et al., *Generation of isogenic D4Z4 contracted and noncontracted immortal muscle cell clones from a mosaic patient: a cellular model for FSHD.* Am J Pathol, 2012. **181**(4): p. 1387-401.
67. Lun, A.T. and G.K. Smyth, *csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows.* Nucleic Acids Res, 2016. **44**(5): p. e45.
68. Falco, G., et al., *Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells.* Dev Biol, 2007. **307**(2): p. 539-50.
69. Lee, K. and L.S. Gollahon, *Zscan4 interacts directly with human Rap1 in cancer cells regardless of telomerase status.* Cancer Biol Ther, 2014. **15**(8): p. 1094-105.
70. Pickett, H.A. and R.R. Reddel, *Molecular mechanisms of activity and derepression of alternative lengthening of telomeres.* Nat Struct Mol Biol, 2015. **22**(11): p. 875-80.
71. Zhang, Q., et al., *Tcstv1 and Tcstv3 elongate telomeres of mouse ES cells.* Sci Rep, 2016. **6**: p. 19852.
72. Liu, L., et al., *Telomere lengthening early in development.* Nat Cell Biol, 2007. **9**(12): p. 1436-41.
73. Baird, D.M., et al., *Telomere instability in the male germline.* Hum Mol Genet, 2006. **15**(1): p. 45-51.
74. Antunes, D.M., et al., *A single-cell assay for telomere DNA content shows increasing telomere length heterogeneity, as well as increasing mean telomere length in human spermatozoa with advancing age.* J Assist Reprod Genet, 2015. **32**(11): p. 1685-90.
75. Heaphy, C.M., et al., *Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes.* Am J Pathol, 2011. **179**(4): p. 1608-15.
76. Henson, J.D. and R.R. Reddel, *Assaying and investigating Alternative Lengthening of Telomeres activity in human cells and cancers.* FEBS Lett, 2010. **584**(17): p. 3800-11.

77. Schwartzentruber, J., et al., *Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma.* Nature, 2012. **482**(7384): p. 226-31.
78. Chawla, R., et al., *Human UPF1 interacts with TPP1 and telomerase and sustains telomere leading-strand replication.* The EMBO Journal, 2011. **30**(19): p. 4047-4058.
79. Wang, F., et al., *Molecular insights into the heterogeneity of telomere reprogramming in induced pluripotent stem cells.* Cell Res, 2012. **22**(4): p. 757-68.
80. Ji, G., et al., *Telomere reprogramming and maintenance in porcine iPS cells.* PLoS One, 2013. **8**(9): p. e74202.
81. Perrem, K., et al., *Coexistence of alternative lengthening of telomeres and telomerase in hTERT-transfected GM847 cells.* Mol Cell Biol, 2001. **21**(12): p. 3862-75.
82. Plantinga, M.J., et al., *Telomerase suppresses formation of ALT-associated single-stranded telomeric C-circles.* Mol Cancer Res, 2013. **11**(6): p. 557-67.
83. Henson, J.D., et al., *DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity.* Nature Biotechnology, 2009. **27**: p. 1181.
84. Henson, J.D. and R.R. Reddel, *Assaying and investigating Alternative Lengthening of Telomeres activity in human cells and cancers.* FEBS Letters, 2010. **584**(17): p. 3800-3811.
85. Henson, J.D., et al., *The C-Circle Assay for alternative-lengthening-of-telomeres activity.* Methods, 2017. **114**: p. 74-84.
86. Kimura, M., et al., *Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths.* Nat Protoc, 2010. **5**(9): p. 1596-607.
87. Lau, L.M., et al., *Detection of alternative lengthening of telomeres by telomere quantitative PCR.* Nucleic Acids Res, 2013. **41**(2): p. e34.

VITA

Rebecca Resnick was born in Seattle, WA, and attended Amherst College where she completed an undergraduate honors thesis. She graduated in 2010 with a Bachelor of Arts degree in Biology. She subsequently spent two years pursuing cancer research at MIT before returning to Seattle to join the Medical Scientist Training Program at the University of Washington. In 2014, she joined the Molecular and Cellular Biology graduate program and began her thesis work in Stephen Tapscott's lab at Fred Hutch. Rebecca received her Doctor of Philosophy in 2018 and returned to medical school in 2019 to complete her training.