© Copyright 2018

Sean C. Shadle

Mechanisms of toxicity in cell models of FSHD

Sean C. Shadle

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

Toshio Tsukiyama

Robert N. Eisenman

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Mechanisms of toxicity in cell models of FSHD

Sean C. Shadle

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

Current consensus places the mis-expression of DUX4 in skeletal muscle as the cause for facioscapulohumeral muscular dystrophy (FSHD). Whereas DUX4 expression normally occurs during early embryonic development, ectopic DUX4 expression is highly toxic in human cells. In my thesis work, using a new cell model with regulated expression of DUX4, I found that expression of DUX4 leads to the stabilization of normally rapidly turned-over mRNAs including cMYC, whose overexpression is associated with cytotoxicity. Moreover, I found that DUX4 expressing cells accumulate large double-stranded RNA (dsRNA) foci within their nuclei, corresponding with induction of the pro-apoptotic PKR dsRNA response pathway. By high-throughput sequencing RNAs immunoprecipitated with anti-dsRNA antibodies I determined that DUX4-induced dsRNAs appeared to originate from intergenic loci that are enriched for *Alu* and LINE-1 elements, endogenous retroviruses, as well as from pericentric sequences enriched for

human satellite II (HSATII) repeats. Similar to pericentromeric repeat expression in the early mouse embryo, bidirectional HSATII transcription appeared to be temporally restricted, with the predominant transcription of one strand preceding the other. DUX4-induced nuclear dsRNA foci correlated with the accumulation of transcripts originating from both HSATII strands. Together, these results identify DUX4 as regulating bidirectional transcription of HSATII repeats that is associated with dsRNA-mediated cell toxicity in FSHD. Furthermore, these findings suggest a similar role for DUX4 in regulating strand-specific HSATII expression in the early embryo.

TABLE OF CONTENTS

List of Figures

List of Supplemental Figures

List of Tables

Acknowledgements

I owe my advisor, Stephen, who is an excellent scientist and mentor, my deepest thanks. I must also thank both current and past members of the Tapscott lab for guiding me along my path. I would like to thank my committee, especially Bob Eisenman (and his lab) for allowing me to sit in and present during their lab meetings. I greatly appreciate others at the Fred Hutch for being so generous with their time, reagents and, most importantly, advice. I am also indebted to Lisa Kursel, who has always been there for me during my graduate school journey.

Chapter 1. *Introduction*

1.1 *Facioscapulohumeral muscular dystrophy*

Facioscapulohumeral muscular dystrophy (FSHD) is a disease that typically manifests as asymmetric weakness of the face, shoulders and upper arms, usually beginning during adolescence or later. Other muscle groups such as those in the lower extremities can also be affected, leading to wheelchair dependency in some patients [1]. The disease develops slowly and is characterized by lengthy stable periods interspersed with rapid bursts of progression [2]. Over the course of the last several decades, the complex genetics underlying FSHD have been slowly worked out and genetic testing now constitutes a major component of clinical diagnosis. While initial attempts to determine the molecular basis for FSHD were stymied by the intricate genetics involved, it is now widely accepted that skeletal muscle mis-expression of the double homeobox domain protein, DUX4, causes FSHD. Below, I will briefly review the major discoveries which helped to clarify the molecular genetics of FSHD and follow with a summary of what is currently understood regarding the regulation of DUX4 expression, the normal function for DUX4 in development, and the role of DUX4-mediated toxicity and its role in disease pathology, which is the major focus of this dissertation.

1.1.1 *Genetics of FSHD*

FSHD is a neuromuscular disease that follows an autosomal dominant inheritance pattern. Early studies in the 1990s used linkage analysis in FSHD families to provide the first evidence that the disease locus likely mapped to chromosomal region 4q35 [3]. This segment of DNA resides adjacent to the telomere and contains a polymorphic repeat structure that is highly variable in length between individuals [4]. Even with this region identified, it was still a number of years before the molecular genetics of FSHD were more completely understood.

Intriguingly, it was determined that FSHD was likely not caused by a protein coding mutation, as is the case in many other inherited diseases. Rather, it was determined that FSHD patients contained fewer of the 4q35 repeat units, known as D4Z4, than unaffected people [4,5]. Paralleling the work on the phenomenon of repeat mediated repression in fruit flies, it was posited that the D4Z4 contraction in FSHD might be associated with the loss of “position effect” repression of adjacent genes [6,7]. Indeed, this line of thought led to the investigation of several candidate FSHD genes which reside adjacent to the 4q35 D4Z4 as the cause for the disease [8]. Evidence against this position effect model for FSHD mounted, however, as transcriptional upregulation of D4Z4 proximal genes in patients could not be consistently demonstrated [9]. Moreover, it was shown that at least one D4Z4 unit was required for disease as distal 4q monosomy is not associated with FSHD [10]. Combined, these findings suggested that the D4Z4 repeat itself might contain the pathogenic sequence.

The D4Z4 is a macrosatellite (>100 bp) repeat with each unit defined as a 3.3 kilobase (kb) KpnI restriction fragment [4]. The D4Z4 is usually arrayed as 11-150 tandem head-to-tail repeats in unaffected individuals but is contracted to 1-10 repeats in most FSHD patients [5]. Early work led to the discovery that the D4Z4 contains an open reading frame for a “double homeodomain” transcription factor named DUX4 [11]. Unfortunately, early attempts to detect the expression of DUX4 in FSHD muscle were unsuccessful [12], hampering progress linking DUX4 to the disease. Further complicating the link between DUX4 and FSHD was the observation that chromosome 10q also contains a D4Z4 repeat array that is not associated with the development of FSHD; contractions of the 10q D4Z4 repeat were not linked with the disease [13]. Serendipitously, FSHD families with meiotic recombination of a pathogenic 4q35 allele onto chromosome 10q were found [13], precluding the possibility that a chromosome 4 D4Z4 proximal gene was responsible for the disease and indicating that a necessary disease associated DNA segment likely lay somewhere in the sequence telomeric to the 4q35 D4Z4.

Indeed, the landmark paper which detailed these findings demonstrated that individuals with FSHD must possess the chromosome 4 D4Z4-distal polymorphism called “4qA” which allows for proper polyadenylation of the *DUX4* transcript [13]. D4Z4 contractions on the “4qB” allelic background do not readily permit polyadenylation of the *DUX4* transcript and are not associated with FSHD. Thus, the recombination of a 4qA repeat onto the contracted 10q D4Z4, which cannot normally be polyadenylated, explained the case of the aforementioned FSHD families with a chromosome 4 to 10 recombination.

Subsequently, a publication provided the first evidence for expression of DUX4 in FSHD but not control cultured patient muscle cells in roughly 0.1% of myonuclei [14], suggesting stochastic, rare reactivation in FSHD. Because the reactivation rate was relatively low, this explained why expression was not readily detected during previous attempts. As a transcription factor, DUX4 was found to bind a consensus TAAYBBAATCA motif (where “Y” is either pyrimidine base and “B” is C or G or T) and activate the expression of early stem cell and germline developmental genes [15]. Notably, DUX4 can bind to and initiate transcription of repeat sequences such as pericentromeric satellite repeat DNA [16] as well as a subset of endogenous retroviral elements [15], which occasionally serve as novel enhancer elements for adjacent genes [16]. It is presently unclear whether or how the transcriptional activity of DUX4 including the activation of repeats contributes to FSHD, but emerging hypotheses are described in detail below.

Combined, this large body of research led to a unifying genetic model centering on the muscular mis-expression of the DUX4 transcription factor as a necessary event for development of FSHD. Still, questions remain concerning both normal regulation of DUX4 as well as its capacity for pathogenesis.

1.1.2 *Regulation of the D4Z4*

Although FSHD is most commonly caused by a D4Z4 repeat contraction and subsequent derepression, approximately 5% of FSHD cases exhibit contraction-independent D4Z4 derepression in the permissive 4qA background [17,18]. Nearly all of these cases are associated with heterozygous mutations to *SMCHD1* [18] which encodes a facilitator of female X-inactivation in mice [19]. SMCHD1 can directly bind to and repress the D4Z4 locus [18,20], although the detailed mechanism of its repressive activity on the D4Z4 is, at present, not entirely understood. Interestingly, levels of SMCHD1 protein, but not RNA appear to decrease during muscle differentiation in culture, correlating with increased DUX4 expression [20]. Though these results need substantiation *in vivo*, they raise the possibility that diminishing SMCHD1 protein levels can “poise” skeletal muscle to be particularly susceptible to stochastic reactivation of DUX4, perhaps explaining why FSHD is a disease of the muscle. Furthermore, because reducing DUX4 levels in FSHD muscle is of therapeutic interest, further work has been done to identify other transcriptional regulators of the D4Z4.

CpG hypomethylation of the D4Z4 is correlated with derepression in FSHD and, fittingly, mutations to the DNA methyltransferase enzyme, *DNMT3B*, appear to be involved in a small portion of non-contracted cases [21]. This indicates that DNA methylation likely plays an important role in repressing the D4Z4. Accordingly, our lab showed that components of the Nucleosome Remodeling Deacetylase (NuRD) complex, which can bind methylated DNA and recruit repressive histone deacetylases and nucleosome remodelers, bind to the D4Z4 and mediate its repression [22]. Moreover, we found that DUX4 activates the expression of an inhibitor of the NuRD complex, MBD3L2, which suggests an ability for DUX4 to auto-regulate its own expression.

While these findings strengthened our mechanistic understanding between DNA methylation and regulation of the D4Z4, the multitude of other factors which have also been

shown to regulate the D4Z4, including HP1γ and cohesin [23], CTCF [24] and Dicer/argonaut [25], highlight the complexity of D4Z4 regulation (see “regulators of DUX4 expression”, Fig 1) and imply other mechanisms of regulation.

In aggregate, the multiple levels of D4Z4 repression is reminiscent of X-inactivation in mammals, which does not appear to rely on a singular complex, but rather “stacks” of imperfect repressive complexes to achieve repression [26]. The redundancy implicit to this model necessitates that if one member were lost then repression would remain largely intact, as is observed with the rare low-level reactivation of DUX4 in *SMCHD1* mutant backgrounds. In further agreement with this idea, individuals with both a loss of function SMCHD1 mutation and a contracted (<11 D4Z4 unit) repeat tend to have more severe FSHD [27], a potential ramification of losing two layers of repression that normally synergize to repress the D4Z4.

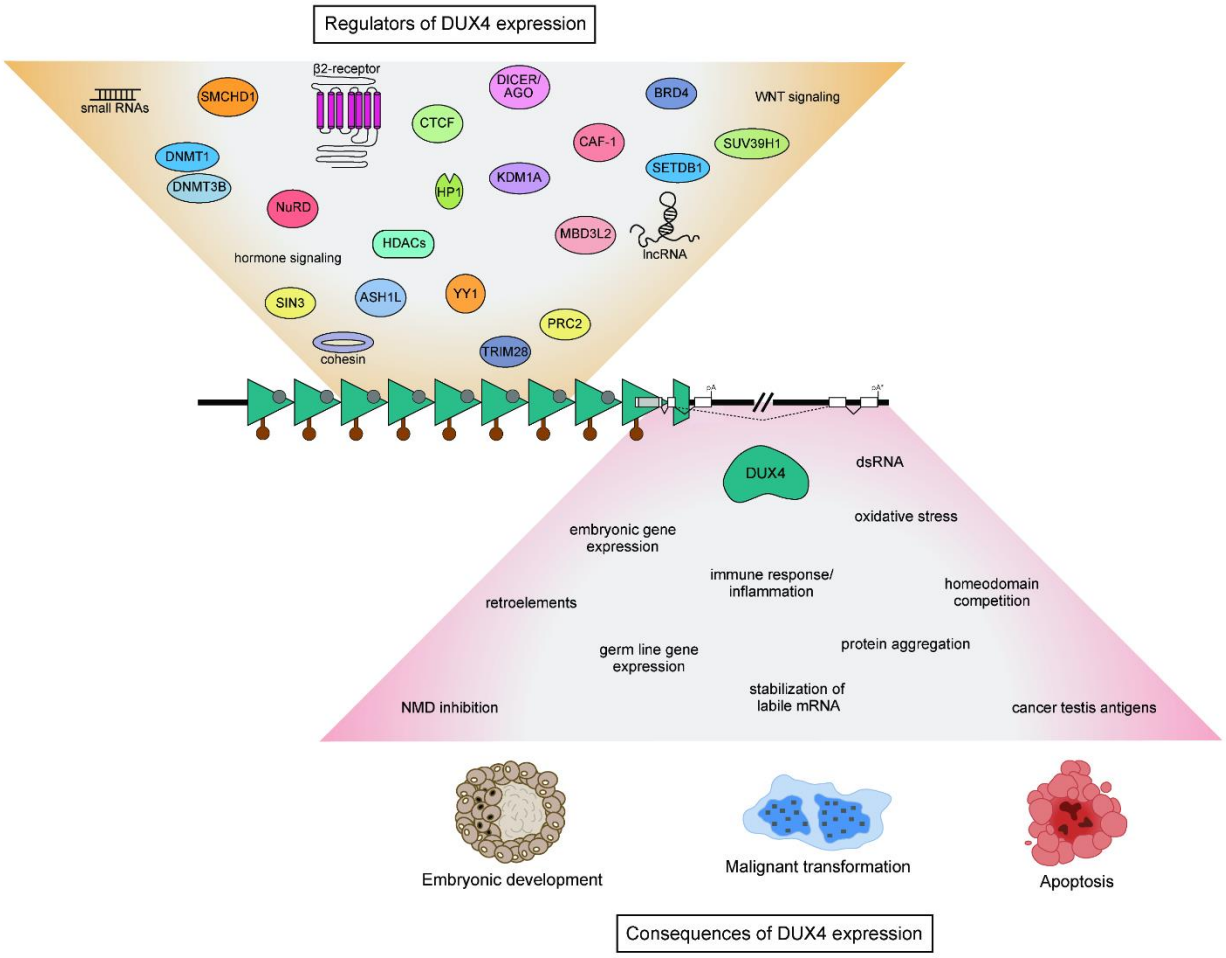


Figure 1 Causes and effects of DUX4 expression. The DUX4 gene (white boxes with the open reading frame in gray) is encoded within each repeat unit (teal triangles) of the D4Z4 macrosatellite array. In healthy individuals the array is marked by high levels of CpG DNA methylation (brown lollipops) and transcriptionally repressive histone modifications (gray circles), which contribute to DUX4 silencing and are reduced in people affected by FSHD. Other factors and pathways known to regulate DUX4 expression are depicted in the tan funnel above the D4Z4 array. Production of stable DUX4 mRNA depends on the presence of a permissive haplotype located immediately distal to the last D4Z4 repeat that contains a functional polyadenylation site (pA), or on a polyadenylation sequence (pA*) found in downstream exons that is likely utilized in the germline. The expression of DUX4 protein is associated with many cellular phenotypes, which are outlined in the inverted pink funnel below theD4Z4 array. Depending on the context, DUX4 may or may not lead to normal development, inappropriate apoptosis or malignant transformation. Adapted from Campbell, *et. al* [28].

1.2 *DUX4 in embryonic development and activation of repeat sequences*

Consistent with its recognized ability to transcriptionally activate early stem cell genes [15], recent studies have shown that DUX4, and its mouse ortholog Dux, are expressed during embryogenesis at the early cleavage stage of development [29,30]. Gene expression analysis suggests that a core set of developmentally important targets are activated by DUX4 in human and Dux in mouse [31]. This indicates a conserved role for the double homeobox factors in activation of the embryonic genome during the transition from maternal to zygotic gene expression (termed embryonic genome activation, or EGA). Emphasizing its critical developmental importance, knockout of mouse *Dux* halts development at the cleavage stage [30], presumably because of a failure to commence EGA.

Data from several studies suggests that DUX4's role in EGA may be in part mediated through activation of repetitive elements including endogenous retro-elements and satellite repeats. Mouse embryonic stem cells (mESCs) can transition into a 2-cell or "2C-like" state spontaneously in culture which is thought to represent a reversion to an earlier stage expression program [32]. The 2C-like state is characterized by reactivation of MERVL endogenous retroviral sequences which can serve as promoters for EGA genes [33]. Similar to how DUX4 can bind to and activate retroelements such as HERVL and MaLRs in human, mouse Dux can bind to MERVL and activate its expression [29,31]. Moreover, overexpression of mouse Dux is highly effective at converting mESCs into 2C-like cells [29], further suggesting that Dux can drive the cleavage state. One morphological consequence of 2C-like transition is the loss of the DAPI-dense repressive "chromocenters" [34], large condensed heterochromatic regions. Counter-intuitively, the transcription of repeats in (peri)centromeric regions is often necessary for maintenance or establishment of this heterochromatin. Indeed, expression of mouse pericentric repeats has been observed in 2-cell mouse embryo [35] a developmental time point where dramatic chromatin reorganization takes place. Suggesting a conserved developmental

role, both human DUX4 and mouse Dux activate transcription of pericentromeric satellite repeats [16,29]. Zygotic depletion of these satellite transcripts in mice halts developmental progression at the cleavage stage [35], further illustrating the developmental necessity of these Dux-induced noncoding repeat transcripts.

1.3 *DUX4 and FSHD pathogenesis*

DUX4 is toxic to many human and mouse cell lines in which it has been overexpressed and it exhibits toxicity *in vitro* using mouse models of FSHD [28]. There are many known downstream consequences of DUX4 expression which might contribute to its toxicity (see "consequences of DUX4 expression", Fig 1), though a consensus model for toxicity is currently lacking. A major aim of my thesis work is to mechanistically understand DUX4 toxicity and relate it to DUX4 expression as well as to its role in the early embryo. In this section, I will briefly highlight current hypotheses as to how DUX4 expression may lead to cytotoxicity.

Early work expressing DUX4 in an immortalized mouse muscle cell line led to the discovery that DUX4 toxicity might operate through the activation of oxidative stress pathways [36]. More recently, treatment of cells with antioxidants demonstrated some capacity to alleviate DUX4 toxicity in human tumor cell lines as well [37]. While it is attractive to speculate that DUX4 expression leads to the accumulation of reactive oxygen species (ROS) via activation of a target or set of targets, it is still not clear which pathways activated by DUX4 would be responsible for ROS accumulation.

A non-mutually exclusive hypothesis for DUX4 pathogenesis is that DUX4 can compete with and repress homeodomain transcription factors, PAX3/PAX7 [36], factors which are critical for generating skeletal muscle lineage. Mechanistically, DUX4 may compete with PAX3/PAX7 for targets based on similarities of their homeodomains [36]. In line with this idea of competitive binding, the overexpression of mouse Pax3 or Pax7 alleviated toxicity of DUX4 in mouse muscle cells expressing human DUX4 and FSHD muscle shows repressed PAX7 target gene

expression [38]. Moreover, PAX7 target repression correlates with activation of a hypoxic response, potentially connecting DUX4 to the oxidative stress response [38].

Another more recent finding demonstrated that exogenous DUX4 expression can lead to nuclear focal aggregation of TDP-43 [39], a protein which is found to be aggregated as detergent insoluble, nuclear “inclusions” in neurological diseases such as amyotrophic lateral sclerosis [40]. One known function for TDP-43 includes pre-mRNA splicing [41], a molecular process which is, intriguingly, also disrupted in DUX4 expressing cells [42]. Furthermore, the *Caenorhabditis elegans* ortholog of TDP-43, TDP-1, appears to suppress the accumulation of RNAs with double-stranded secondary structure (dsRNAs), which was offered as an explanation for some of the splicing abnormalities seen in TDP-1 (and TDP-43) mutants [43]. Disruption of TDP-43 via knockdown leads to dsRNA accumulation [43] and is associated with activation of the pro-apoptotic interferon response [44], an innate immune pathway which evolved to respond to viral infection and can be triggered by dsRNA accumulation. With TDP-43 accumulation, it could be hypothesized that dsRNAs play a role in DUX4 toxicity, yet prior to my thesis work, it was unknown whether DUX4 expression could modulate endogenous dsRNA levels and the associated downstream pathways.

In humans, endogenous double-stranded RNAs are often comprised of repeat sequence transcripts [45]. As the expression of repeats would presumably be detrimental to many cell types, they are constitutively repressed through mechanisms such as DNA methylation [46]. The idea that inappropriate repeat activation by DUX4 is connected to FSHD has been previously put forth [15], and essentially takes the following form: induction of endogenous retroviruses by DUX4 may lead to activation of an innate immune response and/or genome instability in FSHD muscle. In agreement with the idea that repeat activation can trigger an innate immune response, tumor cell lines treated with a DNA demethylating agent, 5-azacytidine, reactivate repetitive elements and the bidirectional transcription of endogenous retroviruses in these studies was correlated to an induction of the interferon response, which

appeared to be provoked by repeat dsRNA accumulation [47,48]. Other repeat transcripts such as pericentric satellite RNAs are bidirectionally transcribed [49] and can likely form dsRNA in both the early embryo [50] and certain tumors [51]. Additionally, destabilization of replication forks by RNA-DNA hybrids formed from transcription of satellite repeats can potentially induce DNA damage, genomic instability and cell death [52]. Cumulatively, it is apparent that the reactivation of repeats by DUX4 might be detrimental to cells by either invoking a cellular response to dsRNA and/or activating a DNA-damage response that could lead to cytotoxicity, though detailed studies exploring repeat-mediated DUX4 toxicity are lacking.

1.4 *Dissertation overview*

My thesis work aims specifically to address how DUX4 causes cellular toxicity. To approach this line of research, I first set out to build a model system of FSHD which would faithfully recapitulate endogenous characteristics of DUX4 expression in patient muscle cells. I therefore produced DUX4 inducible clonal cell lines which would be ideal for scaling and reproducibility for screening purposes. Likely because of silencing mechanisms which control expression of DUX4 (described above), my initial attempts to produce a DUX4 inducible myoblast cell line were unsuccessful. In Chapter 2, I describe the approach that was used to overcome this barrier by altering the codon structure of DUX4. Comparison of this inducible DUX4 to other datasets, including the endogenous expression of DUX4 in FSHD myocytes, indicated high transcriptomic similarities, validating its use as a model system to study downstream effects of DUX4 expression. Importantly, expression of DUX4 using this inducible system leads to profound cell death.

In Chapter 3, I utilized the DUX4 inducible system to conduct a largely unbiased high-throughput screen using siRNAs targeting the human “druggable” genome to address which pathways might be involved in triggering DUX4 toxicity. Using this approach, I uncovered an unexpected role for DUX4 in stabilizing MYC mRNA. Given that MYC overexpression can lead

to cellular apoptosis, and that knockdown of MYC or its dimerization partner, MAX, in DUX4 expressing cells diminishes cell death, I hypothesize that its stabilization can contribute to DUX4 toxicity. Intriguingly, I also found that knockdown of either PKR or RNASEL, two mediators of the innate immune response to double-stranded RNAs, can rescue cells from DUX-mediated toxicity. This finding led to the surprising discovery that DUX4 expressing cells can induce massive nuclear focal aggregation of dsRNAs. Because repeat transcripts are known to form dsRNA, I posited that DUX4 activation of repeat sequences can lead to the formation of these large dsRNA aggregates

In Chapter 4, I asked which specific sequences make up DUX4-induced dsRNAs using an antibody-based dsRNA immunoprecipitation approach coupled with high-throughput RNA sequencing. As expected, dsRNAs in DUX4 expressing cells were largely composed of repeat sequences and include those specifically activated by DUX4. Pericentric “HSATII” satellite repeats which are activated by DUX4, comprised a unique fraction of DUX4-induced dsRNAs and perfectly coincided with nuclear dsRNA aggregates. As with pericentric repeat expression in the cleavage embryo of mice, DUX4 expression led to predominant transcriptional activation of one strand of HSATII repeats, followed temporally by the opposite strand, suggesting dsRNA formation via bidirectional transcription. Supporting a role for bidirectional HSATII transcripts in DUX4 toxicity, the timing of dsRNA formation and transcription of both HSATII strands corresponded well with the activation of the dsRNA responder PKR as well as death of the expressing cells. Finally, I demonstrated that variation within the HSATII repeat sequence creates an ideal binding motif for DUX4, indicating that activation is likely direct.

As expression of satellite repeats is vital to embryonic development [35], it is expected that a critical role of DUX4 in the early embryo is in activating expression of HSATII noncoding RNAs and other repeats. While repeat expression might be essential for early human development, activation of these normally repressed sequences could be incompatible with

somatic tissues such as muscle because of their ability to form dsRNA and/or cause genome instability.

Chapter 2. *A Model system for studying FSHD toxicity*

A version of this chapter has been previously published as:

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. Human molecular genetics. 2016 Aug 15;25(20):4419-31.

**These authors contributed equally to this work*

2.1 *Abstract*

Facioscapulohumeral dystrophy (FSHD) is caused by the mis-expression of the double-homeodomain transcription factor DUX4 in skeletal muscle cells. Many different cell culture models have been developed to study the pathophysiology of FSHD, frequently based on endogenous expression of DUX4 in FSHD cells or by mis-expression of DUX4 in control human muscle cells. Although results generated using each model are generally consistent, differences have also been reported, making it unclear which model(s) faithfully recapitulate DUX4 and FSHD biology. In this study, we systematically compared RNA-seq data generated from three different models of FSHD—lentiviral-based DUX4 expression in myoblasts, doxycycline-inducible DUX4 in myoblasts, and differentiated human FSHD myocytes expressing endogenous DUX4—and show that the DUX4-associated gene expression signatures of each dataset are highly correlated (Pearson's correlation coefficient, r ~ 0.75-0.85). The few robust differences were attributable to different states of cell differentiation and other differences in experimental design. Our study describes a model system for inducible DUX4 expression that enables reproducible and synchronized experiments and validates the fidelity and FSHD relevance of multiple distinct models of DUX4 expression.

2.2 *Introduction*

Facioscapulohumeral dystrophy (FSHD) is a prevalent form of muscular dystrophy that is currently untreatable and incurable. FSHD is caused by the derepression of the D4Z4 macrosatellite array at chromosome 4q35 [4], resulting in the ectopic expression of the DUX4 gene encoding a germline transcription factor. DUX4 expression is toxic to somatic cells in culture and leads to muscle atrophy in vivo [36,53–56]. These data, as well as strong genetic evidence demonstrating an essential requirement for at least one copy of a polyadenylation-competent DUX4 gene in FSHD, implicate DUX4 as the primary driver of FSHD [13].

Consequently, determining whether there is a core set of molecular pathways dysregulated by DUX4 that correlate with transcriptional abnormalities in FSHD cells is of paramount importance to uncover the mechanisms behind DUX4-induced myopathy and develop effective models of the disease for therapeutic development.

The gene expression signature of DUX4 and FSHD has been characterized by several studies, many of which used different model systems of DUX4 expression and/or FSHD disease models and tissues. Such studies have reported diverse genes and biological pathways that are affected in FSHD and by DUX4 expression, including early stem cell and embryonic genes [15,57], cancer testis antigens [15,57], retroelements and repetitive elements [15,16], genes involved in RNA processing and splicing [15,57], inflammatory response [15], myogenesis [15,36,37], Wnt/β-catenin signalling pathway [58,59], oxidative stress response [37], protein homeostasis [39], RNA quality control [42], and PAX targets [36,59], among others. However, not all of these affected genes and/or pathways have been consistently reported by all studies. These differences might reflect the complexity of the disease and raise the possibility that disease models that reproduce all of the major aspects of FSHD might remain out of reach [60].

An alternative perspective is that because DUX4 expression causes FSHD, models based on the expression of DUX4 in skeletal muscle cells should reproduce the major cell-autonomous transcriptional features of FSHD. This perspective is supported by the finding that muscle biopsies from FSHD-affected individuals mis-express the same set of genes that are also upregulated in cultured FSHD muscle cells and in control skeletal muscle cells transduced with a DUX4-expressing vector [61]. However, a recent study reported a rather limited overlap between the gene expression signature associated with expression of endogenous DUX4 in FSHD myocytes and the set of DUX4-regulated genes that were previously identified by transduction of non-FSHD human myoblasts with a lentivirus expressing DUX4 [57]. Although the authors carefully noted that the differences might arise from different biological and/or technical variables, this finding raised the concern that exogenous expression of DUX4 in non-

FSHD skeletal muscle cells might not be a good model for gene expression changes that occur in FSHD muscle cells.

FSHD is a complex disease and it is important to determine whether different model systems yield disparate results or whether they show convergence on similar biological pathways. Differences in DUX4-induced genes in different model systems of FSHD could arise from true biological differences, distinct cell types or contexts, and/or technical differences, such as the use of different data analysis strategies. Here, we used a consistent data analysis pipeline to determine the gene expression profile of three different cell culture models of FSHD: control muscle cells with an inducible DUX4, control muscle cells transduced with a lentiviral vector expressing DUX4, and FSHD muscle cells expressing endogenous DUX4. We report that all three cell culture models exhibited highly similar gene expression patterns and that the few differences, such as viral immune response or myogenic differentiation, were attributable to the differences in experimental design. These results indicate that expressing DUX4 in control skeletal muscle cells accurately recapitulates DUX4-associated differences in gene expression observed in FSHD muscle cells that endogenously express DUX4, and license multiple distinct models of DUX4 expression as relevant tools for studying FSHD biology.

2.3 *Results*

2.3.1 *Codon altering DUX4 enables inducible expression*

The D4Z4 repeat locus is normally hypermethylated and epigenetically silenced in somatic tissue. The DUX4 open reading frame (ORF) has 130 individual CpG sites that might be subject to DNA methylation, suggesting that the high (73%) GC content of the DUX4 cDNA could contribute to gene silencing. We therefore speculated that reducing the CpG content of the DUX4 cDNA might be required for efficient transgene expression. Hence, we re-coded the DUX4 cDNA to reduce the total number of CpG sites while preserving the protein sequence (Fig. 2A). The codon-altered and wild-type DUX4 ORF sequences were then cloned into the

pCW57.1 lentiviral vector (Addgene plasmid #41393) such that a doxycycline-inducible promoter regulates DUX4 and the puromycin resistance gene (puroR) is constitutively driven by the PGK promoter. Clonal cell lines expressing wild-type or codon-altered DUX4 were isolated from immortalized human control myoblasts (MB135) transduced with the corresponding DUX4 constructs using puromycin selection. None of the fifteen tested wild-type DUX4 clones exhibited apoptosis following doxycycline induction, whereas three of five tested codon-altered DUX4 clones exhibited complete cellular detachment/death (Fig. 2B). These data suggest that DUX4 is not efficiently expressed in wild-type DUX4 cells, and that reducing the CpG content of DUX4 cDNA overcomes this barrier to expression ($P = 0.001$ by the two-sided binomial test for equality of proportions).

To confirm that DUX4 was induced by doxycycline, we measured transgene mRNA and DUX4 protein levels, as well as levels of ZSCAN4 mRNA, a direct transcriptional target of DUX4, in five individual codon-altered DUX4 and five wild-type DUX4 myoblast clones upon doxycycline induction. None of the five wild-type DUX4 clones showed induction of DUX4 mRNA, DUX4 protein, or ZSCAN4 mRNA, whereas all of the codon-altered DUX4 clones showed induction of the transgene (Fig. 2C). Three of the five codon-altered DUX4 clones—the same clones that exhibited cell death upon doxycycline addition (Fig. 2B)—showed induction of full-length DUX4 protein and activated ZSCAN4 transcription (Fig. 2C and D). In contrast, the two codon-altered DUX4 clones that did not exhibit cell death upon doxycycline addition expressed truncated forms of the DUX4 protein and did not activate ZSCAN4 expression (Fig. 2C), suggesting that the absence of cell death in these clones was due to a deletion or truncating mutation that prevented DUX4 activity and subsequent cytotoxicity. Together, these results indicate that decreasing the GC content of the DUX4 ORF likely facilitates DUX4 mRNA expression from an integrated DUX4 transgene. Although it remains to be determined, the mechanism might be due to more efficient silencing of the WT-DUX4 transgene because the inducible expression of the WT-DUX4 was initially robust and declined over time following

transduction compared to the CA-DUX4, and transduction with WT-DUX4 lentivirus produced fewer puromycin resistant colonies compared to CA-DUX4 lentivirus used at equivalent titers (Supplementary Material, Fig. S1A and B), however, further work will be needed to determine whether this represents epigenetic silencing or another process.

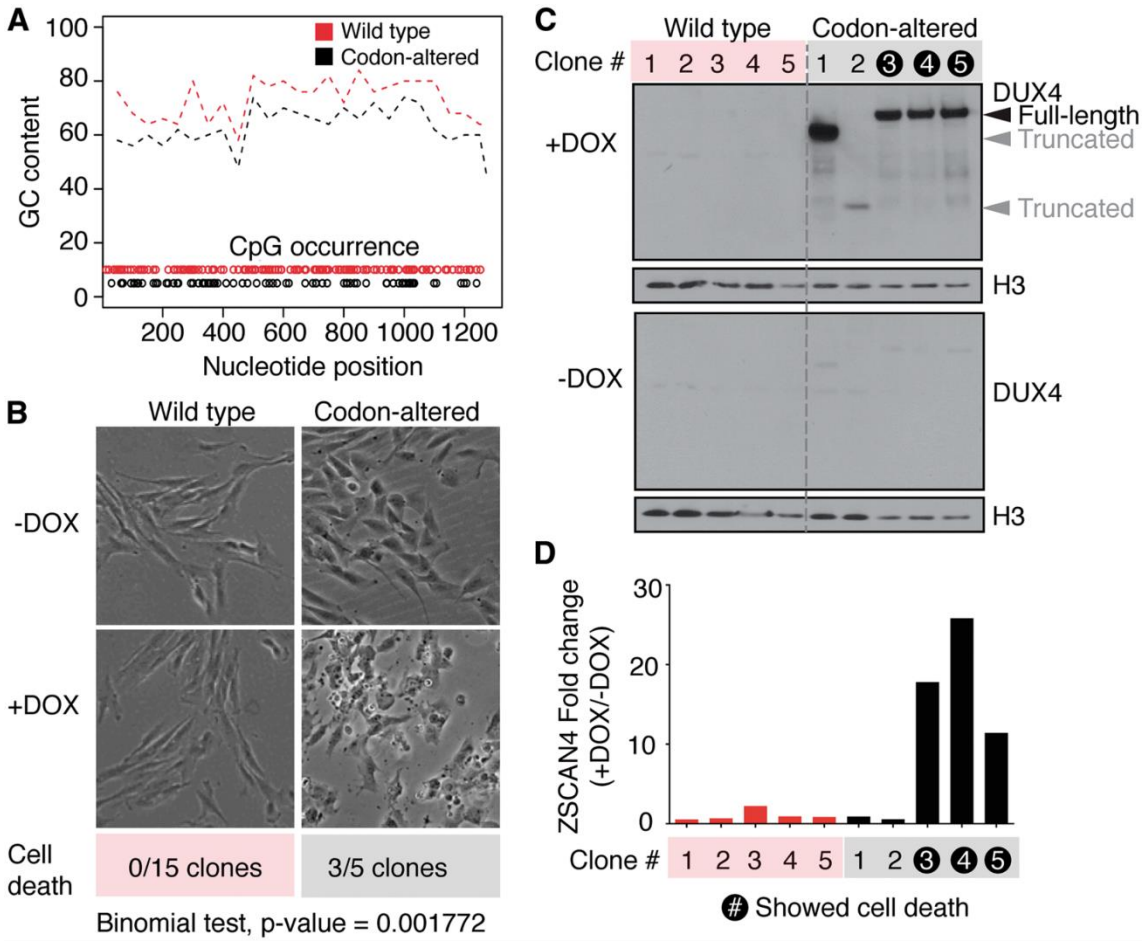


Figure 2 Codon altering allows stable, inducible expression of DUX4 in human myoblasts. (A) Graphical depiction of GC percentage and CpG occurrence of the codon-altered (black) and wild-type (red) DUX4 coding regions. GC percentage was calculated over 50 base pair sliding windows. The positions of CpG dinucleotides are indicated by open circles. (B) Phase contrast images of monoclonal cells encoding wild-type or codon-altered DUX4 expression constructs, with or without doxycycline induction. The number of clonal cell lines that exhibited cell death among all that were tested is shown. (Binomial test for equality of

proportions, P Value = 0.001). (C) Western blot analysis for DUX4 expression on lysates from 5 clones encoding wild-type or codon-altered DUX4, with or without induction with doxycycline for 8 hours. Histone 3 (H3) serves as a loading control. Black arrowhead indicates full-length DUX4 product. (D) qRT-PCR analysis of a DUX4 transcriptional target, ZSCAN4, shown as fold-change over uninduced cells in the various clones. The clones that exhibited cell death upon doxycycline induction are highlighted.

2.3.2 *Codon-altered DUX4 faithfully reproduces transcriptional and post-transcriptional dysregulation previously reported for wild-type DUX4*

We next confirmed that codon-altered DUX4 induction drove the same transcriptional response and inhibition of RNA quality control that have been previously reported for DUX4 [15,16,42]. We conducted RNA-seq on a codon-altered DUX4 myoblast cell line 14 hours after doxycycline induction (referred to as 'iDUX4' hereafter) or the same time point without doxycycline. We sequenced each sample to a depth of ~50 million reads to allow accurate measurement of isoform expression. A scatter plot of gene expression showed robust DUX4-induced expression of hundreds of genes as well as mild, but statistically significant, repression of a smaller set of genes (Fig. 3A). Known targets of DUX4, such as ZSCAN4, several PRAME genes, LEUTX, TRIM43 and KHDC1L, were highly activated (Fig. 3B). Some of the most downregulated genes were involved in the immune response, such as IL7R, and the extracellular matrix, both pathways that were previously shown to be affected by DUX4 expression (by [15] and [57], respectively). Specific classes of endogenous retroviral elements and repetitive genomic sequences, including HSATII, were also upregulated, consistent with previous reports [16] (Fig. 3C and D). Finally, RNA isoforms that contain premature termination codons and are normally degraded via nonsense-mediated RNA decay (NMD) exhibited high expression in cells expressing DUX4 (Fig. 3E and F), as previously reported [42]. Together, these data demonstrate that codon-altered DUX4 faithfully recapitulates the transcriptional and

post-transcriptional changes that previous studies have reported for expression of wild-type DUX4 in muscle cells.

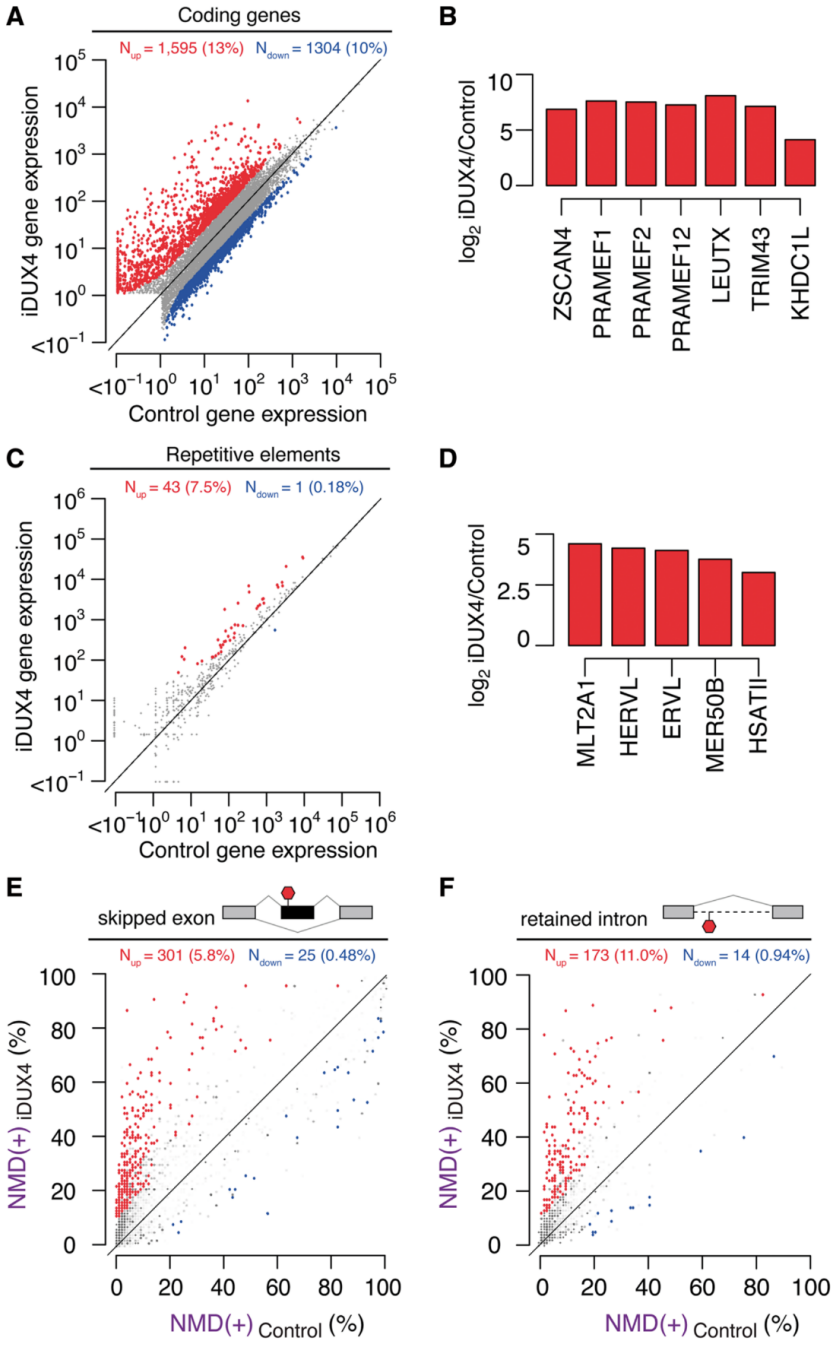


Figure 3 Inducible expression of codon-altered DUX4 activates germline antigens, endogenous retrotransposons and repetitive elements and inhibits RNA quality control. (A) Scatter plot of gene expression (in transcripts per million) in control (uninduced) versus

iDUX4 (doxycycline-induced) myoblasts. Red/blue, genes exhibiting increases/decreases of > 2.5 fold. (N) Numbers of genes with increased/decreased expression; (percentages) fraction of genes that are affected by DUX4 expression. (B) Relative mRNA levels of known DUX4 transcriptional targets in iDUX4 versus control myoblasts expressed as log2 fold-change. (C) Scatter plot of repetitive element expression (in transcripts per million) in control (uninduced) versus iDUX4 (doxycycline-induced) myoblasts. Red/blue, repeat elements exhibiting increases/decreases of > 2.5 fold. (N) Numbers of repetitive elements with increased/decreased expression; (percentages) fraction of repetitive elements that are affected by DUX4 expression. (D) Relative levels of known DUX4-activated repetitive elements in iDUX4 versus control myoblasts expressed as log2 fold-change. (E, F) Isoform ratios of predicted NMD substrates generated by cassette exon alternative splicing (E) or intron retention (F), comparing iDUX4 versus control myoblasts. Red/blue, cassette exons (E) or retained introns (F) exhibiting increases/decreases of ≥10% in isoform ratios for the isoforms that are predicted NMD substrates. Events that do not change significantly are rendered transparent.

2.3.3 *Comparison of inducible, viral and endogenous DUX4 expression systems*

A recent study of the transcriptome regulated by endogenous DUX4 in differentiated FSHD muscle cells [57] reproduced many of the changes in RNA abundance previously shown to be mediated by DUX4 in myoblasts transduced with a DUX4-expressing lentivirus [15,16,42]. However, this study also reported that a large component of the gene expression signature in these FSHD myocytes only occurred in the endogenous expression system. These results raised the possibility that the delivery of exogenous DUX4 to control muscle cells might not faithfully recapitulate the changes in RNA expression mediated by endogenous DUX4 in FSHD muscle cells. However, the different datasets were acquired using different gene expression profiling platforms and analyzed using different statistical methodologies, preventing direct assessment of similarities and differences in the consequences of exogenous versus endogenous DUX4 expression.

We therefore performed a systematic comparison of DUX4-regulated changes in the transcriptome in our inducible codon-altered DUX4 expression system (iDUX4), the endogenous DUX4 expression system (enDUX4), and cells transduced with lentivirus constitutively expressing DUX4 (vDUX4). The specific datasets used in this comparison are as follows: iDUX4 represents a new dataset generated from the MB135 immortalized human myoblasts with the doxycycline inducible codon-altered DUX4 (iDUX4), performed in biological triplicate fourteen hours after DUX4 induction in growth media, with uninduced cells as a control; enDUX4 represents the published dataset of differentiated FSHD myocytes that do or do not express endogenous DUX4, as determined using a DUX4-responsive fluorescent reporter and flow sorting [57]; vDUX4 represents a published dataset wherein two different myoblast cell lines (MB135 and 54-1) were transduced with a lentiviral construct that drives constitutive DUX4 expression via the PGK promoter and maintained in growth media for 24 hours (MB135) or 36 hours (54-1) prior to harvesting RNA [42,61]. More information about the datasets, including the RNA-seq methodology and the number of mapped reads in each sample, is presented in Table 1.

Distinct RNA-seq analysis strategies and statistical methodologies can produce very different results, including the identification of only partially overlapping sets of differentially expressed genes in a given dataset [62]. To avoid this confounding factor, we quantified differential gene expression in the enDUX4, iDUX4, and vDUX4 datasets using a common read mapping and analysis pipeline (Fig. 4A; described in Materials and Methods).

We first compared the global extent of gene induction versus repression associated with DUX4. Plotting average gene expression versus DUX4-associated change in gene expression (“MA” plots) revealed that DUX4 preferentially caused gene upregulation, as expected, in all three datasets. However, the global extent of gene induction or repression varied between the different models of DUX4 expression (Fig. 4B and D, Supplementary Material, Table S1). Endogenous DUX4 expression was associated with more modest changes in gene expression

Table 1 Description of samples and datasets used in this study

Sample names	Sample description	RNA-seq details	Number of replicates	RNA-seq read count (control, dux4)
iDUX4, control (Pilot dataset in Figure 3)	Stable DUX4 expression induced with doxycycline; uninduced sample serves as Control	50bp SR	1	**iControl**: 57,537,306; **iDUX4**: 50,000,472
iDUX4, control	Stable DUX4 expression induced with doxycycline; uninduced sample serves as Control	100bp SR	3	**iControl**: 19,409,784; 17,221,415; 13,447,636; **iDUX4:** 16,185,774; 18,680,623; 18,129,148
enDUX4, Control	FSHD patient-derived cell line expressing DUX4 spontaneously during differentiation. DUX4-positive cells isolated by sorting for DUX4-induced BFP.	50bp PE, treated as 50bp SR	6	**enControl:** 22,972,559; 17,296,151; 22,392,500; 22,745,205; 28,530,475; 20,704,644. **enDUX4**: 20,653,257; 17,695,221; 19,343,364; 24,964,611; 21,690,419; 20,372,625
vDUX4, Control	DUX4 expression via lentiviral delivery; samples expressing GFP via lentiviral infection serve as control.	100bp SR	2	**vControl**: 76,520,414 (MB135); 74,669,369 (54-1); **vDUX4**: 97,840,848 (MB135); 70,138,030 (54-1)

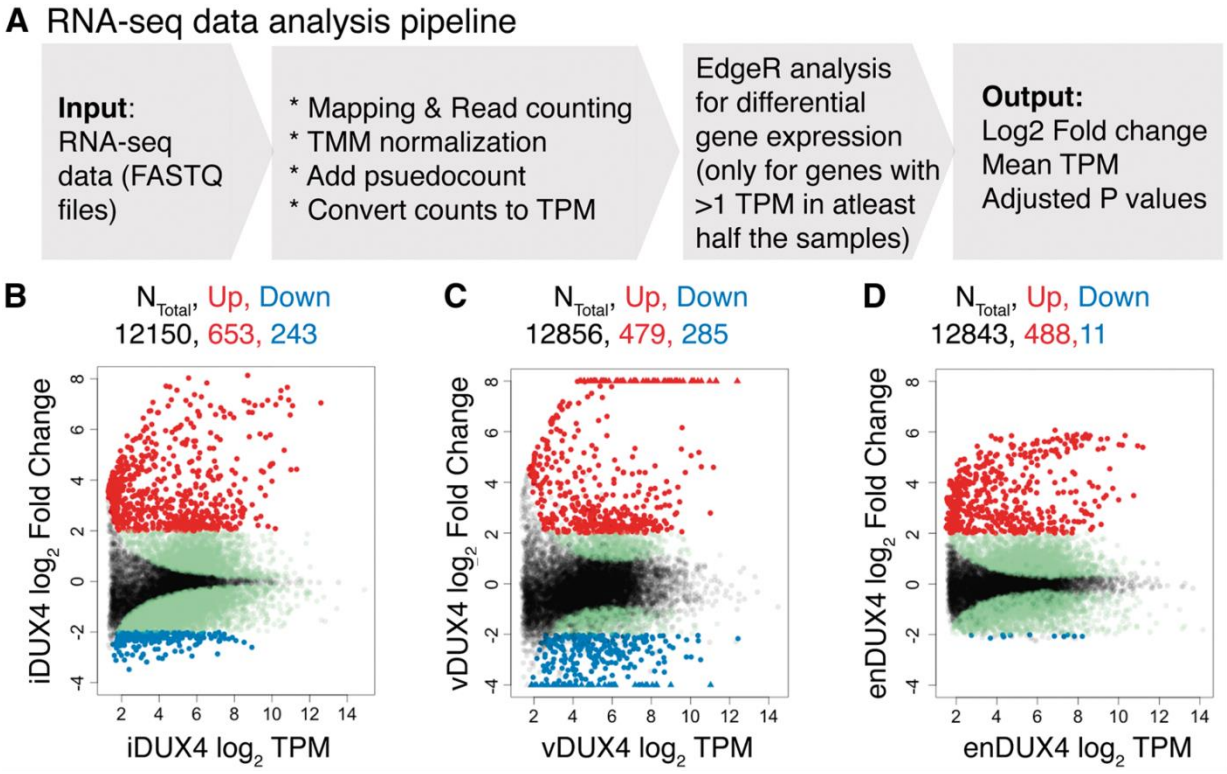


Figure 4 Transcriptional response of endogenous and exogenous DUX4 expression in human myoblasts. (A) Schematic representation of the RNA-seq data analysis pipeline. (B–D) MA plots for inducible, viral and endogenous DUX4-induced transcriptomes. Genes upregulated by more than 4-fold in red; genes downregulated by more than 4-fold in blue; genes with a significant adjusted P-value (< 0.05) that do not meet 4-fold cutoff for differential expression in green and the genes with adjusted P-value > 0.05 in black. For the vDUX4 sample, genes with log2 fold-change >8 or < -4 are plotted as 'triangles' at the top and bottom edges of the plot, respectively.

relative to ectopic DUX4 expression, particularly for the few genes exhibiting downregulation in DUX4-expressing cells. In contrast, a substantial number of genes were downregulated in the vDUX4 dataset, likely due to suppression of the innate immune response to the lentiviral transduction by DUX4 (described in detail below).

We next tested whether identical sets of genes were expressed in each dataset (irrespective of DUX4 expression). We noted that each dataset contained a set of genes that were not identified as expressed in the other two samples (Supplementary Material, Table S2). (We defined expressed genes as those with a mean expression of at least one transcript per million (TPM), the standard filter recommended by the edgeR package for RNA-seq data analysis [63]). A Venn diagram of the overlap of detected genes among the three datasets (Fig. 5A) showed that the number of dataset-specific genes was particularly large for enDUX4 ($n = 535$). In comparison, vDUX4 had fewer data-set-specific genes ($n = 304$), while iDUX4 elicited even less data-set-specific gene expression ($n = 111$). There are two possible reasons for a gene to be identified as differentially expressed in a given sample: a gene might be robustly expressed only in one dataset, or a gene might be expressed near the detection threshold in all datasets and therefore not reproducibly detectable across all samples. An MA plot highlighting only the dataset-specific genes present in the iDUX4, vDUX4 and enDUX4 datasets showed that most, but not all, such genes were expressed at very low levels in all datasets (Fig. 5B–D). Therefore, to eliminate the noise generated by genes expressed at low levels (based on these MA plots), we limited our analysis to genes expressed above a minimal threshold, defined as an average of 8 TPM (log2 TPM of 3; dotted lines in Fig. 5B–D).

We next sought to understand the biological functions of the minorities of genes that were uniquely expressed in particular datasets. We identified 19 and 123 such genes, respectively, in the vDUX4 and enDUX4 datasets. Gene Ontology (GO) analysis of the 19 genes uniquely expressed in vDUX4 showed that they were mostly genes involved in the viral immune response, consistent with the viral mode of transgene delivery used for vDUX4 (Fig. 5E). The genes unique to enDUX4, on the other hand, were preferentially involved in muscle differentiation or muscle function (Fig. 5F). This enrichment for myogenic processes is likely due to the fact that the enDUX4 dataset assayed differentiated myocytes, whereas the iDUX4 and vDUX4 datasets were derived from undifferentiated, replicating myoblasts.

Of the 123 genes uniquely and robustly detected in the enDUX4 dataset, 15 were upregulated by enDUX4 (log2 fold-change > 2) Fig. 5B–D), indicating that they might be DUX4 targets activated only in the context of myogenesis. If so, then despite their low number, such

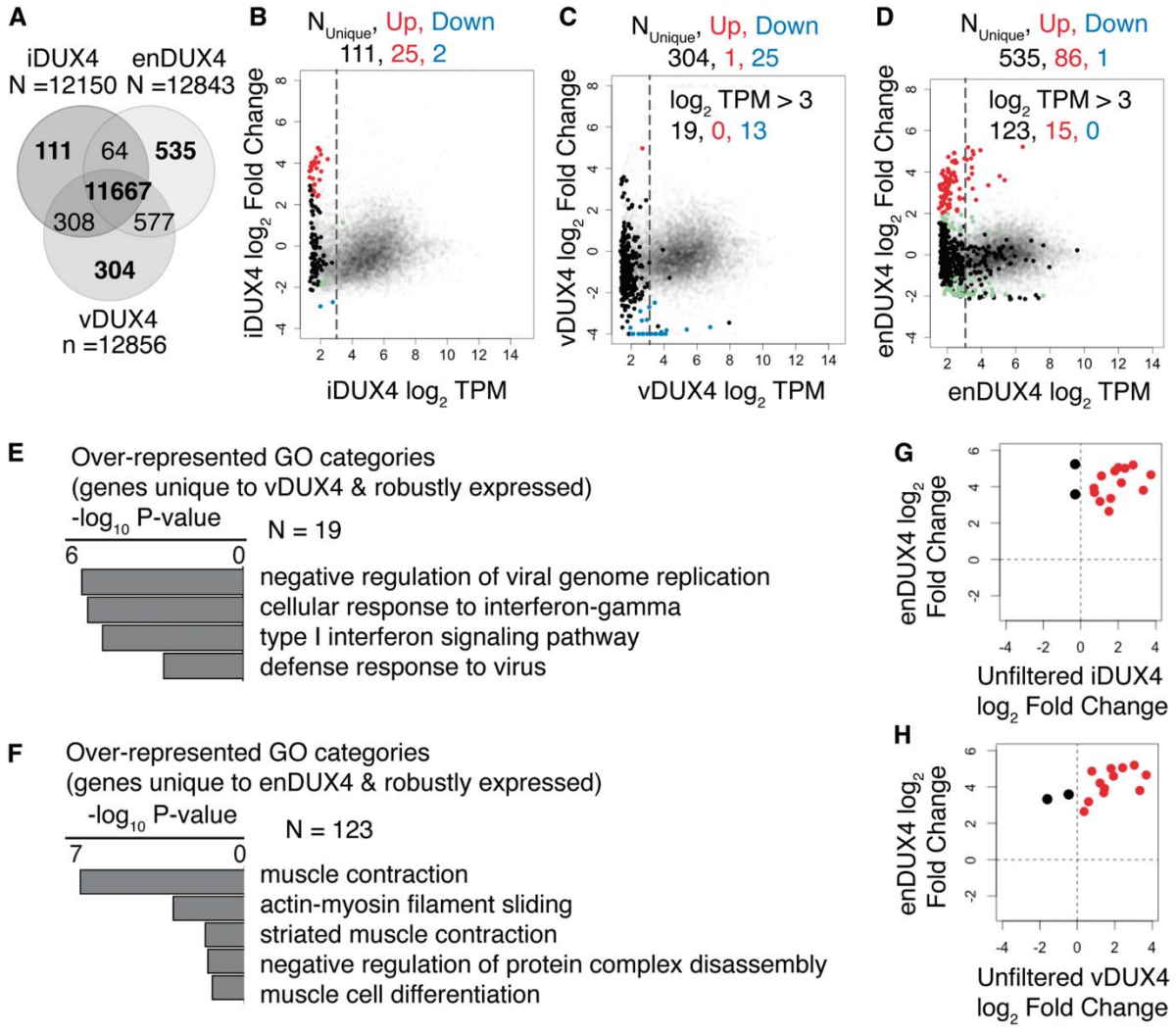


Figure 5 Genes unique to endogenous DUX4 expression are most relevant to muscle differentiation. (A) Venn diagram showing the overlap between the detected genes in iDUX4, enDUX4 and vDUX4 samples. (B–D) MA plot for inducible, viral and endogenous DUX4-induced transcriptome, highlighting the genes that were uniquely detected in each of the samples. Dotted line represents log2 TPM of 3. Color scheme is same as that of Figure 3B–D. (E–F) Gene Ontology (GO) analysis for the genes unique to vDUX4 and enDUX4,

respectively, and expressed at a level above 8 TPM (log2 TPM of 3). (G-H) Scatter plot of log2 fold-change of genes robustly and uniquely upregulated by enDUX4 versus log2 fold-change for the corresponding genes by iDUX4 (G) and vDUX4 (H) calculated without applying a filter for low expressing genes. Black dots represent the discordant genes.

genes could be very relevant to the FSHD disease process since DUX4 is presumed to be expressed during myogenic differentiation in vivo. Hence, we asked how iDUX4 and vDUX4 affected the expression of these 15 genes, which are poorly expressed in myoblasts and so did not pass the filter for robust expression. We repeated the edgeR analysis for these samples without applying a filter for low expression. We found that 12 of these 15 genes were activated by iDUX4 and vDUX4 (Fig. 5G and H). The three discordant genes—MBD3L4 and FAM151A in iDUX4 and AIRE and FAM151A in vDUX4—showed uncertain RNA-seq read assignment in the enDUX4 samples (data not shown). Accurate read mapping to members of multigene families, such as MBD3L4 and FAM151A, and subsequently quantifying gene expression is a known challenge in bioinformatics [64]. For subsequent analyses, we restricted comparisons to only those genes reliably detected and quantified with an average expression greater than 8 TPM in one of the samples being compared, as appropriate.

In summary, while some genes do exhibit dataset-specific expression patterns, our analyses indicate that these differences are primarily due to the distinct basal transcriptomes of myoblasts (iDUX4) versus differentiated myocytes (enDUX4), as well as gene expression changes in response to viral transduction of myoblasts (vDUX4).

2.3.4 *Endogenous and exogenous DUX4 produce highly similar transcriptional responses*

We next measured the similarity in differential gene expression caused by expression of iDUX4, vDUX4, and enDUX4. Venn diagrams are commonly used to assess similarities and differences between datasets and to represent the overlap between lists of significantly up or downregulated genes. For this purpose, a log2 fold-change cutoff of two is often chosen [57].

However, it is clear from the MA plots (Fig. 4B–D) that DUX4 activates genes to different extents in the different datasets, such that a particular gene might be induced two-fold in one dataset but three-fold in another. Therefore, using a single, arbitrarily defined fold-change cutoff to select up- or downregulated genes might make the datasets seem artificially dissimilar. We therefore used a more statistically rigorous approach to further explore the differences between the DUX4 expression systems. We measured the percentage overlap between the sets of significantly up- or downregulated genes across a wide range of fold-change cutoffs (Fig. 6A–D). For instance, of all of the genes that were upregulated by more than four-fold (adjusted P-value < 0.05) by vDUX4, 75% (310 genes) and 51% (215 genes) were also upregulated by more than four-fold in iDUX4 and enDUX4, respectively (Fig. 6A–B). As the (arbitrarily defined) fold-change cutoff is relaxed for the comparator group to two-fold, the sets of upregulated genes become increasingly concordant (97% and 81% overlap). This dependence on the fold-change threshold is even more striking for the sets of downregulated genes. There is poor overlap at a cutoff of four-fold, but reasonable concordance at a comparator group cutoff of two-fold (Fig. 6C and D). Moreover, pairwise scatter plots of the fold-changes of the three datasets showed that transcriptional changes caused by DUX4 were indeed highly similar (Pearson's correlation coefficient, $r \sim 0.75–0.85$; P-values < 2.2e-16), despite major differences in cell type (myoblasts versus differentiated muscle cells), timing, expression levels, culture conditions, and the mechanism of expressing the DUX4 protein (Fig. 6E–G). Together, these statistical analyses indicate that DUX4-responsive genes are consistently induced or repressed across the three datasets, although the magnitude of induction or repression may differ.

2.3.5 *Biological differences between DUX4 expression systems arise from the distinct cellular contexts used*

Our comparisons of DUX4-regulated genes revealed highly similar patterns of induction and repression across all datasets, but there were small subsets of genes that responded to

DUX4 expression only in particular datasets. In order to understand the origins of these differences, it is important to consider the different cellular contexts in which DUX4 was expressed in each dataset.

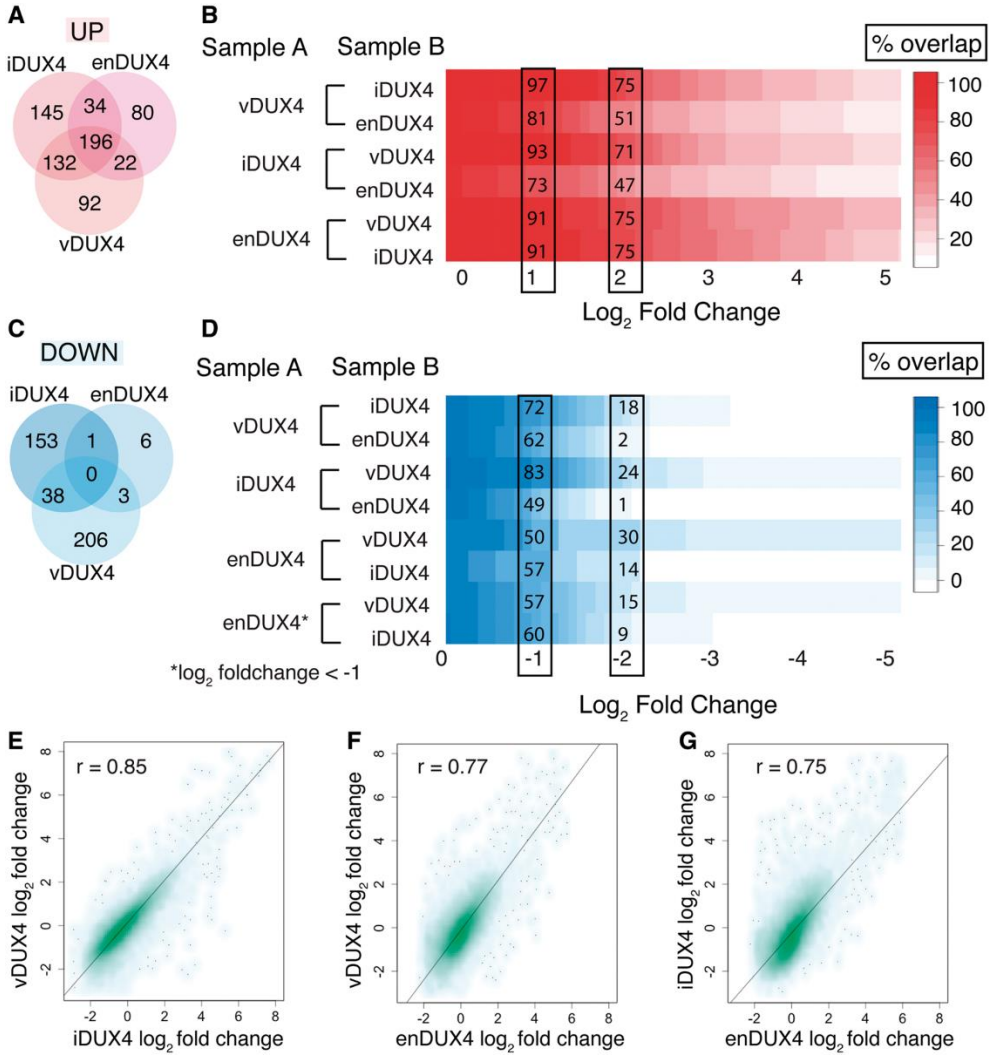


Figure 6 Regulated gene sets show significant overlap between samples. (A, C) Venn diagram of upregulated genes showing the overlap between genes with > 2 log2 fold-change (A) or < -2 log2 fold-change (B) and an adjusted P-value < 0.05 in the three datasets. (B) Percent overlap plot shows the overlap of gene sets that are > 2 log2 fold upregulated in Sample A with a significant adjusted P value over a sliding scale of 0 to 5 log2 fold upregulation in Sample B. (D) Percent overlap plot shows the overlap of gene sets that are < -

2 log2 fold downregulated in Sample A with a significant adjusted P value over a sliding scale of 0 to -5 log2 fold downregulation in Sample B. (E–G) Scatter plot of log2 fold-change of quantifiable genes in inducible versus viral DUX4 expression (E), endogenous versus viral DUX4 expression (F), and endogenous versus inducible DUX4 expression (G). r - Pearson's correlation coefficient.

For example, in the vDUX4 dataset, the lentiviral delivery of DUX4 induced an antiviral innate immune response. In the enDUX4 dataset, the muscle cells were differentiated in the presence of the calcium chelator EGTA to prevent fusion and facilitate FAC sorting, whereas vDUX4 and iDUX4 were expressed in undifferentiated myoblasts. When we plotted normalized log2 fold-changes of all of the genes expressed in the iDUX4 dataset versus the vDUX4 or enDUX4 datasets, we identified a small subset of genes that were comparatively over- or under-expressed in each sample (Fig. 7A and B). For the genes that were comparatively under-induced in the vDUX4 dataset compared to the iDUX4 dataset, the top five enriched GO categories all corresponded to the cellular defense response (Fig. 7C). This enrichment is consistent with our expectation that a lentiviral vector will induce an immune response, as well as the fact that lentiviral GFP induces a stronger response relative to lentiviral DUX4, as DUX4 suppresses the innate immune response [15]. Genes that were comparatively more induced in the iDUX4 dataset compared to the enDUX4 dataset were enriched for processes including cell development and differentiation (Fig. 7D), suggesting that the degree of fold-change might be affected by the state of muscle differentiation.

To confirm that myogenic differentiation is the major contributor to the small differences between the DUX4-induced genes in the iDUX4 and enDUX4 datasets, we induced codon-altered DUX4 with doxycycline (DOX) in the MB135 myoblasts in growth media (GM) and the same cells differentiated into myotubes in differentiation media (DM). RT-qPCR analysis of a set of genes that were relatively repressed in the enDUX4 dataset compared to the iDUX4 dataset showed a similar trend of relative repression in differentiated iDUX4 myotubes compared to

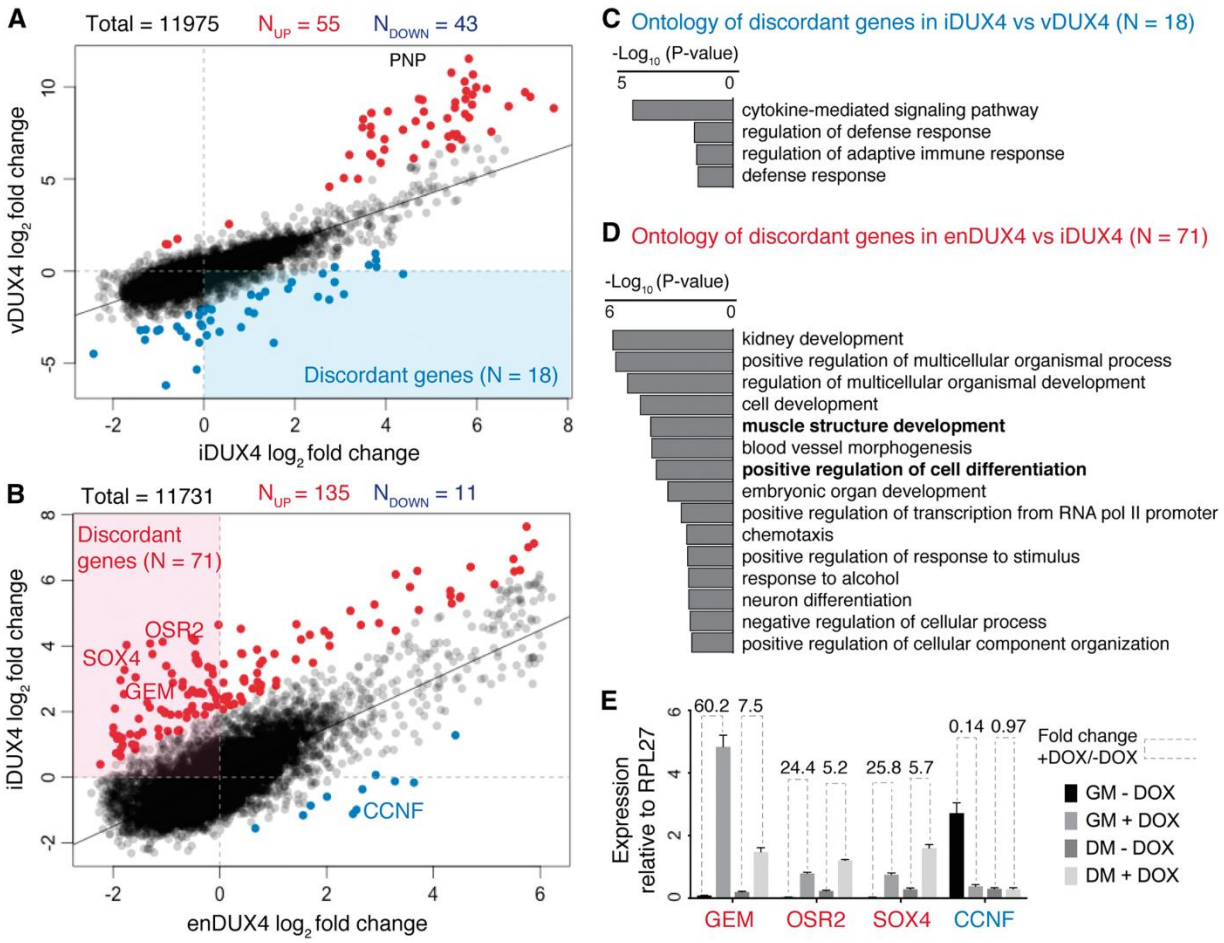


Figure 7 Differentially regulated genes appear most relevant to the gene expression programs underway during DUX4 expression. (A) Scatter plot of scaled and centered log2 fold-change values of iDUX4 and vDUX4. Line represents a linear model and the genes marked in red are more over-expressed in vDUX4 compared to iDUX4 (residual > 2) and those in blue are more under-expressed in vDUX4 compared to iDUX4 (residual < -2). (B) Scatter plot of scaled and centered log2 fold-change values of enDUX4 and iDUX4. Line represents a linear model and the genes marked in red are more over-expressed in iDUX4 compared to enDUX4 (residual > 2) and those in blue are more under-expressed in iDUX4 compared to enDUX4 (residual < -2). (C–D) GO category analysis of significantly under-expressed genes in vDUX4 compared to iDUX4 (C) and significantly over-expressed genes in iDUX4 compared to enDUX4 (D). (E) qPCR data for a few candidate discordant genes in control and iDUX4 cells in growth media (GM) versus differentiation media (DM).

undifferentiated iDUX4 myoblasts (Fig. 7E). Together, these results suggest that the differences in gene expression between the enDUX4 and vDUX4 or iDUX4 datasets were primarily due to differences in baseline gene expression in undifferentiated versus differentiated cells, rather than differences in the intrinsic activities of endogenous or exogenous DUX4.

2.3.6 *The expression pattern common across the different modes of DUX4 expression recapitulates context-independent functions of DUX4*

Having explored the differences between the endogenous and exogenous DUX4 expression systems, we next sought to characterize the gene expression program that is common to the different systems. To this end, we performed Gene Ontology analyses of the genes that were induced or repressed by more than two-fold by iDUX4, vDUX4 and enDUX4 (Fig. 8A and Supplementary Material, Table S4). The genes that were upregulated by DUX4 in all three expression systems were enriched for proteins involved in transcription, RNA processing, splicing and transport (Fig. 8B), as has been observed before [15,57]. The downregulated genes were involved in viral defense, cell proliferation, and apoptosis, among other gene classes (Fig. 8C). Next, we asked if the FSHD biomarkers previously identified via transcriptome analysis of FSHD muscle biopsies [61] showed activation across all three datasets. Out of the 67 biomarker genes identified by Yao et al. [61], 47 were annotated and/or detectably expressed in the iDUX4, vDUX4 and enDUX4 datasets and all showed high upregulation by both exogenous and endogenous DUX4 as shown by the MA plot (Fig. 8D–F). In conclusion, the core activity of DUX4 is highly similar across different modes of DUX4 expression and the three model systems compared in this study each recapitulate key transcriptome changes that can be found in FSHD muscle.

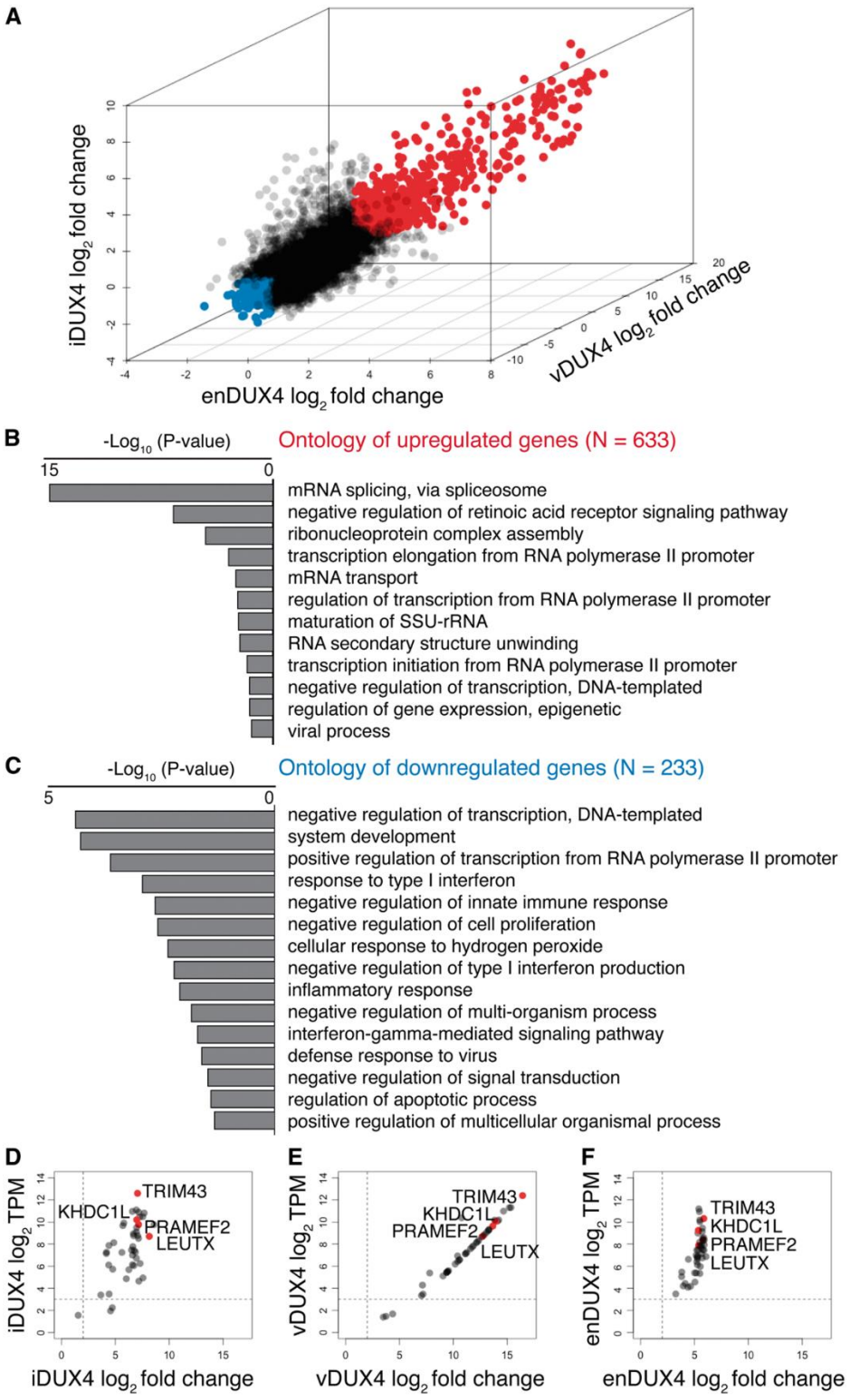


Figure 8 Gene sets common to endogenous and exogenous DUX4 expression highlight the core functions of DUX4. (A) 3D scatter plot for the three datasets (iDUX4, enDUX4 and

vDUX4) highlighting the genes upregulated by more than 2 log2 fold-change in all samples in red and those downregulated by more than 2 log2 fold-change in all samples in blue. (B) GO analysis of the upregulated genes (marked ‘red' in 6A). (C) GO analysis of downregulated genes (marked ‘blue' in 6A). (D–F) MA plot for 47 biomarkers identified by Yao et al. [61] for the iDUX4 (D), vDUX4 (E) and enDUX4 (F) datasets; The four high-confidence biomarkers (LEUTX, PRAMEF2, TRIM43, KHDC1L) are marked in ‘red'. The horizontal dotted line represents TPM of 8; the vertical dotted line represents fold change of 4.

2.4 *Discussion*

It is important to compare different cellular models of DUX4 expression to determine whether particular models are more or less appropriate for studying FSHD biology. In this study, we found that quite different models of DUX4 expression yielded very similar patterns of DUX4-induced transcriptional changes. While there were differences between systems, such differences were largely explained by the differentiation state of the cells or culture conditions used, which gave rise to distinct patterns of basal gene expression. Overall, the high degree of overlap between DUX4-regulated genes identified across distinct cellular contexts strongly indicates that all three models recapitulate important aspects of FSHD biology.

The current study was motivated, in part, by a prior publication that identified differences in gene expression among different models of FSHD [57]. While single fold-change cutoffs are useful for defining gene sets of interest within a dataset, they are not statistically robust when comparing multiple different datasets, as arbitrary cutoffs generate “edge effects”. This statistical effect might explain why this previous study identified relatively little overlap between the differentially expressed genes among different models of FSHD. Correlation analyses conducted here provide strong assurance that studies using different model systems of FSHD can be compared with confidence, provided that the biological state of the cells or variables induced by the experimental design are taken into account.

Consistent with the above conclusions, it is further reassuring that our study, using a newly developed doxycycline-inducible codon-altered DUX4, identified the same DUX4-induced gene expression changes and RNA processing abnormalities reported by prior studies [15,16,42]. Our earlier attempts to create an inducible DUX4 were hampered by the difficulty of establishing a stable integrant that was efficiently induced, which was true for multiple different viral preparations and transduced cell types. Decreasing the CG content and the number of CpG dinucleotides substantially increased the recovery of clones with an inducible DUX4, suggesting that the high CG content of the DUX4 coding sequence might trigger silencing, however, the precise mechanism for the more efficient induction of the stably integrated CA-DUX4 needs further study. The similarity of the genes regulated by the codon-altered DUX4 to genes identified in studies using the wild-type DUX4 indicates that the gene expression changes are due to the DUX4 protein and not the RNA, consistent with our prior studies [65]. Finally, we expect that the human myoblast cell line expressing doxycycline-inducible DUX4 described here will be a useful tool for FSHD research. This cell line allows the isolation of a large number of clonal cells with synchronized induction of DUX4 expression, which we expect will improve reproducibility and allow for accurate temporal dissection of the molecular events following DUX4 expression.

2.5 *Materials and Methods*

Accession codes

The raw sequence reads for the enDUX4 expression experiments were downloaded from the NCBI sequence read archive (SRA) database under accession number SRP058319 (9). Data generated in this study are available through the NCBI SRA database under accession number GSE85461.

Cell culture

Proliferating human myoblasts were cultured in F10 medium (Gibco/Life Technologies) supplemented with 20% fetal bovine serum (Thermo Scientific), 10ng bFGF (Life Technologies), 1µM dexamethasone (Sigma) and 50U/50µg penicillin/streptomycin (Life Technologies). Differentiation into myotubes was initiated by switching the fully confluent cell monolayer into a low-serum media such as DMEM with 1% horse serum (Life Technologies), supplemented with 10 µg/ml each of insulin and transferrin for 48 hours. To induce DUX4 expression in differentiated iDUX4 cells, doxycycline was added in the last 14 hours of differentiation. 293T cells were cultured in DMEM (Gibco/Life Technologies) supplemented with 10% fetal bovine serum and 50U/50µg penicillin/streptomycin as above.

Codon-altered and wild-type inducible constructs

Wild-type DUX4 was subcloned into the pCW57.1 vector, a gift from David Root (Addgene plasmid #41393) by restriction enzyme digest, using the NheI and SalI sites of the pCW57.1 vector. The codon-altered DUX4, which has ~73% identity to wild-type DUX4, was synthesized by IDT custom gene synthesis and subcloned into pCW57.1 such that the only discrepancies between the codon-altered and wild-type constructs are within the coding region itself. Sequence is in Supplementary Material, Fig. S2.

Generation of clonal cell lines expressing DUX4

Lentivirus with the inducible wild-type and codon-altered DUX4 transgenes were generated by transfection of the appropriate pCW57.1 vector into 293T cells, along with the packaging and envelope plasmids pMD2.G and psPAX2 using lipofectamine 2000 reagent (ThermoFisher). To generate clonal lines, control human myoblasts, MB135, immortalized with hTERT and CDK4, were plated at low density and transduced with lentivirus at a low multiplicity of infection (MOI < 1) in the presence of polybrene. Cells were selected and maintained in puromycin. After the selection was complete, remaining cells were allowed to grow and form colonies. Individual

clones that were well isolated were picked using cloning cylinders, about 10 days after transfection, and expanded in the presence of puromycin. Five individual codon-altered DUX4 clones and 15 wild-type DUX4 clones were picked and used in this study.

RNA extraction and RT-qPCR

Total RNA was extracted from whole cells using either TRIzol reagent (Ambion) or NucleoSpin RNA kit (Macherey-Nagel) following the manufacturer's instructions. Purified RNA was Dnase I treated (ThermoFisher) and heat inactivated prior to cDNA synthesis. First strand cDNA synthesis was performed using SuperScript III reverse transcriptase (ThermoFisher) and following the manufacturer's instructions. Samples were split into duplicate reactions where one underwent a mock, no enzyme treatment as a control. Quantitative PCR was carried out on cDNA using the standard curve method and SYBR green as the detector. The primers used in this study are listed below:

GEM_1F: GAAAAGAACCCCTGGAACGTG

GEM_1R: TGTACTGGTGGGGCTCTTTC

OSR2_1F: TGCCCAGGTTGACCTTTCTG

OSR2_1R: CTGAGGGGACCAACCCTTTC

SOX4_1F: ATCGCTGTTTGGATTTCCTG

SOX4_1R: ACACTGGTGGCAGGTTAAGG

CCNF_1F: GACCATCTTGAGTCTCCCCG

CCNF_1R: AAGAGCTTCAGGTTCCCTGG

RPL27-1L: GCAAGAAGAAGATCGCCAAG

RPL27-1R: TCCAAGGGGATATCCACAGA

ZSCAN4_F: FTGGAAATCAAGTGGCAAAAA

ZSCAN4_R: RCTGCATGTGGACGTGGAC

KHDC1L_F: CACCAATGGCAAAGCAGTGG

KHDC1L_R: TCAGTCTCCGGTGTACGGTG

Protein extraction and immunoblotting

Cells were directly lysed in 2X gel loading buffer with 4% BME, sonicated, and boiled for 10 minutes. Samples were run on a 4–12% polyacrylamide gel and transferred to a PVDF membrane. Membranes were blocked in 5% milk for one hour before overnight incubation with primary antibody at 4°C. Membranes were incubated with secondary antibody for one hour, and the chemiluminescent signal was detected on film. The antibodies used in this study are rabbit anti-DUX4 (E14-3) and rabbit anti-H3 (Abcam; ab1791).

RNA-seq library preparation and sequencing

The RNA-seq libraries were prepared with polyA-selected RNA (starting with 1 µg of total RNA) using TruSeq RNA Sample Prep Kit (Illumina) either manually (for the pilot dataset in Fig. 3) or using a Perkin Elmer Sciclone NGSx Automated Library Prep Workstation (for the triplicate dataset in Fig. 4). Library size distributions were validated using an Agilent 2100 Bioanalyzer (Agilent Technologies). Indexed libraries were quantified using Qubit® 2.0 Fluorometer (Life Technologies) and pooled for optimal clustering. Sequencing on an Illumina HiSeq 2500 was carried out by the FHCRC Genomics Shared Resource to generate 50 bp single-end reads for the pilot singleton iDUX4 dataset (Fig. 3) and 100 bp single-end reads for the iDUX4 and vDUX4 datasets with replicates (Fig. 4). Image analysis and base-calling were performed using Real time Analysis (RTA) version v1.18 (Illumina), followed by demultiplexing of indexed reads to generate FASTQ files using bcl2fastq Conversion Software v1.8.4 (Illumina).

RNA-seq data analysis

The general pipeline for RNA-seq data analysis is presented in Fig. 4A. Briefly, RNA-seq reads were mapped to the UCSC hg19 (NCBI GRCh37) human genome assembly using bowtie [66], RSEM [67] and TopHat [68], as described by Dvinge et al. [69]. Two mismatches were allowed for 50bp reads and three mismatches allowed for 100bp reads. Differential gene expression analysis was performed using the edgeR package [63] as follows: RNA transcript levels normalized using trimmed mean of M value (TMM; [70]) were input to the edgeR program in transcripts per million (TPM). DUX4-expressing samples were compared to the corresponding control samples in biological triplicates (iDUX4), duplicates (vDUX4) or sextuplicates (enDUX4). Only genes expressed with at least 1 TPM in 50% of the samples in a dataset (i.e. in either the controls or the DUX4 samples) were considered for differential expression to avoid noise from poorly expressed genes. edgeR output of log2 fold-change, average TPM and false discovery rate (corrected for multiple hypothesis testing by Benjamini-Hochberg approach) were used in all subsequent analysis (processed data is provided in Supplementary Material, Table S1). All plots were generated using R plotting functions and/or the ggplot2 package [71]. All statistical tests were also performed using R functions.

GO category analysis

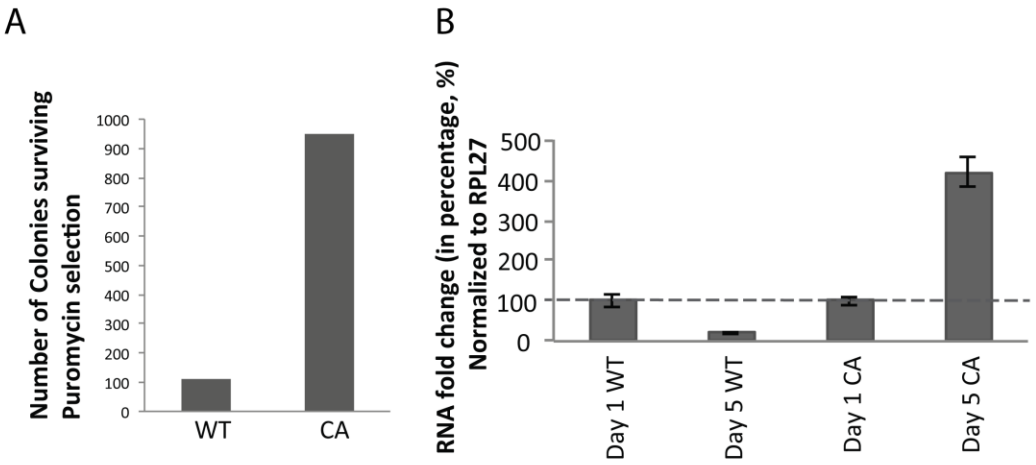
GO category analysis was conducted using the PANTHER classification system (http://pantherdb.org/geneListAnalysis.do; [72]) using the statistical overrepresentation test against all human genes, using the complete GO biological process annotation. P-values were corrected for multiple hypothesis testing using the Bonferroni correction.

Supplementary Material

Supplementary Material and Supplementary Tables for this chapter are available at HMG online.

2.6 *Supplemental Figures*

Supplementary Figure S1



Supplementary Figure S1. Silencing of WT-DUX4 compared to CA-DUX4.
A) Colony counts from HEK293T transduced with WT-DUX4 and CA-DUX4 lentiviral particles and selected for puromycin resistance. The lentiviral production core determines titer by quantitative PCR of integrated viral transduced genes after passage in the absence of selection. Both WT-DUX4 and CA-DUX4 lentivirus were prepared and titered in the same batch and had nearly identical titers. Based on this, equivalent infectious units for each viral prep were used, suggesting that the puromycin resistance gene was not as efficiently expressed from the WT-DUX4 compared to the CA-DUX4. MB135 myoblasts also showed fewer colonies with WT-DUX4 compared to CA-DUX4 following transduction and selection (data not shown). B) Inducible expression of WT-DUX4 decreases over time compared to CA-DUX4. Immortalized MB135 myoblasts were transduced with lentivirus encoding the doxycycline inducible WT-DUX4 or CA-DUX4 and then induced with doxycycline one day or five days after transduction. qRT-PCR analysis of a DUX4 transcriptional target, KHDC1L, was performed on RNA isolated six hours after transduction to measure induced DUX4 activity because measuring DUX4 directly would be contaminated by viral RNA containing the DUX4 gene. The induced fold-change of the KHDC1L target for day one induction was set at 100% for the WT-DUX4 and the CA-DUX4.

Chapter 3. *A genetic screen uncovers the double stranded RNA response and MYC-mediated apoptosis as potential pathways involved in DUX4 toxicity*

A version of this chapter has been previously published as:

Shadle SC, Zhong JW, Campbell AE, Conerly ML, Jagannathan S, Wong CJ, Morello TD, van der Maarel SM, Tapscott SJ. DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. PLoS genetics. 2017 Mar 8;13(3):e1006658.

3.1 *Abstract*

Facioscapulohumeral dystrophy (FSHD) is caused by the mis-expression of DUX4 in skeletal muscle cells. DUX4 is a transcription factor that activates genes normally associated with stem cell biology and its mis-expression in FSHD cells results in apoptosis. To identify genes and pathways necessary for DUX4-mediated apoptosis, we performed an siRNA screen in an RD rhabdomyosarcoma cell line with an inducible DUX4 transgene. Our screen identified components of the MYC-mediated apoptotic pathway and the double-stranded RNA (dsRNA) innate immune response pathway as mediators of DUX4-induced apoptosis. Further investigation revealed that DUX4 expression led to increased MYC mRNA, accumulation of nuclear dsRNA foci, and activation of the dsRNA response pathway in both RD cells and human myoblasts. Nuclear dsRNA foci were associated with aggregation of the exon junction complex component EIF4A3. The elevation of MYC mRNA, dsRNA accumulation, and EIF4A3 nuclear aggregates in FSHD muscle cells suggest that these processes might contribute to FSHD pathophysiology.

3.2 *Introduction*

Facioscapulohumeral dystrophy (FSHD) is a progressive muscular dystrophy caused by mis-expression of the double-homeobox transcription factor DUX4 in skeletal muscle [73]. Normally, DUX4 is not expressed in skeletal muscle nor in most somatic tissues examined [14,74]. Ectopic expression of DUX4 in human and mouse cell lines as well as in vivo injection of DUX4 adenovirus into mouse muscle leads to rapid cellular apoptosis [53,55]. This cell death is dependent on the transcriptional activity of DUX4 because expression of DUX4 with mutations in the DNA binding domain or trans-activation domain do not exhibit toxicity [36,55]. More recently, it was demonstrated that endogenous levels of DUX4 produced in FSHD muscle cells similarly causes cellular death [57].

Apoptosis is known to be a critical cellular process for both tissue homeostasis as well as vertebrate ontogeny where cells in developing organs follow the general guidelines of “proliferation, differentiation and demolition” [75]. It is also appreciated that cellular apoptosis, outside of normal homeostatic or developmental contexts, is involved in autoimmune and neurological diseases [76,77]. As evidenced by numerous mouse knockout lines, programmed cell death is required for productive sperm development where excess or abnormal germ cells are constantly culled to ensure adequate space and nutrients [78]. Previously, immunodetection has identified DUX4 expression in cells in the seminiferous tubule, morphologically resembling spermatogonia or primary spermatocytes [14] and in the thymus [74], both tissues with high rates of apoptosis. Thus, it is possible that expression of DUX4 in skeletal muscle inappropriately activates a program of apoptosis that might otherwise be a 'normal' consequence of DUX4 expression during developmental processes.

As a transcription factor, DUX4 activates many genes that are expressed in stem cells and in the germline [15]. The long terminal repeat (LTR) of a subset of human endogenous retroviruses (ERVs) contain the DUX4 binding motif, and DUX4 binds and activates their transcription, occasionally creating a novel transcription start site for adjacent genes [16]. DUX4 expression also represses the innate immune response [15] and the nonsense mediated decay (NMD) pathway [42], leading to an accumulation of normally degraded RNAs. However, it is not currently understood whether the changes in RNA stabilization following DUX4 expression lead to apoptosis. It has been demonstrated that FSHD muscle biopsies exhibit oxidative stress and mitochondrial dysfunction compared to control biopsies [79] and small molecule screens have demonstrated that compounds which protect cells from oxidative damage also protect against DUX4 toxicity [36,37]. It was also previously shown that Tp53 knockout mice were protected from the effects of DUX4 delivered to skeletal muscle by AAV transduction and that a P53 inhibitor decreased DUX4 toxicity in human HEK293 cells [55]. However, the DUX4-induced apoptotic pathways relevant to human skeletal muscle and FSHD remain poorly understood.

We conducted a small interfering RNA (siRNA) screen to identify genes and pathways necessary for DUX4 toxicity. This screen identified the MYC-mediated apoptotic pathway and components of the double-stranded RNA (dsRNA) innate immune response as necessary for DUX4-induced apoptosis. We found that DUX4 expression resulted in the stabilization of several mRNAs, including MYC, and the accumulation of nuclear dsRNAs. Stabilization of MYC mRNA was associated with a dramatic increase in MYC protein levels and the activation of genes in the MYC-mediated apoptosis pathway, including BCL2L11 and EGR1; whereas the accumulation of dsRNA was associated with nuclear aggregation of EIF4A3 and phosphorylation of the kinase EIF2AK2/PKR and its downstream target, the eukaryotic translational initiation factor subunit EIF2S1/eIF-2α. The similar elevation of MYC mRNA, dsRNA accumulation, and EIF4A3 nuclear foci in FSHD muscle cells suggest that these processes might contribute to FSHD pathophysiology.

3.3 *Results*

3.3.1 *siRNA screen identifies candidate genes necessary for DUX4-induced cell death*

Although a prior study had identified TP53 as necessary for DUX4-induced cell death in some cells [55], we found that human myoblasts with CRISPR mutated TP53 and the RD rhabdomyosarcoma cell line that does not contain a functional TP53 allele [80,81] both succumbed to DUX4-induced cell death as efficiently as primary human myoblasts (Fig S2A–S2D). To establish a screen for genes necessary for DUX4-induced apoptosis, we transduced the RD cell line with a lentiviral vector encoding a doxycycline-inducible DUX4 coding sequence and puromycin selectable marker (see Fig S3A for schematic) and generated a clonal rhabdomyosarcoma cell line (RD-DUX4i) with robust doxycycline-inducible expression of DUX4 that showed cellular toxicity at 24 hours after DUX4 induction (Fig S3B) and more than 95% cell death by 48 hours, as determined by the ATP-based CellTiter-Glo assay (Fig S3C). The dying

cells exhibited an increase in activated caspase 3/7 (Fig S2A) indicating that DUX4 induction leads to apoptosis in the P53 deficient RD cells.

To identify genes and pathways necessary for DUX4-induced toxicity in RD cells we measured cell survival following transfection of an siRNA library targeting 6,961 genes in the human “druggable” genome with a pool of four siRNAs per target gene. The parameters for the screen are summarized in Fig S3D and were optimized as in Fig S4A. As positive controls, we used two unique siRNAs targeting DUX4 which had different efficacies of knockdown (Fig S4B) and, consequently, different effects on cell viability after DUX4 induction (Fig S3E). Transfections were performed in triplicate and cell survival following DUX4 induction was measured by CellTiter-Glo.

Using a stringent mean Z-score threshold of 3.0, 69 siRNA pools significantly increased cell viability in response to DUX4. Targets that exceeded this threshold included some genes previously implicated in cellular apoptosis, for example: Death effector domain containing 2 (DEDD2), Cell death-inducing DFFA-like effector a (CIDEA), MutS homolog 2 (MSH2), C-MYC (MYC) and its dimerization partner MAX, and RNASEL, a mediator of cellular antiviral defense. Using a more traditional Z-score cutoff of 2.0, 353 siRNA pools enhanced cell viability after DUX4 induction. In addition, 30 siRNAs had a Z-score of less than -2.0, suggesting that these genes might protect cells from DUX4-induced death. Among these targets were the DUX4 target genes, TRIM51 (also known as SPRYD5) and TRIM43 as well as the anti-apoptotic BCL2L1 (also known as BCL-X) that can suppress MYC-mediated cell death [82]. As expected, siRNAs targeting TP53 had no effect on cell viabilty (Z-score = -0.03). Network analysis placed MYC/MAX as a 'hub' of protein-protein interactions among the candidate genes with an absolute value Z-score cutoff of 2.0 (Fig S4C). The general results of the screen are summarized in Fig S3E, and the full ranked list is available in S1 Table.

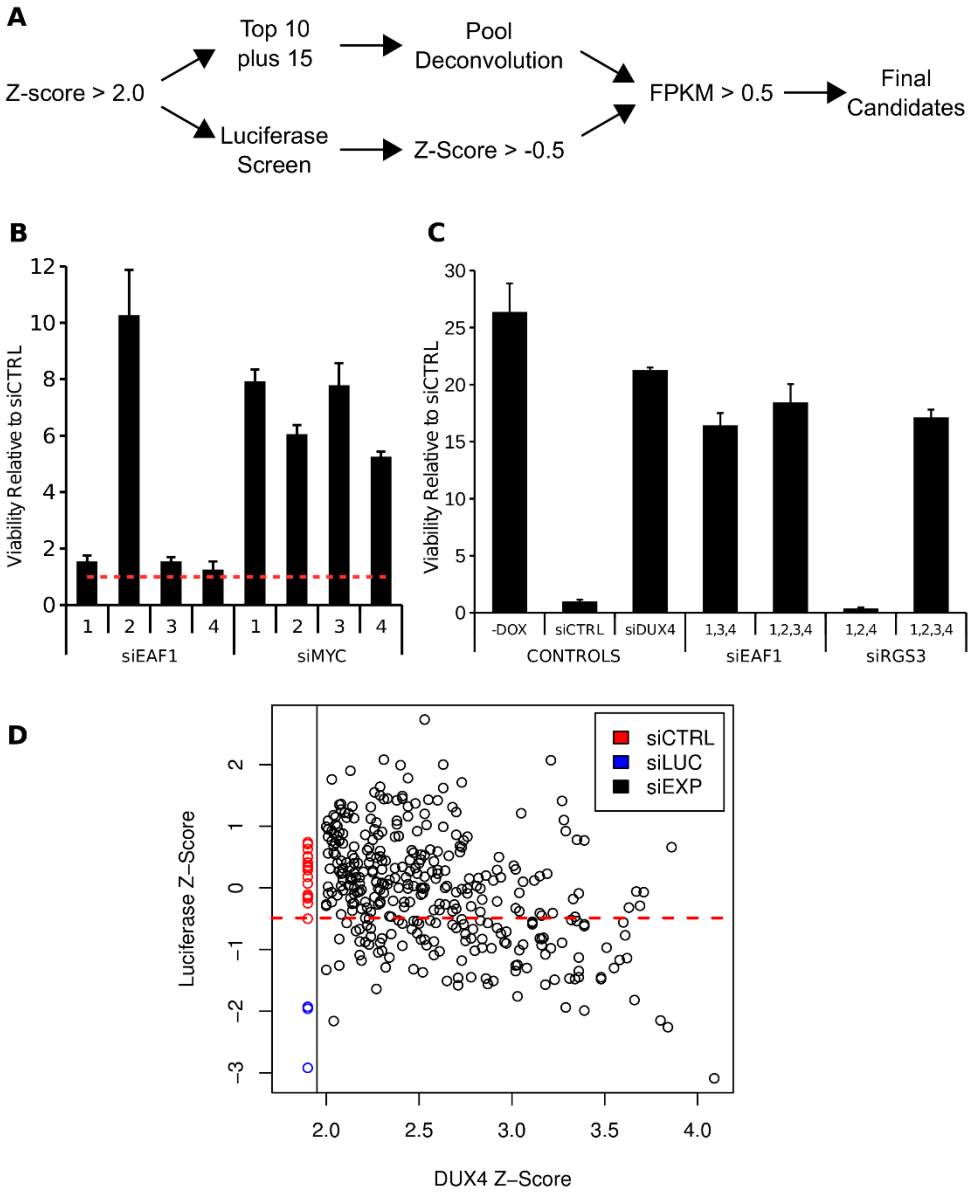


Figure 9 Elimination of siRNA pools with possible off-target activities. (A) Schematic of secondary screening used to filter out unwanted targets from the primary siRNA screen. (B) Example of how siRNA pools from the primary screen were deconvoluted to determine whether the effect on cell viability was dominated by a single siRNA which would suggest off-targeting. In this example, EAF1 has a single dominant siRNA that rescues, whereas multiple siRNAs to MYC rescue. These data depict cell viability using the CellTiter-Glo assay. (C) Pooling of poorly performing siRNAs can 'synergize' to rescue from DUX4 toxicity as with siEAF1 whereas others are clearly dominated by a single siRNA as in siRGS3. These data depict cell viability using the CellTiter-Glo assay. (D) Scatter-plot of mean Z-scores from the

original DUX4 siRNA screen against mean Z-scores produced in the Luciferase screen to determine the siRNA pools that inhibited doxycycline induction of the transgene. Error bars for (B) and (C) depict standard deviation of the mean of three replicate wells.

We proceeded to validate candidate siRNA pools identified in the initial screen using a strategy outlined in Fig 9A. As a first validation step we transfected RD-DUX4i cells with the ten siRNA pools with the highest Z-scores plus 15 additional pools with Z-scores ranging from 1.93 to 3.61. All 25 pools increased viability over the non-silencing control siRNA, indicating that pools identified in the primary screen reproducibly enhanced cell survival after induction of DUX4 (Fig S5A).

As a second validation step we de-convoluted these 25 pools of four individual siRNAs to determine whether multiple siRNAs within each pool contributed to the increased survival of RD cells rather than a single siRNA which might indicate a nonspecific, off-target effect. Of the 25 pools retested, 21 had more than one siRNA that enhanced cell survival by 2-fold or greater (Fig 9B and Fig S5B). For the four pools with a single siRNA that rescued viability, we tested whether the individual non-rescuing siRNAs from that pool might rescue if pooled together, reasoning that synergy among the individually non-rescuing siRNAs might indicate on-target activity. Re-transfection of these non-rescuing siRNAs as pools of three revealed that more than one siRNA against two genes (EAF1 and TLR5) rescued from DUX4 toxicity (Fig 9C and Fig S5C), whereas for the other two genes (TLL1 and RGS3), only a single siRNA had rescuing activity.

Further analysis showed that the rescue conferred by siRNAs in pools where there was a single rescuing siRNA, such as siRGS3, was correlated with the inhibition of doxycycline induction of DUX4 and of a doxycyline inducible luciferase transgene in RD cells (RD-LUCi cells; Fig S5D). This indicated that some of the siRNA pools in the original screen likely

achieved RD-DUX4i cell survival and a high Z-score because a single siRNA in the pool inhibited the doxycycline induction of DUX4 through unknown mechanisms.

Therefore, as a third validation step we tested for the inhibition of the doxycycline induction pathway on the entire set of siRNA pools with an original Z-score ≥ 2.0 using the RD-LUCi cells. 110 of our initial 353 pools suppressed luciferase induction based on a cutoff Z-score of ≤ -0.5 (Fig 9D and S2 Table), the minimum score of the 16 scrambled control siRNAs used in this secondary screen. For example, the RGS3 siRNA pool, which we had already determined had a single siRNA that suppressed doxycycline induction of luciferase, had a mean Z-score of -3.09. Overall, there was a modest but significant inverse correlation between the RD-DUX4i screen Z-scores and the RD-LUCi screen Z-scores (Spearman's rho = -0.396), indicating that some of the pools identified in the original screen were likely secondary to inhibiting the doxycycline induction pathway. Therefore, we eliminated the 110 genes targeted by pools that depress luciferase induction by more than the non-silencing control.

As a final filter, we performed RNA-sequencing on RD-DUX4i cells and required that the targeted mRNA must be expressed at greater than an average FPKM (fragments per kilobase of transcript per million mapped reads) of 0.5 to filter out low or non-expressed genes. For example, although TLR5 apparently rescued based on initial validation criteria (see above), it was expressed below the threshold level and was therefore eliminated. Of the original set of 6,961 genes targeted by the library, 16 genes with a Z-score ≥ 3.0 passed our validation and filtering steps (Table 2). The final list of filtered targets with Z-score ≥ 2.0 is summarized in S3 Table.

Table 2 Top 16 filtered targets with Z-score > 3.0 from RD-DUX4i siRNA screen.

Target ID	DUX4i Z-score	LUCi Z-score	AVG FPKM[a]-DOX	AVG FPKM[a] +DOX	Validation % of -DOX[b]	Mean normalized Fold Rescue[c]
FOSB	3.86	0.66	0.16	8.59	54.7	75.17
RNASEL	3.71	-0.07	0.94	0.54	52.4	63.50
MYC	3.69	-0.29	29.87	696.89	63.9	62.60
FXN	3.67	-0.06	3.51	0.77	61.6	61.70
EAF1	3.63	-0.32	5.89	19.27	34.0	58.60
PHF20	3.39	0.77	4.93	1.36	38.3	44.93
NLGN2	3.36	-0.09	8.85	1.77	NA	43.17
TGS1	3.35	0.78	10.11	12.56	NA	43.10
HSD17B4	3.29	0.92	5.57	2.09	NA	39.97
ITPK1	3.28	1.1	1.41	0.44	NA	39.60
SGK196	3.19	0.23	4.64	2.52	NA	36.47
CDC20	3.15	-0.41	79.51	358.73	NA	33.87
QSOX1	3.15	0.19	7.54	4.37	NA	33.87
THOP1	3.06	0.1	11.19	3.34	NA	30.87
FBN2	3.04	-0.08	0.91	0.51	NA	30.37
APH1A	3.02	-0.33	15.96	4.79	NA	30.00

[a]AVG FPKM, mean fragments per kilobase of transcript per million mapped reads across three replicates.

[b]Validation % of -DOX, mean CellTiter-Glo determined viability of the indicated target as a percentage of the no doxycycline control in the validation experiment as in S3A Fig. 'NA' indicates that the given target was not retested. For comparison purposes, the scrambled control siRNA viability was 4.7% of the -DOX control.

[c]Mean Normalized Fold Rescue, the average ratio between raw target fluorescent measurements from the original screen and the sample median for the 384-well plate as a whole. For comparison purposes, the average ratio for the more robust DUX4 siRNA was 80.0.

https://doi.org/10.1371/journal.pgen.1006658.t001

3.3.2 *DUX4 inhibits MYC mRNA degradation and activates a MYC-mediated apoptotic pathway*

The filtered targets meeting a Z-score threshold of 3.0 included several genes that broadly regulate RNA transcription (MYC, FOSB, EAF1, PHF20, CDC20), protein translation (TGS1) [83], or mitochondrial function (FXN) [84]. MAX, the obligate heterodimer of MYC, also appeared as a candidate target in the initial screen, with a Z-score exceeding 3.0, although MAX was subsequently eliminated from the final candidate list because the pool of siRNAs to MAX demonstrated a modest reduction in doxycycline induction of the luciferase transgene (Z-score = -0.93). It is interesting to note that many of the aforementioned genes were upregulated in our RNA-seq data in RD cells overexpressing DUX4: FOSB (~92 fold), MYC (~56 fold), EAF1 (~8 fold) and CDC20 (~11 fold). Although siRNAs against MYC had a modest effect on transgene expression (luciferase Z-score = -0.29), the siRNAs also rescued viability following lentiviral transduction of a constitutively expressed DUX4 in cells with DUX4 protein levels and

nuclear localization equivalent to the controls (Fig S5E–S5G). Therefore, we decided to investigate the MYC-mediated apoptosis pathway because DUX4 expression resulted in a dramatically increased expression of MYC as well as a set of genes that facilitate a program of enhanced cellular growth or metabolism.

We first verified that MYC protein was upregulated following DUX4 expression. Western blotting and RT-qPCR confirmed that MYC protein and RNA levels increased dramatically following DUX4 expression (Fig 10A, Fig S6A-B), tracking closely with a direct target of the DUX4 transcription factor, MBD3L2 (Fig 10A). The increase in MYC did not appear to correspond to a large change in MYC protein stability as determined by western blotting after cycloheximide (CHX) treatment to block de novo translation (Fig S6C). There was no clear evidence of direct transcriptional activation of MYC by DUX4 based on prior chromatin immunoprecipitation sequencing (ChIP-seq) data [16] which did not identify a DUX4 binding site near the MYC gene (Fig S6A), although PolII ChIP showed a modest increase in promoter-associated and elongating PolII occupancy at the MYC promoter (Fig S6D), about 2-fold and 4-fold, respectively. Most dramatic, however, was a clear enhancement of MYC mRNA stability following DUX4 induction as evidenced by the increased half-life from roughly 44 minutes to 360 minutes in DUX4-expressing cells (Fig 10B). To determine whether this increase in mRNA stability was specific to MYC or a more general effect on mRNA stability, we assessed the stability of two other labile mRNAs, JUN and CITED2, following DUX4 expression. Although these results were less dramatic, the trend indicated an increase in mRNA stability of both genes following DUX4 expression (CITED2 from approximately 45 to 182 minutes and JUN from 52 to 140 minutes; Fig S6E). We conclude that, although the effect might not be specific to MYC, a major contribution to the increased MYC levels was likely through inhibition of mRNA degradation. These results are in agreement with a more general inhibition of mRNA degradation pathways in DUX4-expressing cells, similar to the previously demonstrated inhibition of NMD which leads to the stabilization of numerous labile mRNAs [42].

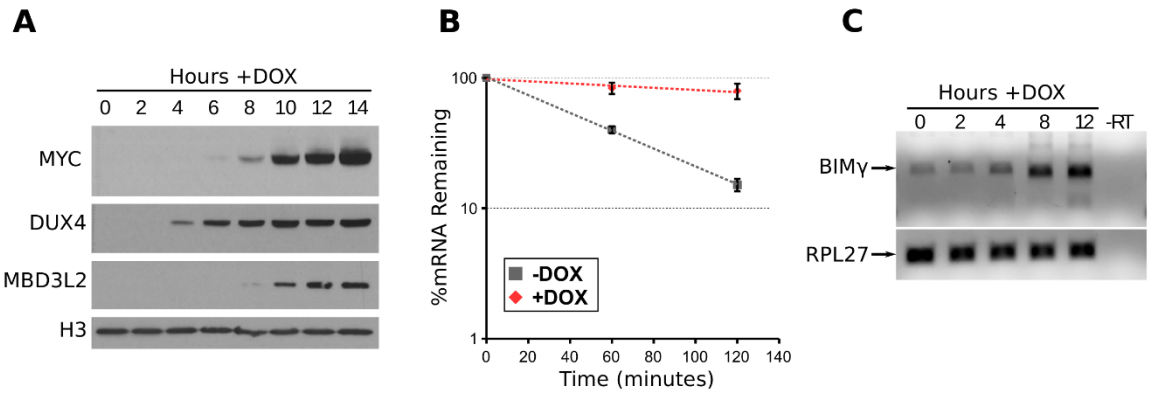


Figure 10 DUX4 increases MYC protein through MYC mRNA stabilization and induces the BIMγ isoform of BCL2L11. (A) Western blot of RD-DUX4i cell extracts following DUX4 induction at the indicated timepoints and using the indicated antibodies. (B) Semi-log plot of MYC mRNA levels in RD-DUX4i cells pretreated +/- doxycyline for 8 hours and in the presence of DRB for the indicated time points. Data are normalized to 18s rRNA levels which we reasoned would not be transcriptionally repressed by DRB, an RNA polymerase II inhibitor. Error bars depict standard deviation of the mean of three independent experiments. (C) RT-PCR of the BIMγ isoform following DUX4 expression in RD-DUX4i cells with the constitutively expressed RPL27 as a control.

Studies in other cells have shown that the MYC-mediated pathway of apoptosis involves MYC activation of EGR1, and then MYC and EGR1 together activate expression of two BH3-only factors BCL2L11 and PMAIP1 (also known as BIM and NOXA, respectively) which antagonize the anti-apoptotic BCL2 proteins and lead to the loss of mitochondrial membrane potential [85]. Analysis of our RNA-seq data revealed that DUX4 induced the expression of EGR1 (moderated fold-change of 48), induced the expression of BIMγ, a specific isoform BCL2L11 with an alternative BH3-like domain [86], and increased total BCL2L11 transcripts roughly 2-fold (Fig 10C and S4 Table). The siRNA library used in our screen contained two separate siRNA pools against different splice isoforms of BCL2L11. One of the two pools rescued RD-DUX4i viability (Z-score = 2.01) and contained three siRNAs that targeted only the BIMγ isoform (3 out of 4 siRNAs), whereas the second pool did not rescue (Z-score = -0.59) and

did not contain siRNAs that targeted only this isoform (Fig S6F). Together our data indicate that DUX4 expression resulted in decreased degradation of the MYC mRNA and increased abundance of the MYC protein with subsequent activation of components of the MYC-mediated apoptotic pathway, which has been shown to sensitize cells to apoptotic stimuli [87].

3.3.3 *DUX4 expression leads to accumulation of dsRNA and activation of an innate immune response*

As noted above, knockdown of RNASEL, a gene involved in innate immunity to viral infection, rescued RD cells from DUX4 lethality (DUX4 Z-Score = 3.71) and had minimal effect on the doxycyline induction of the luciferase transgene (Luciferase Z-Score = -0.07). This rescue was validated as on-target based on our deconvolution strategy (Fig S5B) and prompted us to look for other mediators of the cellular antiviral response in our screen. EIF2AK2/PKR is similarly involved in antiviral response and its knockdown marginally rescued from DUX4 toxicity in our original screen (Z-score 1.93), but it passed all other criteria of our validation and filtering process (Fig S5A-B). These results led us to further postulate that at least part of DUX4 toxicity might be mediated via triggering an innate (antiviral-like) immune response in expressing cells.

Both EIF2AK2/PKR and RNASEL are primary responders of the double stranded RNA innate immune response. Because DUX4 leads to the accumulation of normally degraded RNAs, including those destined for nonsense mediated decay [42], activates the expression of RNAs from repetitive regions and retrotransposons [16], and stabilizes some mRNAs (this study), we considered the possibility that DUX4-expressing cells could increase the abundance of aberrant, endogenously formed dsRNAs. In order to test whether DUX4-expressing cells accumulate dsRNAs, we performed immunofluorescence using the J2 antibody, which recognizes dsRNAs, irrespective of the sequence [88]. These experiments revealed that DUX4-expressing cells had strong, nuclear dsRNA foci which were not present in uninduced cells (Fig 11A). The signal observed after transfection of the dsRNA surrogate, poly(I:C), was similar in

intensity to DUX4 expressing cells, though occupying a more cytoplasmic compartment (Fig 11A). This signal is not a non-specific property of the J2 antibody as we detected similar focal staining in DUX4-expressing cells using the separate monoclonal dsRNA-recognizing antibody K1 (Fig S7A), nor is it an artifact of the doxycycline induction of transgene mRNA expression as our RD-LUCi line did not have similar nuclear staining after doxycycline treatment (Fig S7B). Immunofluorescence of DUX4-induced RD cells using either the J2 or K1 antibodies showed that approximately 5–10% of the nuclei had obvious dsRNA aggregation at 19 hours post induction, whereas the uninduced cells did not exhibit any detectable nuclear staining.

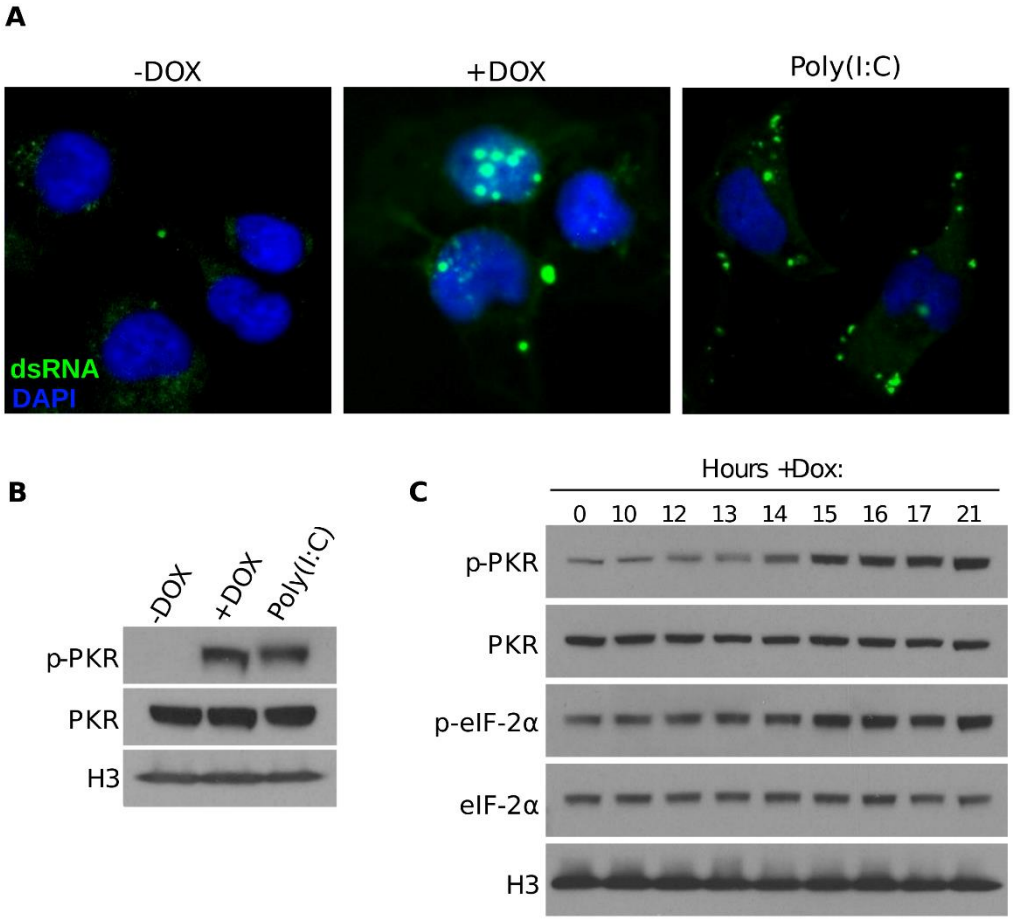


Figure 11 DUX4 induces nuclear dsRNA accumulation and phosphorylation of EIF2AK2/PKR and EIF2S1/eIF-2α. (A) Fluorescence microscopy using the J2 antibody to detect dsRNA in RD-DUX4i cells treated with doxycyline for 16 hours or transfected with

poly(I:C). Note that not all nuclei showed dsRNA aggregation, likely due to asynchronous activation of the DUX4 transgene. (B) Western blot of phosphorylated and unphosphorylated PKR in extracts from RD-DUX4i cells treated with doxycyline for 21 hours or transfected with poly(I:C). (C) Western blot in extracts from RD-DUX4i cells treated with doxycyline for the indicated times and using the indicated antibodies.

Accumulation of dsRNAs activates EIF2AK2/PKR via auto-phosphorylation. EIF2AK2/PKR, in turn, phosphorylates eukaryotic initiation factor 2 alpha (EIF2S1/eIF-2α), which is thought to generally inhibit both host and viral gene translation, culminating in cellular apoptosis [89]. To determine whether dsRNA accumulation leads to EIF2AK2/PKR activation in DUX4-expressing cells, we measured phosphorylated EIF2AK2/PKR in our DUX4-expressing RD cells and observed that, following DUX4 expression, EIF2AK2/PKR was phosphorylated to a level that is comparable to activation via transfection of poly(I:C), whereas total EIF2AK2/PKR levels remained unchanged (Fig 11B). To determine whether the EIF2AK2/PKR activation as a result of DUX4 expression has a functional consequence in our cells, we also blotted for EIF2S1/eIF-2α phosphorylation in a time course following doxycycline induction. This experiment showed that both EIF2AK2/PKR and EIF2S1/eIF-2α became phosphorylated approximately 15 hours post doxycycline addition (Fig 11C).

The profound inhibition of the NMD pathway by DUX4 [42] might be the cause of both the increased MYC mRNA and the accumulation of dsRNA because the NMD endonuclease SMG6 has been shown to degrade MYC as a non-canonical NMD target [90] and NMD inhibition leads to dsRNA accumulation in yeast [91]. Therefore, we used siRNAs to knockdown two components of the NMD pathway, UPF1 and SMG6. Knockdown of either NMD component alone or in combination showed a mild, non-significant ~1.1–1.6 fold increase in MYC mRNA and a ~1.6–2.3 fold increase in the inclusion of an NMD-targeted exon in SRSF3, compared to the more than 24-fold increase in these RNAs following DUX4 induction (Fig S8). Therefore, NMD inhibition through these mechanisms was not sufficient to stabilize the MYC mRNA.

3.3.4 *DUX4 expression in human myoblasts leads to MYC mRNA and dsRNA accumulation with activation of downstream apoptotic pathways*

To determine whether DUX4 induces the same MYC-mediated apoptotic pathways and dsRNA toxicity in human myoblasts, we assessed the expression of MYC, components of the MYC-mediated apoptotic pathway, and the dsRNA innate immune response in an immortalized human myoblast cell line with a doxcycycline-inducible DUX4 (MB135-DUX4i). Our previous RNA-seq data generated using MB135-DUX4i cells fourteen hours after DUX4 induction [92] showed an approximate 3.3-fold increase (based on an EdgeR differential expression analysis) in MYC mRNA abundance and, interestingly, a preferential accumulation of transcripts from the P1 promoter (Fig 12A). Similar to RD cells, DUX4 increased the abundance of EGR1 (5.4-fold) and the BH3-only proteins PMAIP1/NOXA (11-fold) and BCL2L11 (4.7-fold) (see Supplemental Table 1 in reference [92]), and induced expression of the BIMγ isoform of BCL2L11 (Fig 12B). Therefore, the same MYC-mediated apoptotic pathways induced by DUX4 in RD cells were similarly induced by DUX4 in human myoblasts.

To determine whether DUX4 expression in human myoblasts activated the same dsRNA pathway as in RD cells, we stained DUX4-expressing myoblasts with the J2 antibody and found strong focal nuclear dsRNA signal, similar to the observed pattern in RD cells (Fig 12C). Expression of DUX4 also led to phosphorylation of EIF2AK2/PKR as well as EIF2S1/eIF-2α in DUX4 expressing myoblast cells, comparable to levels seen after poly(I:C) transfection (Fig 12D).

Similar to RD cells, knockdown of either MYC or RNASEL rescued the MB135-DUX4i human myoblasts from DUX4-induced cell death, which was measured by counting viable cells at two and four days after induction of DUX4 expression (Fig S9A). However, knock-down of either MYC or RNASEL in human myoblasts also caused a delay in the accumulation of DUX4 protein and the activation of a DUX4 target gene, MBD3L2 (Fig S9B-C). Although DUX4 expression was robust at four days following induction (Fig S9B-C), a time when the MYC

knock-down continued to show strong rescue, these rescue experiments should be interpreted cautiously regarding the individual necessity of MYC or RNASEL for DUX4-induced apoptosis in human myoblasts. In summary, expression of DUX4 in myoblasts, as in RD cells, resulted in increased MYC mRNA and components of the MYC-mediated apoptotic pathways as well as accumulation of dsRNA and activation of the pro-apoptotic dsRNA-sensing innate immune response.

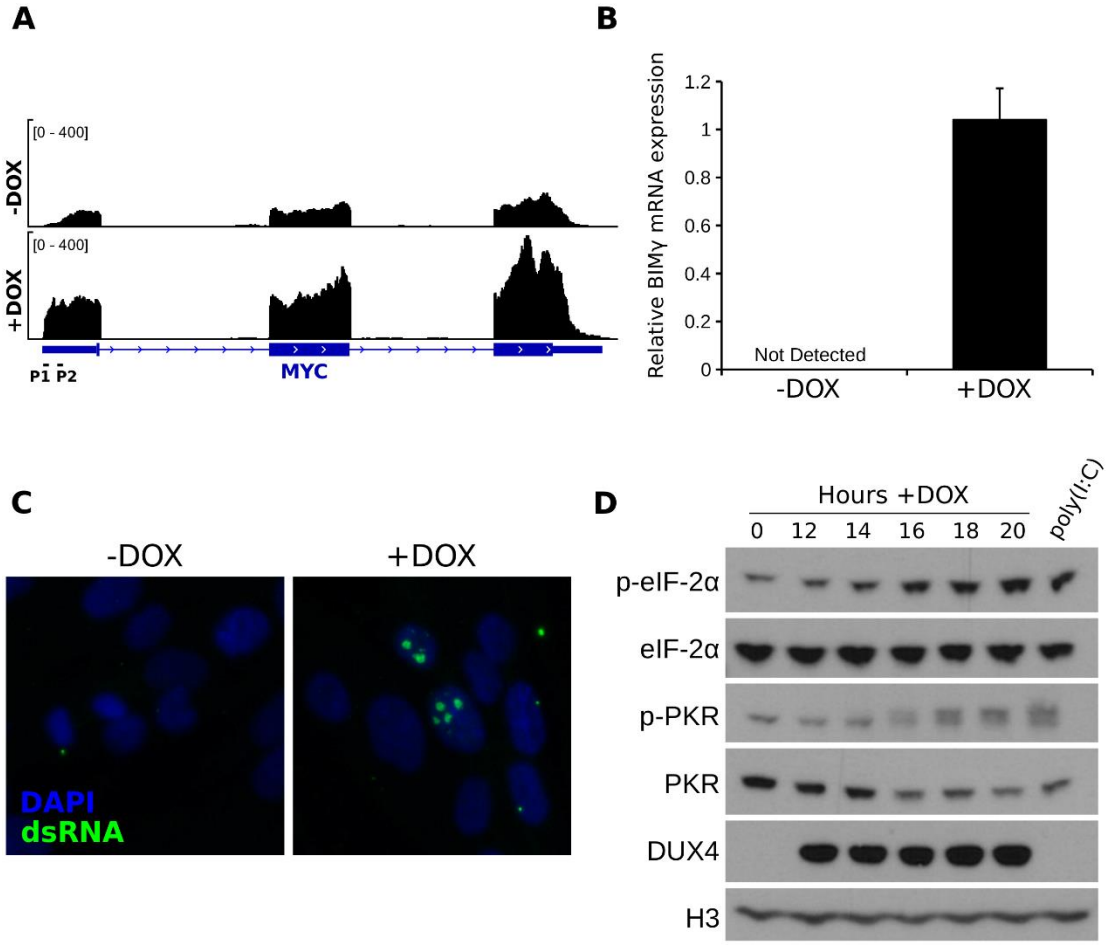


Figure 12 DUX4 induces components of the MYC-mediated apoptotic pathway and the dsRNA response pathway in human muscle cells and FSHD myotubes.(A) Example RNA-seq track showing reads at the MYC locus in MB135-DUX4i cells +/- doxycyline for 14 hours. (B) RT-qPCR using BIMγ isoform specific primers in MB135-DUX4i cells +/- doxycyline for 12

hours. Error bars depict standard deviation of the mean of two independent experiments (C) Confocal microscopy image of MB135-DUX4i cells +/- doxycyline for 20 hours and using the J2 antibody to detect dsRNA. (D) Western blot of MB135-DUX4i cell extracts following DUX4 induction showing phosphorylation of EIF2AK2/PKR and EIF2S1/eIF-2α.

3.3.5 *EIF4A3 aggregates with DUX4-induced dsRNA and EIF4A3 knock-down inhibits NMD*

Next, we determined whether the dsRNA aggregates corresponded to known nuclear subdomains or RNA binding proteins by performing immunofluorescence using the K1 dsRNA antibody and costaining with antibodies to promyelocytic leukemia bodies (PML), paraspeckles (NONO), the exon junction complex (EJC) components EIF4A3 and RBM8A/Y14, and TARDBP/TDP-43, which has been shown to aggregate in DUX4-expressing nuclei [39] and may regulate cellular dsRNA accumulation [43]. Components of PML bodies and paraspeckles were not strictly associated with dsRNA staining (Fig S10A), though NONO did form condensations in the nucleus following DUX4 induction, some of which overlapped with dsRNA staining (Fig S10B). TARDBP/TDP-43 did not show a clear association with the dsRNA foci (Fig S10D).

In contrast, DUX4 strongly induced redistribution of EIF4A3 into aggregates that were almost entirely associated with dsRNA foci as determined by double-staining with either the J2 or K1 antibodies (Fig 13A). In many cases DUX4-induced cells with EIF4A3 aggregates appeared to have reduced nuclear staining beyond the aggregates (Fig 13A) suggesting that the majority of EIF4A3 was redistributed to the dsRNA foci, similar to the depletion of MBNL proteins by the mutant nuclear RNA in myotonic dystrophy [93,94]. The other EJC component tested, RBM8A/Y14, showed some association with the dsRNA foci but was not strictly associated with dsRNA aggregates, nor was it depleted from the rest of the nucleus (Fig S10C).

A time course staining cells at 0, 16, or 24 hours following DUX4 induction with antibodies to either dsRNA or EIF4A3 (Fig 13B) showed the initial accumulation of cytoplasmic dsRNA at 16 hours, a time point where EIF2AK2/PKR and EIF2S1/eIF-2α phosphorylation was

becoming evident (see Fig 12D). The formation of nuclear dsRNA foci initiated at 16 hours and increased through 24 hours. EIF4A3 nuclear foci were present at 16 hrs and more abundant at 24 hours.

DUX4 expression inhibits NMD and at least part of this inhibition is secondary to decreased UPF1 protein levels [42], and knockdown of UPF1 showed a modest increase in NMD-targeted SRSF3 isoform in RD cells (Fig S8). Similarly, knockdown of EIF4A3 or RBM8A/Y14 in MB135 cells resulted in a substantial increase (~2.5–3.8 fold) in the NMD-targeted SRSF3 isoform (Fig 13C), suggesting that the nuclear sequestration of EIF4A3 might contribute to NMD inhibition in DUX4-expressing cells. MYC mRNA levels were not affected by EIF4A3 of RBM8A/Y14 knockdown (Fig 13C).

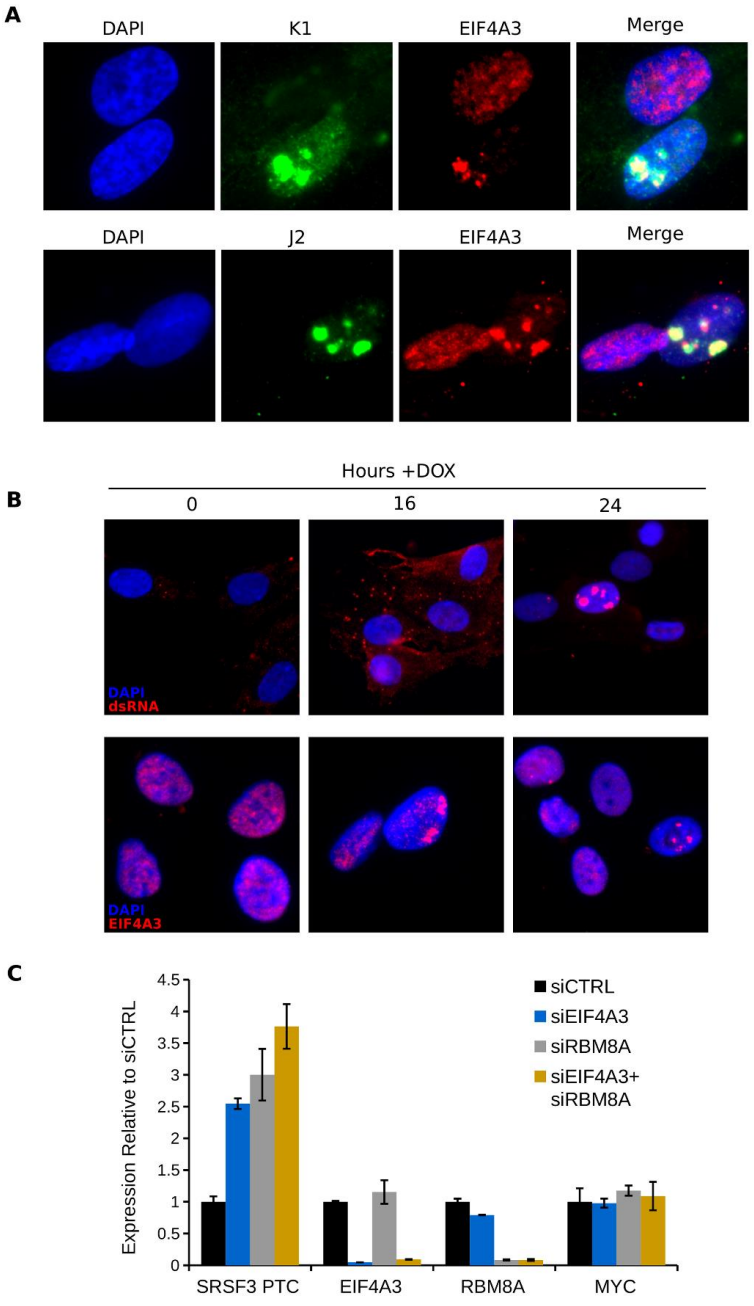


Figure 13 EIF4A3 aggregates with DUX4-induced dsRNAs correlating with inhibition of NMD.(A) Immunofluorescence showing fields of MB135-DUX4i containing nuclei with and without dsRNA aggregates determined using either K1 or J2 antibodies, as indicated, and co-stained against EIF4A3. (B) Time course experiment in MB135-DUX4i cells induced for the indicated period of time and stained against dsRNA (K1) or EIF4A3, as indicated. (C) RT-qPCR using the indicated primers and the corresponding knockdowns, harvested 48 hours

after transfection. SRSF3 PTC indicates SRSF3 isoform normally degraded by NMD. Data are normalized to RPL27A and shown relative to the siCTRL condition. Error bars represent the standard deviation of the mean of three replicate wells.

3.3.6 *FSHD cells have increased MYC expression and foci of dsRNA and EIF4A3*

Endogenous DUX4 is expressed in only a small percentage of cultured FSHD muscle cells [14]. Therefore, to determine whether DUX4 expression in FSHD cells correlates with higher levels of MYC mRNA, we used an RNA-seq dataset from FSHD cells FACS sorted based on the expression of a DUX4-reporter gene [57]. Based on our previous analysis of this dataset (see Supplemental Table 1 in Jagannathan, et al [92]), muscle cells expressing the DUX4-responsive promoter showed an almost 2-fold increase in the level of MYC mRNA (log2 fold-change ~0.9 and adjusted p-value = ~0.003); however, EGR1, PMAIP1/NOXA and BCL2L11 were not significantly upregulated. Therefore, DUX4 expression is associated with higher levels of MYC mRNA in FSHD muscle cells, but the role in activating an apoptotic pathway requires further study.

Immunofluorescence for dsRNAs in differentiated FSHD muscle cells showed either cytoplasmic or nuclear dsRNA staining in some of the cells with DUX4-positive nuclei but not in a control cell line (Fig 14A). In addition, the nuclear dsRNA foci in FSHD muscle cells were associated with EIF4A3 aggregates that were not present in control muscle cells (Fig 14B). Therefore, DUX4-expressing FSHD muscle cells showed increased levels of MYC mRNA and formation of nuclear foci with dsRNA and EIF4A3 accumulation, indicating that the initial screen identified pathways relevant to FSHD biology, and perhaps pathophysiology.

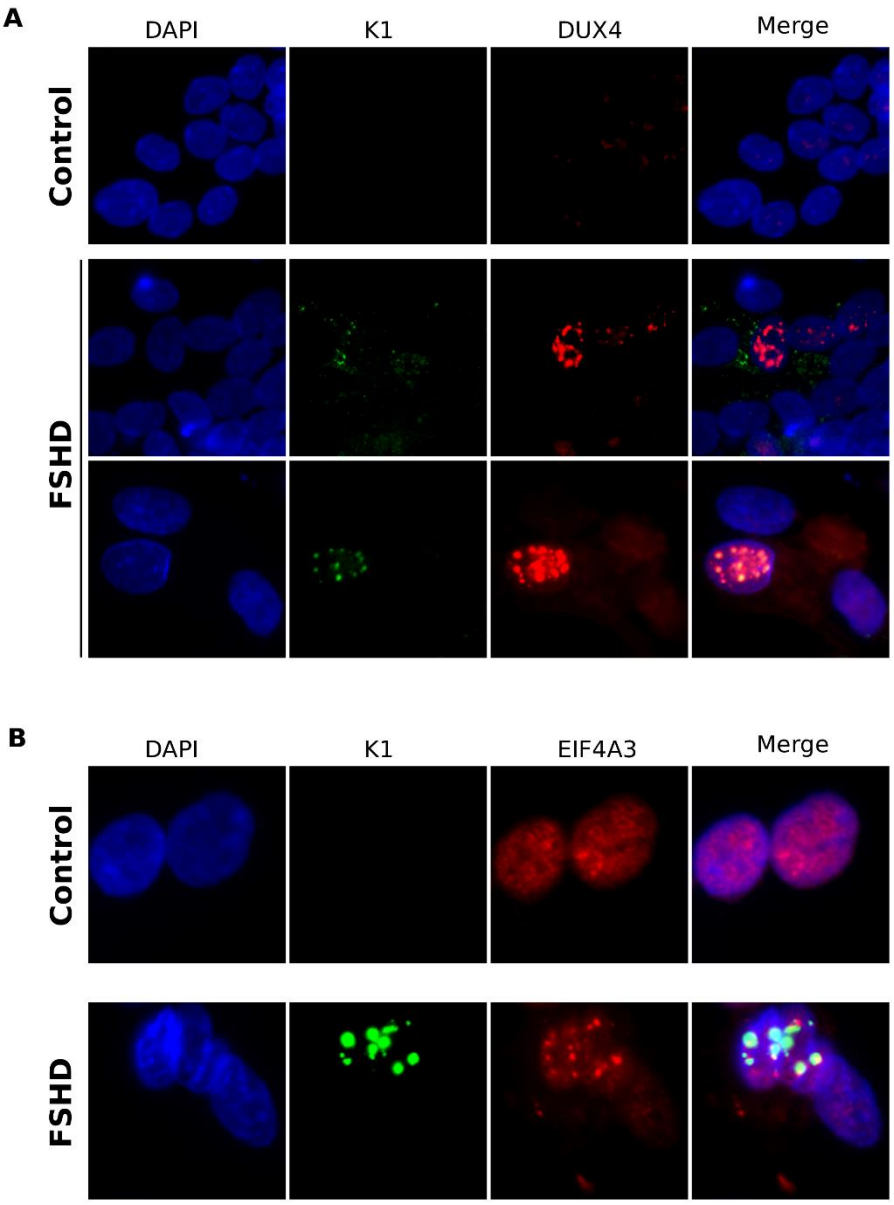


Figure 14 FSHD muscle cells have increased dsRNA and EIF4A3 aggregation. (A) Immunofluorescence showing dsRNA (K1) and DUX4 staining in control (MB135) and FSHD (MB200) muscle cells differentiated for 48 hours. (B) as in (A) but co-stained with K1 and EIF4A3.

3.4 *Discussion*

In this study, an siRNA screen identified that the MYC-mediated apoptotic pathway as well as components of the dsRNA response pathways were involved in DUX4-induced apoptosis in RD cells, and led to the demonstration that DUX4 expression resulted in the stabilization and accumulation of MYC mRNA, the accumulation of nuclear dsRNA, and formation of nuclear EIF4A3 foci. The accumulation of MYC mRNA correlated with higher MYC protein levels and the transcriptional activation of components of the MYC-mediated pro-apoptotic pathway, including EGR1, BCL2L11 and its BIMγ isoform. The accumulation of dsRNA was correlated with the activation of the EIF2AK2/PKR innate immune response and the phosphorylation of EIF2S1/eIF-2α. Inhibiting the expression of individual components of each of these pathways with siRNAs diminished apoptosis in RD cells in response to DUX4. These pathways were also induced by DUX4 in an immortalized human skeletal muscle cell line, with increased MYC RNA and the downstream effectors of MYC-mediated apoptosis (EGR1, PMAIP1/NOXA, and the BIMγ isoform of BCL2L11), as well as the accumulation of dsRNA, EIF2AK2/PKR and EIF2S1/eIF-2α phosphorylation, and nuclear foci consisting of EIF4A3 and dsRNA. Finally, the elevation of MYC mRNA, dsRNA accumulation, and EIF4A3 nuclear foci in FSHD muscle cells suggests that these processes might contribute to FSHD pathophysiology.

Although DUX4 expression caused apoptosis in a P53-dependent manner in some cells [55], we found that P53 was not necessary for apoptosis in either RD cells or human myoblasts. It is possible that different requirements for P53 might be due to differences in the cell type or the mechanism of DUX4 delivery. This would be consistent with the fact that MYC causes apoptosis in a manner that is either P53-dependent or P53-independent, depending on the cellular context [95].

EIF4A3 is a component of the EJC complex that binds the exon-junction of spliced RNAs [96]. Although the mechanism of sequestering EIF4A3 to nuclear foci remains to be determined,

the co-localization of the EIF4A3 foci with dsRNA foci suggests that DUX4 induction and/or stabilization of RNAs with dsRNA secondary structure might overwhelm normal nuclear RNA processing and transport, leading to their accumulation as foci with associated RNA binding proteins. The formation of EIF4A3 nuclear aggregates might contribute to the inhibition of NMD following DUX4 expression, because the formation of nuclear aggregates appears to deplete EIF4A3 from the rest of the nucleus and knockdown of EIF4A3 inhibits NMD. In this regard, FSHD might have some parallels with myotonic dystrophy, where the MBNL RNA binding proteins are sequestered in nuclear foci by the repeat expansion in the mutant RNA [94]. However, our prior demonstration that DUX4 expression leads to degradation of the UPF1 protein indicates that the possible sequestration of EIF4A3 is not the only mechanism by which DUX4 might inhibit NMD, and that simply re-expressing EIF4A3 might be unlikely to fully rescue this defect. Further studies will be necessary to determine the identity of the DUX4-induced dsRNAs and the RNAs that mediate EIF4A3 aggregation.

Although it is currently difficult to diagram a clear epistatic model from our results, there are multiple known interactions between MYC, the RNA-mediated innate immune response, and NMD. It has been shown that MYC overexpression can inhibit NMD, leading to ROS-mediated stress and EIF2S1/eIF-2α phosphorylation [97], both of which are associated with DUX4 expression ([36] and this study). Thus the elevated MYC in DUX4 expressing cells might also contribute to NMD inhibition, feeding forward to further enhance levels of DUX4, which is also an NMD target [42]. Inhibition of NMD might also lead to the accumulation of dsRNAs. In yeast, NMD controls the accumulation of antisense long non-coding derived dsRNAs [91]. It is therefore possible that the profound inhibition of NMD by DUX4 might similarly result in dsRNA accumulation, either from constitutively expressed RNAs or the many repetitive RNA families and LTR-containing ERVs transcriptionally activated by DUX4 [16,61]. Elevated levels of dsRNAs would then lead to an innate immune response, as has been observed in cancer cells that reactivate LTR-containing ERVs after DNA methylation inhibitor treatment [47,48]. While we

observed phosphorylation of EIF2AK2/PKR, which occurs upon binding dsRNA ligand [98], we did not see a robust interferon response (based on our RNA-seq data). However, DUX4 has been shown to inhibit components of the innate immune response pathway, at least in part through upregulation of DEFB103 [15], suggesting that the activation of this pathway by DUX4 might have a normal developmental role.

The apoptotic pathways induced by DUX4 in RD cells and human myoblasts might give insight into its normal role in development. DUX4 is normally expressed in the testis, likely in the early germline cells such as the spermatogonia and primary spermatocytes, and has recently been reported to be expressed in the thymus, both areas of active developmental apoptosis [14,74]. It is possible that one normal function for DUX4 is to elicit cellular apoptosis, and indeed the BIMγ isoform of BCL2L11 that we showed is induced by DUX4 in skeletal muscle is also expressed in the testis [86]. However, the broad transcriptional program activated by DUX4 suggests a role in stem cell function beyond apoptosis, and apoptosis might occur in a cell-context dependent manner. In other words, cells that normally express DUX4 might be primed for apoptosis but also resistant in the correct developmental context. As mentioned above, DUX4 expression blocks the RNA-induced innate immune response to viral infection [15], which might constitute a mechanism for protecting DUX4-expressing cells from toxicity in some normal developmental contexts compared to the expression of DUX4 in skeletal muscle cells. We speculate that DUX4 expression, which leads to elevated MYC levels, might afford an accompanying growth or fitness advantage in the developmental context where cells are competing for resources and survival. This notion is further supported by the fact that the MYC-mediated competitive advantage depends on the innate immune response pathway [99]. This normal process of cell competition is believed to control developmental size and eliminate cells with weaker attributes. Furthermore, the competitive advantage conferred by MYC might also be used in cancers for clonal expansion [100]. In this regard, it is interesting to note that DUX4 expression has recently been shown to be a causal factor in a subset of B-cell leukemias

[101,102]. Further work will be necessary to determine how the DUX4-mediated modulation of MYC levels and RNA accumulation might contribute to aspects of development and the pathophysiology of FSHD.

3.5 *Materials and methods*

Ethics statement

This study used pre-existing de-identified human cell lines from approved repositories and was determined to not be Human Subjects Research by the FHCRC Institutional Review Board. The commonly used RD rhabdomyosarcoma cell line was obtained from the ATCC (www.atcc.org). Myoblast samples were obtained from the Fields center at the University of Rochester (https://www.urmc.rochester.edu/fields-center.aspx) and patients gave informed written consent prior to sample use. For the use of third party data, no additional ethics approval was required and original ethics approval can be viewed in Jagannathan, et al [92] and Rickard, et al [57].

Cell culture

RD-DUX4i cells were grown in DMEM (Gibco) supplemented with 10% FBS (Hyclone), 1% penicillin/streptomycin (Gibco) and 2.0 μg/ml puromycin (Sigma). The DUX4 transgene was induced with 1.0 μg/ml of doxycycline hyclate (Sigma). Unaffected (MB135) and FSHD (MB200) immortalized human myoblasts were cultured in F10 medium (Gibco/Life Technologies) supplemented with 20% FBS and 1% penicillin/streptomycin as well as 10 ng/ml recombinant human FGF (Promega) and 1 μM dexamethasone (Sigma). Human myoblasts were differentiated by culturing in knockout serum replacement medium (Gibco/Life Technologies).

Cloning, virus production and monoclonal cell line isolation

The DUX4 and firefly luciferase genes were subcloned into the NheI and SalI sites of the pCW57.1 vector (Addgene plasmid #41393). Lentiviral particles were produced by co-transfecting 293T cells with the pCW57.1 vector, pMD2.G (Addgene plasmid #12259) and psPAX2 (Addgene plasmid #12260) using Lipofectamine 2000 (ThermoFisher) and following the manufacturer's instructions. Human RD cells were transduced and selected using 2.0 µg/ml puromycin (Sigma) at a sufficiently low multiplicity of infection to allow for clonal lines to be isolated using cloning cylinders. The MB135-DUX4i monoclonal cell line was produced as described in [92].

Caspase activity assay

RD-DUX4i cells were seeded in a 96-well plate, induced with doxycycline for 48 hours, then assayed for caspase 3/7 activity using Caspase-Glo Assay Technology (Promega) according to the manufacturer's instructions.

Cell viability and luciferase assays

RD-DUX4i cells were reverse transfected using LipofectamineRNAiMAX (Invitrogen) reagent in 96-well plates with 6,000 cells/well and 1.25 pmol/well of siRNAs (flexitube, Qiagen). Cells were induced with doxycycline at 1.0 µg/ml for 48–96 hours prior to assaying, depending on the experiment. Viability was assessed using CellTiter-Glo assay (Promega) per the manufacturer's instructions. Luciferase activity was assessed using ONE-Glo EX Luciferase Assay System (Promega), per the manufacturer's instructions.

MB135 human myoblast TP53 CRISPR knockout

Duplexed oligonucleotide containing the guide RNA sequence against TP53 was cloned into lentiCRISPRv2 plasmid (Addgene plasmid #52961), essentially using the Zhang lab protocol [103]. We used the sequence: CGCTATCTGAGCAGCGCTCA for targeting purposes. Cells

were transduced with lentiviral particles produced as above, selected with 2.0 µg/ml puromycin and then clonally reseeded such that isolated colonies could be picked. Mutations were subsequently identified by Sanger sequencing Topo-cloned (ThermoFisher) PCR amplicons which flanked the targeted region. Cells were treated with 10.0 µM of actinomycin D (Sigma) in DMSO to test induction of P53.

siRNA library and screen parameters

The Human Druggable Genome siRNA Set V4.0 was purchased from QIAGEN and contains 27,844 unique siRNAs multiplexed to target 6,961 genes in “druggable” functional groups such as receptors, nucleic acid binders, kinases, transcription factors, and signaling molecules. The master library was diluted with nuclease-free water (HyClone) into a 96-well parent library at a concentration of 0.5 pmol/µl and then converted to a 384-well child library set. All library manipulations were performed with a Beckman Biomek FX liquid handling robot using filtered sterile tips (ART BioRobitix). 2.5 µl of 0.125 pmol/µl siRNA solution was added to each of three replicate 384-well assay plates (ThermoFisher Scientific). Opti-MEM (Gibco) and RNAiMAX transfection reagent (ThermoFisher Scientific) were combined at 3.0 µl/ml RNAiMAX to Opti-MEM and subsequently distributed at 12 µl/well. 25.0 µl of cells at a concentration of 60,000 cells/ml of growth medium was then added to each well and plates were incubated for 24 hours at 37°C. Medium was then aspirated using a BioTek ELx405 Select cell washer and doxycycline-containing growth medium was added at 25.0 µl/well using the BioTek Micro Flo Select and incubated for an additional 72 hours. CellTiter-Glo reagent was added at 20.0 µl/well and luminescence was read on a microplate reader (Perkin Elmer EnVision 2104). Analysis of data was performed using the CellHTS2 Bioconductor package [104] and was normalized by using the median method. Protein interaction network analysis was performed using ConsensusPathDB software considering only high-confidence protein interactions [105].

RNA half-life experiments

RD-DUX4i cells were pre-incubated in the presence of doxycycline for 8 hours prior to treatment with DRB (Sigma) at 75 μM for up to two hours and levels were quantified using real time qPCR.

RNA isolation and real time qPCR

Cells were seeded at 66 x103 cells/well in 12-well plates. RNA was isolated using TRIzol reagent (ThermoFisher Scientific) according to manufacturer's instructions. Isolated RNA was treated with DNaseI (ThermoFisher Scientific), heat inactivated and reverse transcribed into cDNA using Superscript III (ThermoFisherScientific) following the manufacturer's protocol. Real time qPCR was performed using SYBR green reagent (ThermoFisher Scientific) for quantification.

RNA-seq library preparation and data analysis

RNA-seq of RD-DUX4i cells was performed on RNA samples collected from 3 experimental replicates induced for 16–18 hours with doxycycline. Illumina TruSeq libraries were prepared with 500 ng total RNA per sample using the standard library preparation protocol. 100 bp single-end sequencing was performed using the HiSeq 2500 platform by the FHCRC Genomics facility. Base calling was performed with Real Time Analysis software version 1.18.66.3. Raw reads were aligned to UCSC hg38 using Bowtie2 version 2.2.6 and Tophat version 2.1.0 [68]. Gene counts were calculated using the Gencode version 22 annotation file (obtained from the UCSC genome browser) and the GenomicAlignments version 3.3 Bioconductor package [106] in R version 3.4.0 using the intersection-strict mode. Differential expression analysis was performed using DESeq2 version 1.12.4 [107]. FPKM reads were calculated using a custom R script. Processed data were visualized using IGV software [108]. Raw and processed data have been deposited onto the NCBI Gene Expression Omnibus under the accession number GSE87495.

Western blotting

Reduced and boiled samples (typically 10–20 µg total protein per assay) were run on NuPage 4–12% precast polyacrylamide gels (Life Technologies) and transferred to PVDF membrane (Life Technologies). After blocking in 5% milk in PBST, the membrane was incubated with appropriate antibodies (described below) in block solution overnight at 4°C. Membranes were then incubated with appropriate HPRT-conjugated secondary antibodies in block solution for one hour at room temperature and chemiluminescent substrate (ThermoFisher Scientific) was used for detection. Densitometry, when performed, was achieved using ImageJ software [109].

dsRNA immunofluorescence

For all experiments involving RD cells, the cells were fixed with 4% paraformaldehyde (Electron Microscopy Sciences) in PBS for seven minutes at room temperature prior to permeabilization with 0.5% Triton X-100 (Sigma) in PBS. Cells were then incubated with 1% BSA in PBS block solution for 30 minutes at room temperature and then incubated overnight with J2 or K1 antibody at a concentration of 2 µg/ml at 4°C. Appropriate FITC- or TRITC-conjugated secondary antibodies (Jackson ImmunoResearch) and DAPI stain (Sigma) were then used prior to visualization. For the human myoblast experiment, conditions were essentially the same, except that cells were fixed with 2% paraformaldehyde and co-incubated overnight with J2 or K1 and anti-DUX4 (E5-5) antibody. Cells were imaged on a Zeiss AxioPhot or, if indicated, a Leica TCS SP5 II confocal microscope and channel merging was performed using ImageJ software.

Chromatin immunoprecipitation

Cross-linked ChIP coupled with MNase digestion was performed in triplicate on RD-DUX4i cells post doxycycline induction as described in [110] with two minor modifications: decreased MNase concentration (15 units of Worthington MNase incubated at 37°C for 12 min) and greater

sonication intensity (4 pulses at 30% amplitude, 15 seconds per pulse with 1 min rest between pulses). Briefly, 10 million RD-DUX4i cells were harvested per sample, 12.5 hours after the induction of DUX4. After crosslinking with 1% formaldehyde for 10 min, samples were treated with MNase followed by sonication. 10 percent of the soluble chromatin from each sample was reserved as input and the remainder of each sample was divided into three equal aliquots for the mock, 8WG16, and Ser2P ChIPs. Pol II antibodies were pre-bound to Protein G dynabeads (ThermoFisher Scientific) for 4 hours at room temperature, according to the manufacturer's directions and incubated with chromatin overnight.

Antibodies (name, company, batch if available)

Histone H3, Abcam ab1791; MBD3L2, Abcam ab107999, lot GR126890-2; MYC (9E10), Santa Cruz Biotech SC-40; dsRNA, SciCons J2, batch J2-1505; dsRNA, SciCons K1, batch K1-1502; PKR, Cell Signaling D7F7, lot 1; phosphorylated PKR, Abcam ab32036, lot GR155191-6; eIF-2α, Santa Cruz Biotech SC-11386; phosphorylated eIF-2α, Abcam ab32157; 8WG16 UnphosphorylatedPolII, Abcam ab817, lot GR153063-402; Serine2 phosphorylated PolII, Abcam ab5095, lot GR271493-1; PML, Santa Cruz Biotech SC-5621, lot B1216; eIF4A3, Abcam ab180573, lot GR148643-3; nmt55/p54nrb (NONO), Abcam ab133574, lot GR97976-7; Y14 (RBM8A), Abcam ab181038 lot GR152694-1; TDP43, Proteintech 1078-2-AP; p53 (DO-1), Santa Cruz Biotech SC-126, lot l2316; rabbit monoclonal antibodies against DUX4 (E14-3 and E5-5) were produced in collaboration with Epitomics and are described elsewhere [111].

Primers (name and sequence)

MYC_5prime_chip_1F	GAGGCTATTCTGCCCATTTG
MYC_5prime_chip_1R	TCGGTGCTTACCTGGTTTTC
Chr2_neg_F1	TCACACTTCAGGAAAGCCCC
Chr2_neg_R1	GCAGGCCAGTTTGGGAAAAG

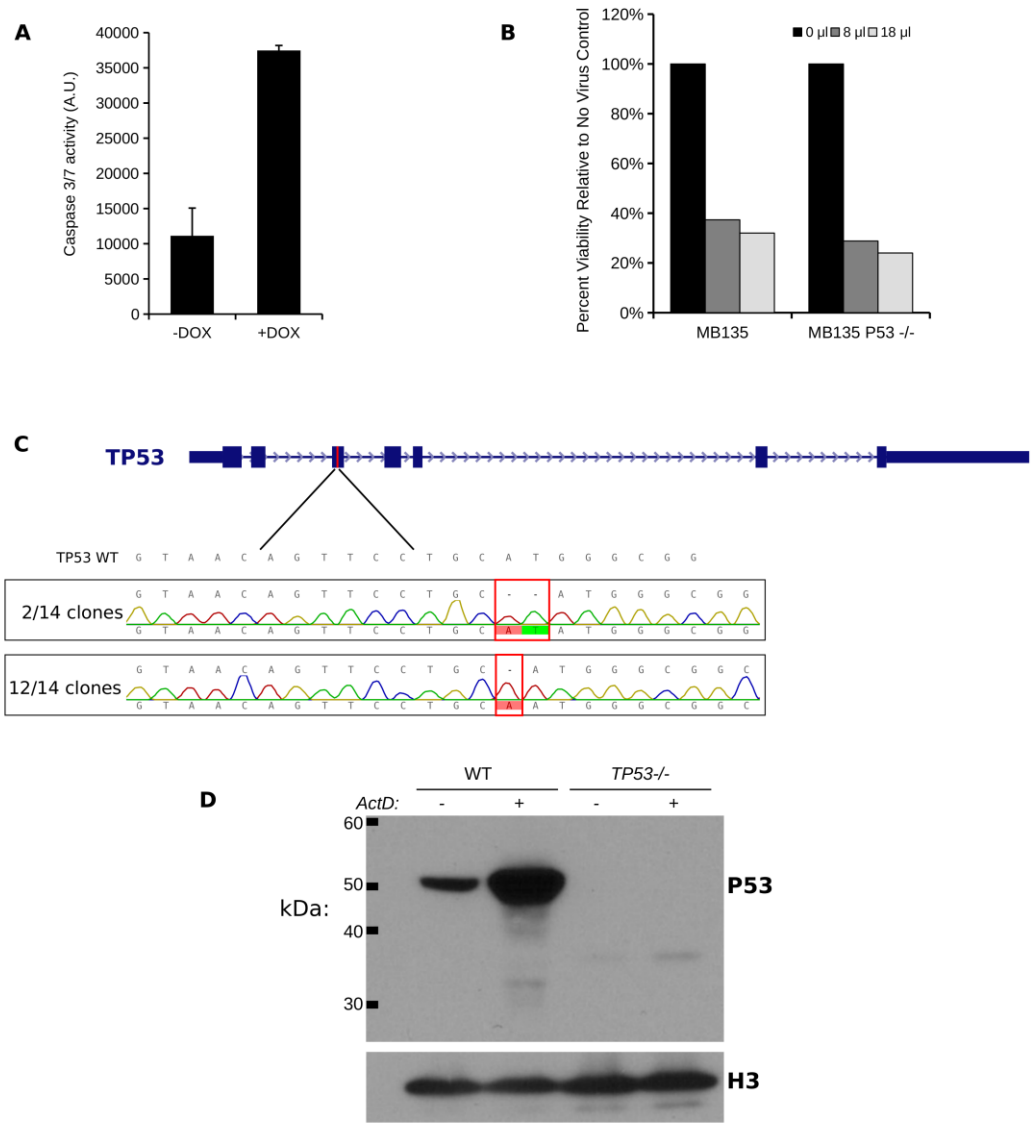
Leutx_5'_F1	CTGCAGCACACAGCTGATCG
Leutx_5'_R1	CTTGCCTTCGCCCAACTTAC
C-MYC-1F	TACAACACCCGAGCAAGGAC
C-MYC-1R	AGCTAACGTTGAGGGGCATC
JUN_1F	GGAGACAAGTGGCAGAGTCC
JUN_1R	CCAAGTTCAACAACCGGTGC
CITED2_1F	AAAGGGAACGGCTCCGAATCTG
CITED2_1R	GCCATCATATGGTCTGCCATTTC
LG185-F (Zscan4)	TGGAAATCAAGTGGCAAAAA
LG186-R (Zscan4)	CTGCATGTGGACGTGGAC
hDUX4_ex2-3 F2	CGGAGAACTGCCATTCTTTC
hDUX4_ex2-3 R2	CAGCCAGAATTTCACGGAAG
Bim_exon2_F_2	GGGCCCCTACCTCCCTAC
Bim_exon5_R_1	TGGTGGTGGCCATACAAATC
RPL27-1L	GCAAGAAGAAGATCGCCAAG
RPL27-1R	TCCAAGGGGATATCCACAGA
18s_rRNA_F	GTAACCCGTTGAACCCCATT
18s_rRNA_R	CCATCCAATCGGTAGTAGCG
JZ69-smg6-fwd	TGCTTACTTAAGGAGTCCGCC
JZ70-smg6-rev	TCAGGTCCGGGACAAAGGAA
RKB_321 (SRSF3 PTC)	GGGTGGTGAGAAGAGACATGA
RKB_322 (SRSF3 PTC)	CTTGGAGATCTGCGACGAG
RKB_328 (UPF1)	CAGCTCGCAGACTCTCACTTT
RKB_329 (UPF1)	TGCGTCTGGCTAGGAAGAGT
pCW57.1_DS2F (transgene)	CACCACCACCACCACAAGG
pCW57.1_DS2R (transgene)	GAACGGACGTGAAGAATGTG

JZRT_17-RNASEL-1F	TGAGGGACTGTCTGAGTGACC
JZRT_18-RNASEL-1R	TCAGATTTTCGTGTTTTGATGTCGG
JZ80-EIF4A3-1F	TGGCTCCCACAAGAGAGTTG
JZ80-EIF4A3-1R	CTTCCTGATGTCCTCGCCAA
JZ78-RBM8A-1F	GCGAAGATTTCGCCATGGAT
JZ79-RBM8A-1R	TCATAATCCTCACGCATCCGC

Supplementary Material

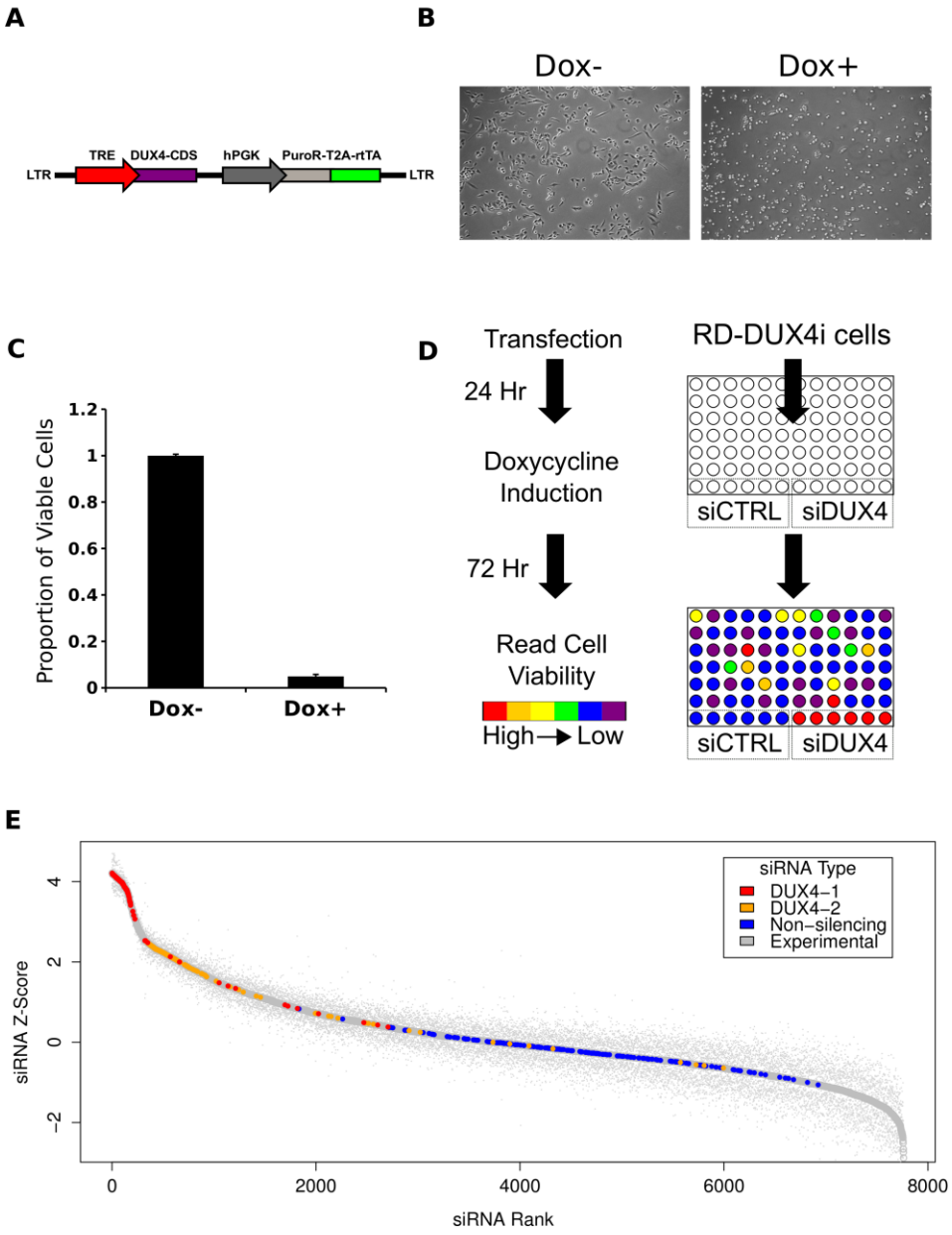
Supplemental Tables for this chapter are available on the PLOS Genetics website

3.6 *Supplemental Figures*



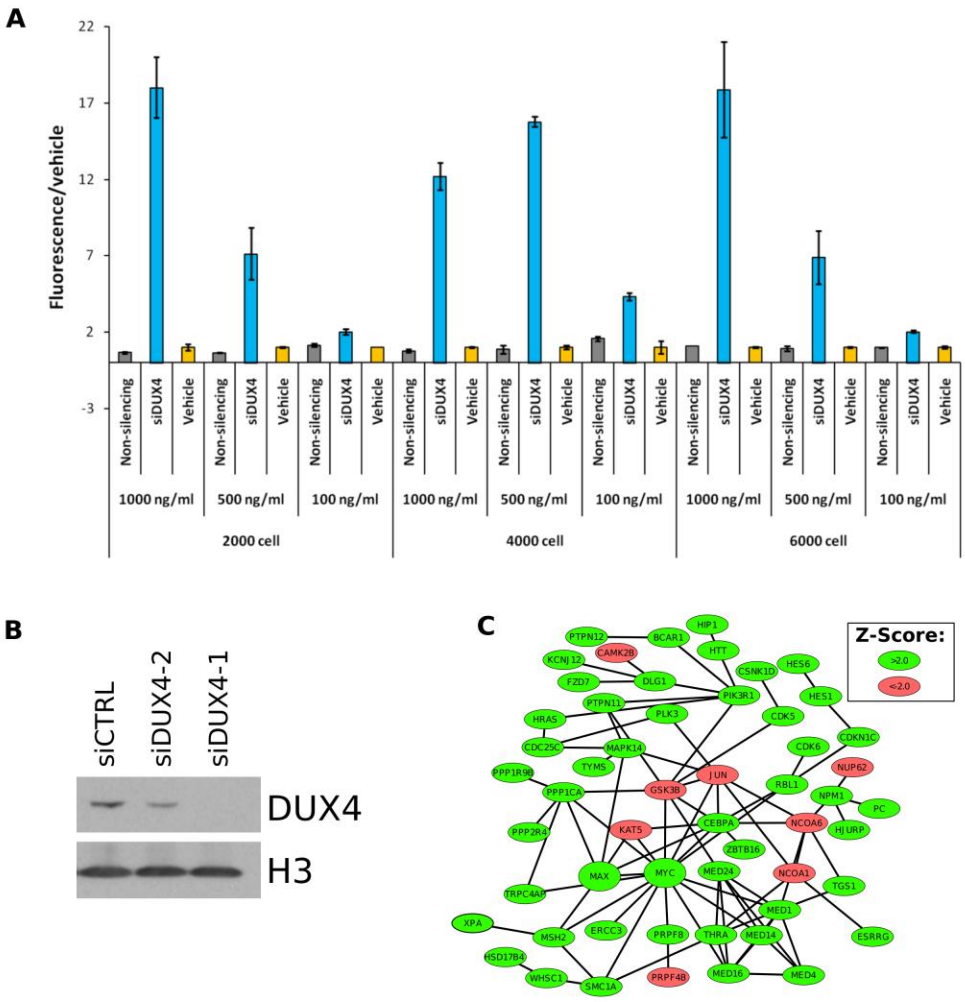
Supplemental Figure 2 DUX4 induces cell death in TP53 deficient cells. (A) Caspase 3/7 activity assay (Caspase-Glo) 48 hours following doxycycline induction in RD-DUX4i cells. Error bars represent the standard deviation of the mean of three replicate wells. (B) CellTiter-Glo viability assay of parental (unmodified) MB135 and TP53 knockout MB135 (MB135 P53-/-) immortalized human myoblast cell lines transduced with the indicated volume of DUX4 lentiviral expression vector. Note that, because DUX4 induces cell death, it is not possible to conventionally titer the virus. (C) Representative Sanger sequencing results of knockout cell line depicting the region of TP53 target site where an indel was induced by non-homologous end

joining from CRISPR/Cas9 directed cleavage. Two of 14 topo-cloned PCR amplicons had a two nucleotide insertion whereas the other 12 amplicons had a single nucleotide insertion at the cleavage site and no wild-type sequences were observed indicating that our TP53 knockout MB135 cell line has a frameshift mutation in both alleles. (D) Western blot showing P53 levels in WT (parental) and TP53 knockout MB135 cell line. P53 was induced by actinomycin D (ActD) which was added to growth medium for 24 hours prior to harvesting and serves as a positive control for detecting the endogenous levels of P53.

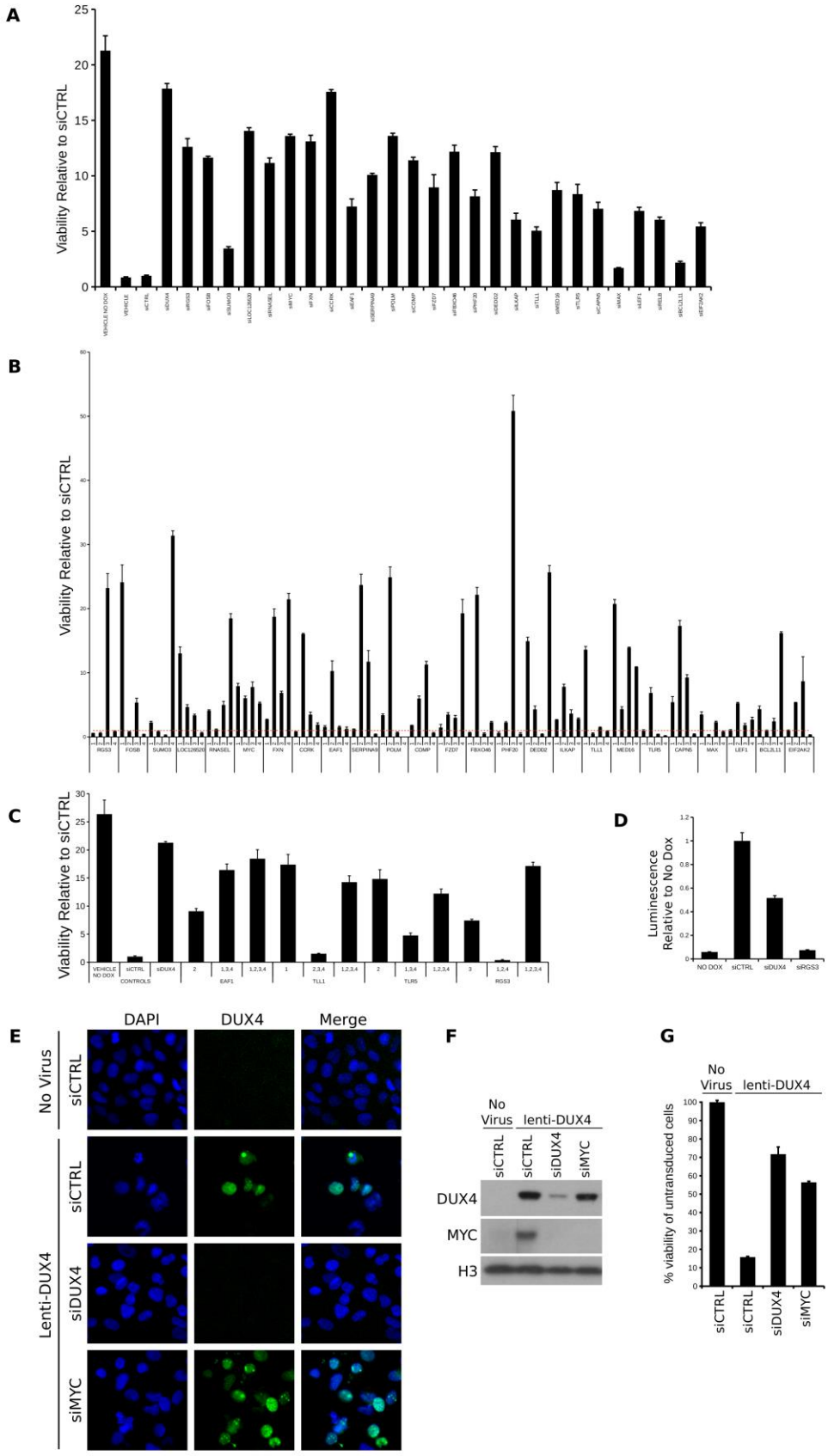


Supplemental Figure 3 siRNA screen identifies targets that diminish DUX4 toxicity in RD cells. (A) Schematic of the all-in-one pCW57.1 inducible lentiviral system used to express DUX4. Explanation of abbreviations used: TRE: tetracycline response element; CDS: coding DNA sequence; hPGK: human phosphoglycerate kinase 1 promoter; PuroR-T2A-rtTA: co-expressed puromycin N-acetyltransferase resistance gene, 2A peptide which yields separate translation of the tetracycline controlled transactivator. (B) Phase contrast images showing morphology of RD-DUX4i cells 24 hours +/- doxycyline. (C) CellTiter-Glo (ATP-based) assay 48 hours +/- doxycyline as a measure of cell viability. Data are relative to the “Dox-” condition. Error

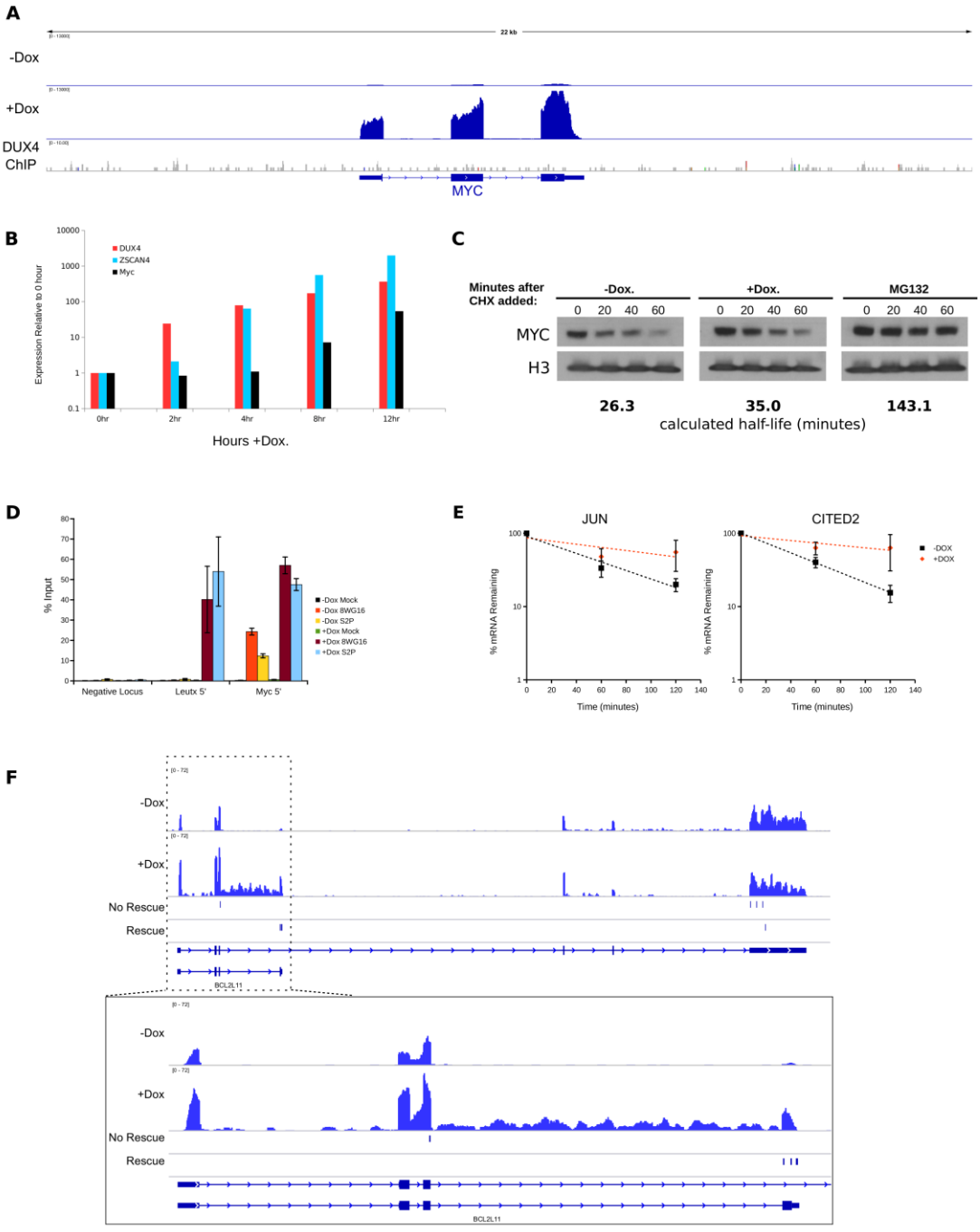
bars represent the standard deviation of the mean of three replicate wells. (D) Schematic showing optimized parameters used for the full scale siRNA screen. Briefly, cells were transfected in multi-well plates for 24 hours and subsequently induced to express DUX4 for 72 hours before cell viability was recorded using CellTiter-Glo reagent. (E) Plot ranking all individual siRNA targets from the siRNA screen. The mean of three triplicate wells (large points) and minimum and maximum values of triplicate wells (smaller points above and below) are shown. Note that DUX4-1 siRNA was more robust at knocking down the DUX4 transgene than DUX4-2 siRNA (see also Fig S4B).



Supplemental Figure 4 Optimization and network analysis of the siRNA screen. (A) CellTiter-Glo viability assay depicting an example of our strategy used to optimize parameters for the full-scale siRNA screen. In this example we varied cell number and dose of doxycyline (concentration in ng/ml). Error bars represent the standard deviation of the mean of three replicate wells. (B) Western blot of inducible DUX4 transgene expression 24 hours following indicated siRNA transfection and subsequent 5 hour induction. (C) ConsensusPathDB induced network module analysis of protein-protein interactions from |Z-score| > 2.0 of unfiltered screen results.

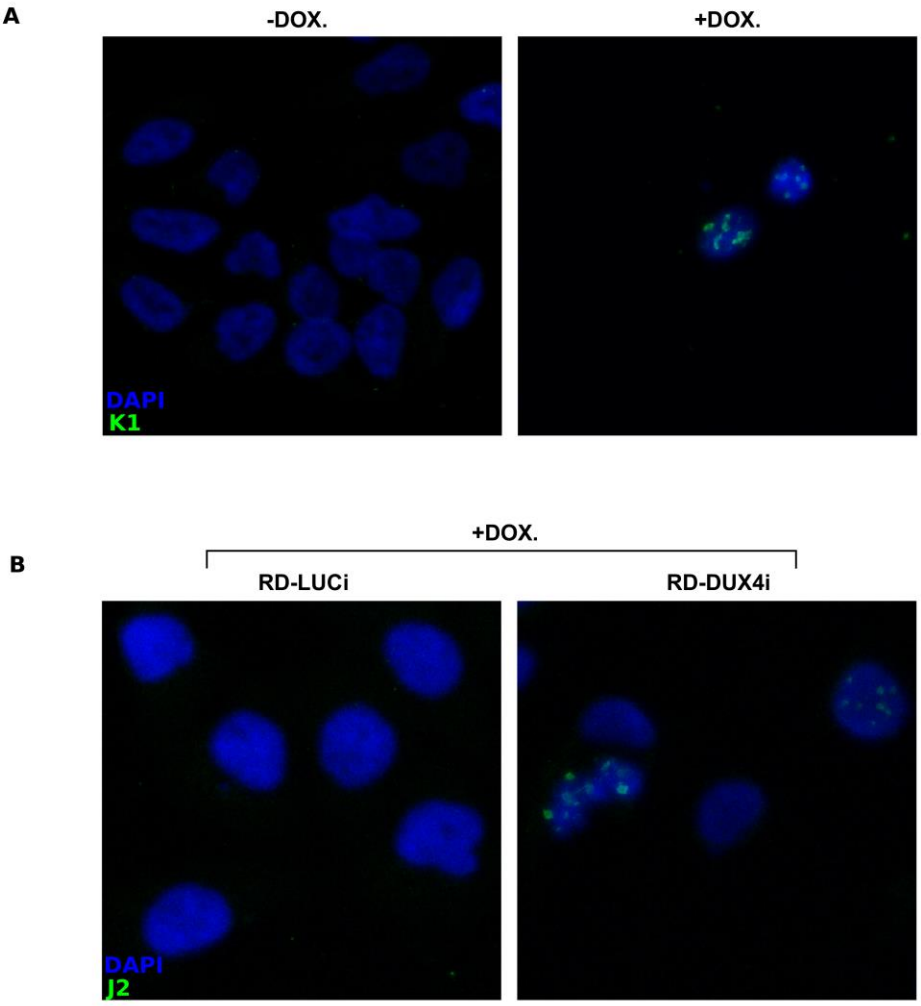

A
Viability Relative to siCTRL
B
Viability Relative to siCTRL
C
Viability Relative to siCTRL
D
Luminescence Relative to No Dox
NO DOX
siCTRL
siDUX4
siRGS3
E
DAPI
DUX4
Merge
No Virus
siCTRL
Lenti-DUX4
siCTRL
siDUX4
siMYC
F
No Virus
lenti-DUX4
siCTRL
siCTRL
siDUX4
siMYC
DUX4
MYC
H3
G
No Virus
lenti-DUX4
% viability of untransduced cells
siCTRL
siCTRL
siDUX4
siMYC

Supplemental Figure 5 Validation, deconvolution, and synergy screens of siRNA pools. (A) CellTiter-Glo viability assay of select rescuing targets from RD-DUX4i siRNA screen following transfection of indicated siRNA pools in order to determine reproducibility of the original experiment. Viability is shown relative to the siCTRL condition. (B) Deconvolution of pools as in (A). The red dotted line is set at 1.0 as a reference. (C) Viability assay testing pooling of 'non-rescuing' siRNAs from (B) in order to determine whether these siRNAs could 'synergize' or if the response was dominated by a single siRNA (likely off-target result). (D) RD-LUCi cells were treated with siRNAs for 24 hours and induced with doxycycline prior to reading luminescence of luciferase transgene. Error bars in all graphs represent the standard deviation of three replicate wells. (E) Immunofluorescence of DUX4 in RD cells that were transfected with the indicated siRNAs and, after 24 hours, transduced with lenti-DUX4 (pRRLSIN vector backbone with a human PGK promoter driving DUX4 expression). Images were taken 42 hours following DUX4 transduction, when clear viability differences between knockdown conditions were evident. Note that siMYC appeared to have no clear effect on either nuclear localization or overall expression of DUX4 compared to the control knockdown. (F) Western blot showing DUX4 and MYC protein levels following the indicated knockdowns at 18 hours after transduction of lenti-DUX4. (G) CellTiter-Glo viability assay following the indicated knockdowns at 48 hours after transduction of lenti-DUX4.

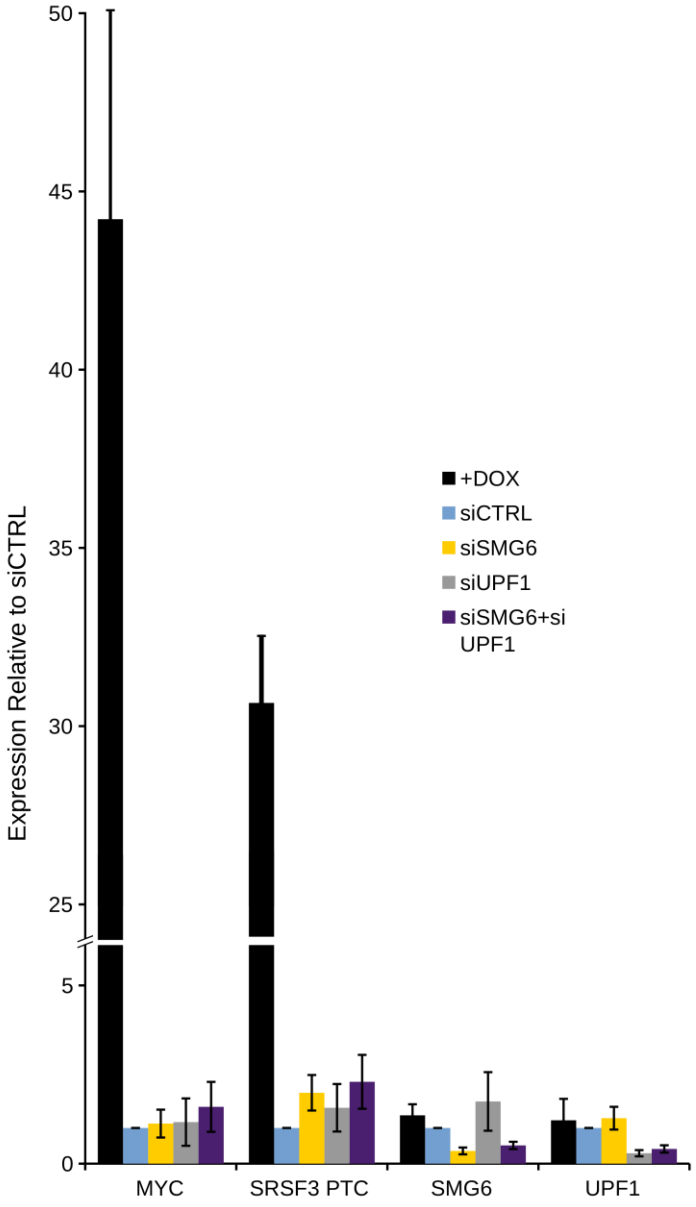


Supplemental Figure 6 Determination of DUX4 binding and activation of MYC, RNA stabilization, and the shorter BIMγ isoform of BCL2L11. (A) Track showing RNA-seq or ChIP-seq reads mapped 22kb at and surrounding the MYC locus. Note that there is no apparent DUX4 occupancy near the canonical MYC promoter nor elsewhere. (B) RT-qPCR data of MYC and ZSCAN4 (a direct transcriptional target of DUX4) following doxycyline induction in RD-

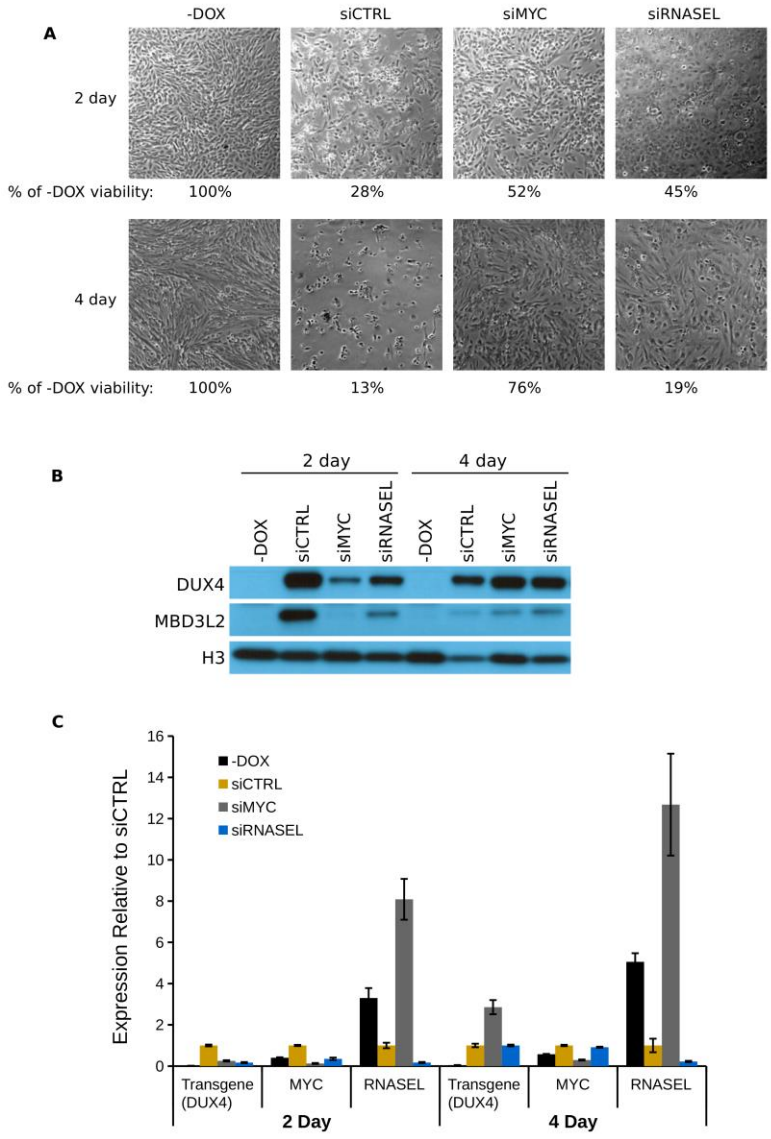
DUX4i cells. (C) Western blot for protein half-life experiment of MYC in RD-DUX4i cells. Cells were treated with or without doxycyline for 8 hours prior to the addition of the translation inhibitor, cyclohexamide (CHX). MG132, a proteosomal inhibitor, was included as a positive control and added during CHX addition. Densitometry was used to estimate relative protein levels compared to the zero-hour time point and data were fitted onto a semi-log plot in order to estimate the half-life of each condition. (D) ChIP-qPCR of unphosphorylated (8WG16) and Serine-2 phosphorylated forms of RNA polymerase II. The “negative locus” is a primer-set with no known annotated transcripts and serves as a negative control. Error bars represent the standard deviation of three replicate ChIP experiments. (E) mRNA half-life experiment as in Fig 10B. (F) RNA-seq track showing the location of the rescuing and non-rescuing BCL2L11 (BIM) pool of siRNAs in RD cells +/- doxycycline. The shorter splice form induced by DUX4 is the BIMγ isoform of BCL2L11 and the rescuing pool has three siRNA that specifically target this isoform.



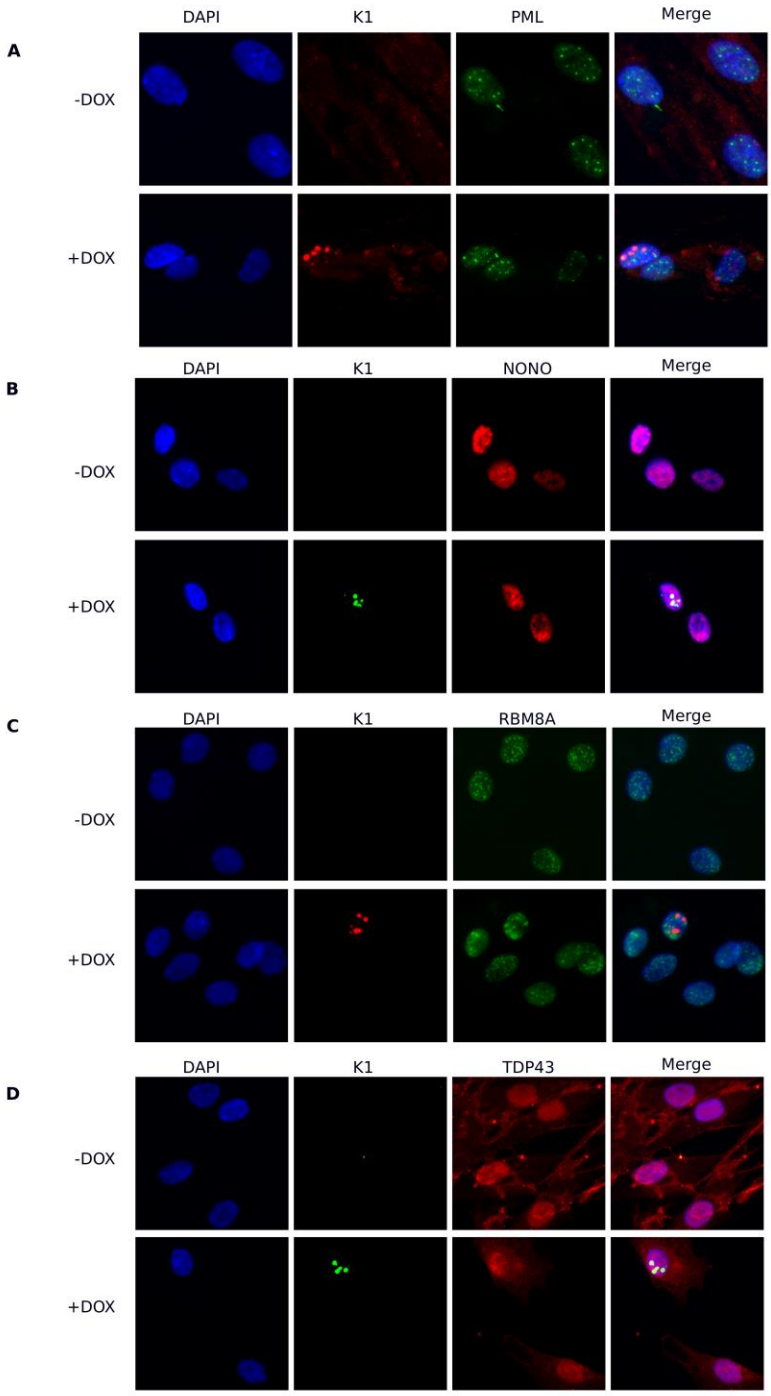
Supplemental Figure 7 dsRNA in RD cells expressing DUX4. (A) Confocal microscopy immunofluorescence of RD-DUX4i cells +/- doxycyline using the K1 antibody to verify the presence of dsRNAs in DUX4 expressing cells. (B) Confocal microscopy immunofluorescence of RD-LUCi and RD-DUX4i cells with doxycycline and stained with the J2 antibody. Note that, based on previous observations, we believe induction of DUX4 occurs asynchronously in the cell population using the doxycycline-inducible system, explaining why there is not universal nuclear dsRNA positivity at the time point assayed.



Supplemental Figure 8 siRNA knockown of NMD factors modestly elevates MYC and a canonical NMD target. RT-qPCR data on RD-DUX4i cells either induced with doxycycline (Dox+) or uninduced in the presence of the indicated siRNA and using the primer sets annotated on the X-axis. Error bars represent the standard deviation of three separate experiments.



Supplemental Figure 9 Knockdown of RNASEL and MYC rescue human myoblasts from DUX4 toxicity, but delay transgene expression. (A) Bright field images showing MB135-DUX4i cell morphology after DUX4 expression at 2 and 4 days post-induction. The ATP-based CellTiter-Glo assay gave unusually high readings in control myoblast cells for unknown reasons and thus we used viable cell counts as an alternative. (B) Western blot showing a DUX4 target gene, MBD3L2, and DUX4 transgene expression in MB135-DUX4i cells at the indicated time-points after induction. (C) RT-qPCR to confirm knockdown of target mRNAs. Data are normalized to RPL27A and shown relative to the siCTRL condition. Error bars represent the standard deviation of the mean of three replicate wells.



Supplemental Figure 10 Immunofluorescence panel of DUX4-induced dsRNAs with PML, NONO, RBM8A, and TDP-43. MB135-DUX4i cells were induced for 24 hours with doxycycline prior to co-staining with antibodies against dsRNA (K1) and (A) PML. (B) NONO (paraspeckles). (C) RBM8A. (D) TDP-43, none of which show exclusive overlap with dsRNA accumulation.

Chapter 4. *DUX4-induced transcripts from pericentric human satellite II repeats form double-stranded RNA*

This chapter is being prepared for publication as:

Shadle SC, Chao-Jen W, Bennett SR, Karreman NA, Campbell AE, van der Maarel SM, Bass, BL, Tapscott SJ. DUX4-induced transcripts from pericentric human satellite II repeats form double-stranded RNA. In preparation.

4.1 *Abstract*

The DUX4 transcription factor is normally expressed in the cleavage stage embryo and regulates genes involved in zygotic genome activation. Mis-expression of DUX4 in skeletal muscle, however, is toxic and causes facioscapulohumeral muscular dystrophy (FSHD). We recently showed DUX4-induced toxicity is due, in part, to the accumulation of double-stranded RNAs (dsRNAs) with concomitant activation of the PKR viral response pathway. Here, using dsRNA immunoprecipitation coupled with next-generation sequencing, we determined that DUX4-induced dsRNAs originate from intergenic regions enriched for *Alu* and LINE-1 elements, endogenous retroviruses, and pericentric human satellite II (HSATII) repeats. DUX4-induced HSATII dsRNAs are formed via temporally controlled bidirectional expression with predominant transcription of one strand preceding the other. Whereas DUX4 activation of dsRNAs derived from intergenic regions such as HSATII contributes to toxicity in FSHD, in the early embryo these dsRNAs might facilitate heterochromatin formation, as has been proposed for cleavage stage expression of mouse major satellites.

4.2 *Introduction*

The Double Homeobox 4 (DUX4) transcription factor is normally expressed in the testis, likely the germline cells [14], and in the cleavage stage embryo where it regulates genes in the early wave of zygotic genome activation [28–31]. In contrast, the mis-expression of DUX4 in skeletal muscle causes facioscapulohumeral muscular dystrophy (FSHD) [13,14,65]. In skeletal muscle cells, inappropriate expression of DUX4 activates the transcription of repetitive elements characteristic of the cleavage stage embryo such as endogenous retroviruses (ERVs) and the pericentric human satellite II (HSATII) repeats [15,16,29,31], sequences which are normally silenced in most tissues. In mice, the repressive patterning of major satellite repeats is established during cleavage stage development and correlates with their bidirectional transcription [35]. It has been postulated that accumulation of these bidirectional transcripts

leads to the formation of double-stranded RNA (dsRNA) that might direct heterochromatin formation [50], similar to centromeric silencing mechanisms in fission yeast [112]. However, it is currently not known which specific factor(s) direct the transcription of HSATII repeats nor whether they form dsRNA in human cells.

We recently reported that the expression of DUX4 in skeletal muscle cells resulted in the accumulation of dsRNAs, first in the cytoplasm and then as nuclear aggregates that co-localized with the exon junction binding factor EIF4A3 [113]. The accumulation of DUX4-induced dsRNAs correlated with PKR and eIF2α phosphorylation, both proapoptotic characteristics of the cellular innate immune response typically triggered by viral dsRNAs [114]. Knockdown of *EIF2AK2* (PKR) or *RNASEL*, another mediator of the dsRNA response pathway, decreased the amount of cell death in DUX4-expressing cells [113]. In contrast to the activation of the PKR pathway, expression of DUX4 efficiently blocks the interferon response [15]. The inhibition of the interferon arm of the dsRNA response pathway by DUX4 may constitute a protective response mechanism, suggesting that DUX4 might also induce dsRNA during normal developmental expression.

In this study, we identified the dsRNAs induced following DUX4 expression in human skeletal muscle cells. Confirming their dsRNA character, these identified DUX4-induced transcripts were modified in a manner consistent with dsRNA-specific editing by the adenosine deaminases that act on RNA (ADAR) enzymes [115]. Analysis of RNA from FSHD muscle cells and RNA-seq data from cleavage stage human embryos, the stage when endogenous DUX4 is normally expressed, also showed expression of these dsRNA-forming transcripts and sequence modifications indicative of dsRNA editing by ADARs. In contrast to the mainly intronic origin of edited endogenous human dsRNAs [116–118], a large portion of DUX4-induced dsRNAs originated from intergenic, non-protein coding regions of the genome. The DUX4-induced dsRNAs were modestly enriched for LINE-1 elements and LTR-containing repeats and massively enriched for HSATII satellite sequences. We further demonstrate that DUX4-induced

HSATII sequences formed dsRNAs upon their bidirectional expression, which occurred in a temporally-controlled manner, with predominant transcription of a single strand preceding transcription of the opposite strand. This formation of HSATII RNA foci was associated with accumulation of EIF4A3 and ADAR1 into dense nuclear foci and correlated with the formation of polycomb bodies. While the induction of HSATII dsRNAs in muscle cells contributes to DUX4-mediated toxicity through activation of the PKR innate immune response pathway, in early development these dsRNAs might broadly alter RNA processing and chromatin repression.

4.3 *Results*

4.3.1 *Double-stranded RNA immunoprecipitation and sequencing identifies regions of dsRNA induced by DUX4*

Previous studies have identified double-stranded RNAs by high-throughput sequencing of immunoprecipitated RNAs (dsRIP-seq) which were pulled down with the widely used J2 dsRNA antibody [117,119,120]. To identify dsRNAs which are induced by DUX4 we performed dsRIP-seq on the doxycycline (DOX) responsive DUX4 inducible human myoblast cell line (MB135-iDUX4) with either the J2 or the independent K1 dsRNA-recognizing antibodies. We first compared enrichment of the J2 or K1 dsRNA IPs to the mock IgG immunoprecipitation in order to identify dsRNA-enriched regions within each condition. A correlation heatmap of read count data showed high similarity between the K1 and J2 antibodies within each condition (Fig S11), indicating that both antibodies identified a largely overlapping subset of dsRNAs. Therefore, we combined the J2 and K1 data to identify candidate dsRNA regions for subsequent analyses.

To identify dsRNAs which were induced by DUX4, we asked which dsRNA containing transcripts were differentially enriched between -DOX and +DOX conditions. Using an absolute log2 fold-change threshold of 4.0 and an FDR-adjusted p-value of 10^{-10}, 870 regions showed increased dsRNA enrichment in DUX4 expressing cells (see Fig S12 for chromosomal

locations) compared to 193 regions which showed decreased enrichment (Fig 15A and Supplementary file 1), confirming a general induction of dsRNA enriched regions following DUX4 expression in this cell line.

To verify that we successfully enriched for dsRNAs we examined dsRNA-specific RNA editing. The ADAR enzymes convert adenosine to inosine specifically in dsRNAs which manifests as A to G mismatching in RNA-seq reads [121]. We therefore examined mismatches in strand specific, ribosomal depleted, total RNA-seq dataset of MB135-iDUX4 myoblasts expressing DUX4. Within DUX4-induced dsRNA regions, the A-G mismatch type made up more than 60% of all 12 possible called mismatches compared to the ~22% frequency of A-G mismatches called in the entire genome-mapped reads (Fig 15B), confirming that the K1 and J2 antibodies successfully enriched for dsRNAs. We further validated that the A-G mismatches represented editing events rather than rare polymorphisms by using primer sets to amplify and sequence the genomic DNA and the cDNA in three regions within dsRNAs containing A-G mismatches. This analysis showed that the genomic DNA sequence of MB135-iDUX4 cell amplicons matched the reference genome, whereas the cDNA sequences contained the expected A-G mismatches (Fig 15C-D and Fig S13). Combined, the above results validate our dsRNA IP strategy and indicate that DUX4 expression leads to the induction of dsRNA transcripts which are subject to dsRNA-specific ADAR editing.

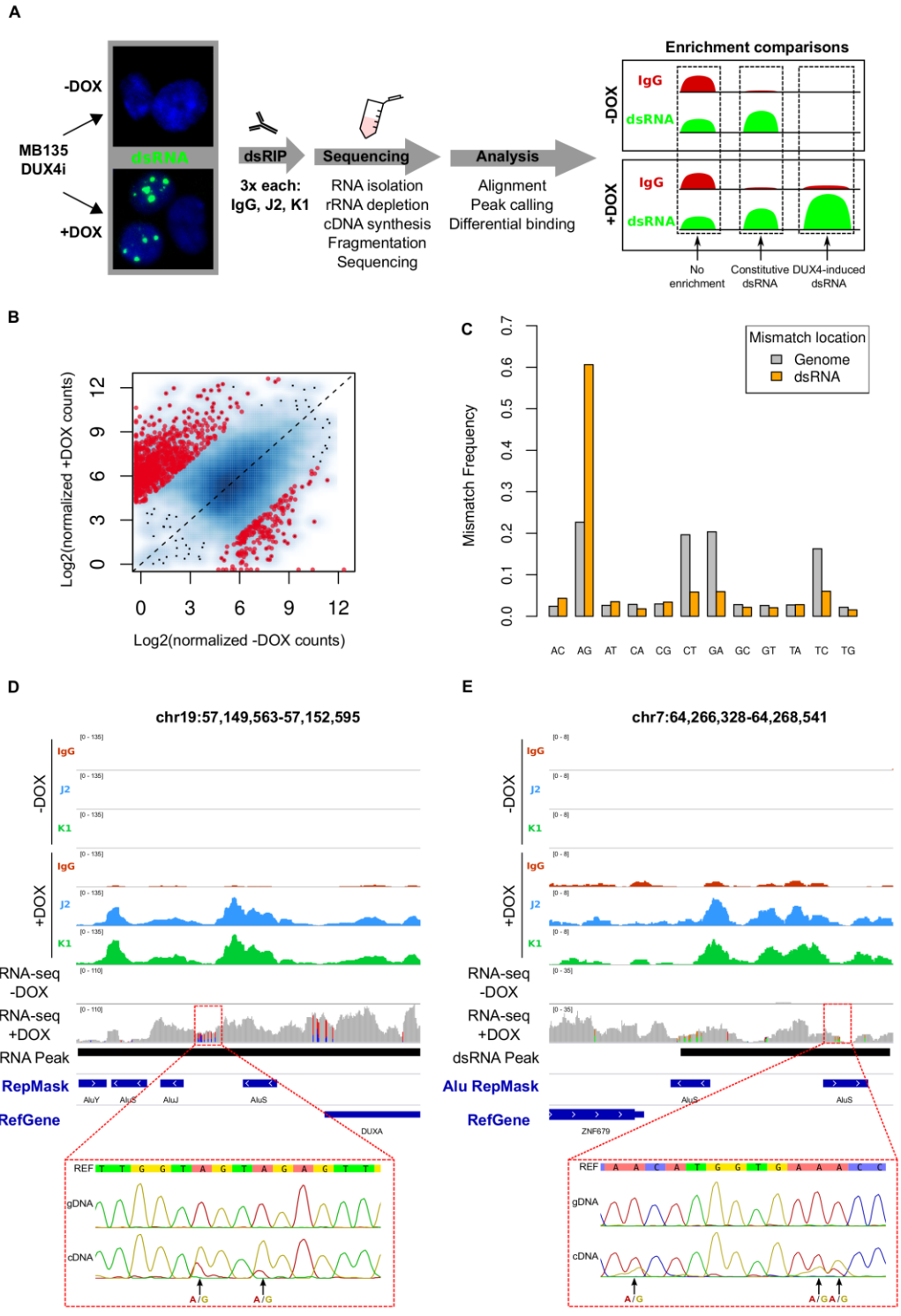


Figure 15 Double-stranded RNA immunoprecipitation and sequencing of DUX4-induced dsRNAs enriches for A-I edited transcripts. (A) Schematic of experimental outline for dsRNA immunoprecipitation and sequencing. Briefly, -/+ doxycycline (DOX) induced MB135-iDUX4 cells were lysed and subjected to immunoprecipitation of either matched antibody isotype

control (IgG) or J2 or K1 antibodies, performed in triplicate. Libraries were constructed from isolated RNA and informatics analysis consisted of broad peak calling followed by differential enrichment testing. Within each (-/+ DOX) condition, dsRNA "enrichment" was determined against the IgG control, which essentially represents total RNA abundance. Comparison of enrichment between conditions provides a way to measure level of induction following DUX4 expression. Three separate cultures of MB135-iDUX4 cells were used per condition. (B) Scatterplot comparing normalized levels of dsRNAs in +/- DOX conditions, depicted as the log2 counts plus a pseudocount value of 1.0. Red points highlight dsRNAs that show evidence for statistically significant differential enrichment and are limited to values with a threshold absolute log2 fold change of 4.0 and FDR-adjusted p-value of 1x10-10. (C) Mismatch frequencies of each of the 12 possible mismatch types within DUX4-induced dsRNAs in the +DOX stranded RNA-seq dataset within the indicated locations. (D-E) Browser screenshots in the indicated hg38 regions of normalized read counts for -/+ DOX immunoprecipitations. These tracks represent merged read counts for the indicated triplicate datasets. "RNA-seq -/+ DOX" refers to the stranded RNA-seq dataset and is shown as raw read counts such that mismatches (highlighted in non-grey colors) can be visualized. DUX4-induced dsRNA regions are depicted by black bars. Sanger sequencing results of amplicons designed across the annotated regions (red dashed lines) in either genomic DNA (gDNA) or complementary DNA (cDNA) of +DOX MB135-iDUX4 cells, with AG mismatches highlighted. Here, the reference strand is shown as the strand of origin for the transcripts, based on the stranded RNA-seq dataset.

4.3.2 *DUX4-induced double-stranded RNAs are enriched for non-coding intergenic RNAs that are activated by DUX4*

We next asked which annotated genomic features DUX4-induced dsRNAs overlap with. Constitutive dsRNAs, which were expressed in the presence or absence of DUX4 expression (see Fig 16A), recapitulated the known intronic enrichment as well as a slight 3-prime UTR enrichment for A-I edited dsRNA (Fig 16B and [116–118]). Conversely, DUX4-induced dsRNAs had a profile shifted largely towards intergenic regions, including regions 5 kb upstream and 5 kb downstream of annotated genes (Fig 16B), such as in the example shown upstream of *TP53BP2* (Fig 16C), or previous examples downstream of *DUXA* (Fig 15D) and *ZNF679* (Fig

15E). Other DUX4-induced intergenic dsRNA enriched regions were embedded within larger DUX4-induced intergenic transcripts that extended upwards of 500 kb in length (Fig 16D and Fig S14A-C).

Based on previously published ChIP-seq [15], many of these transcripts showed evidence for binding by DUX4 at endogenous retroviral LTR elements near the beginning of the transcript. Indeed, DUX4 bound significantly closer ($p = 6.83x10^{-10}$) to DUX4-induced dsRNA regions than constitutive dsRNA regions (Fig S15), suggesting a direct activation for many of the identified DUX4-induced dsRNAs.

Collectively, these data indicate that, in contrast to the largely intronic origin of dsRNAs in mammalian cells, a substantial portion of DUX4-induced dsRNAs are comprised of sequences embedded within non-coding intergenic DUX4-induced transcripts. At least some of these transcripts appear to be directly activated by DUX4 as they originate in close proximity to DUX4 binding sites.

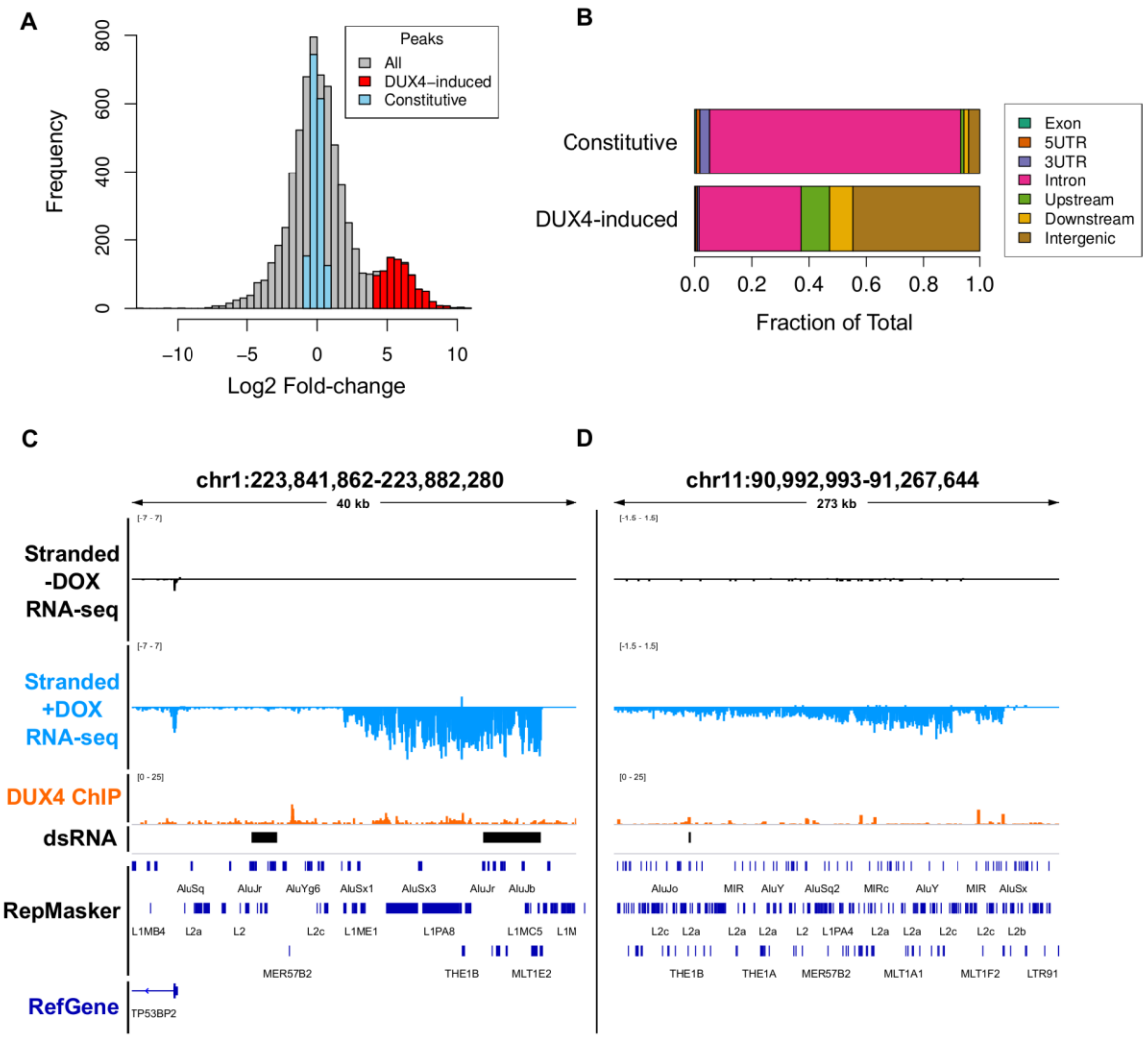


Figure 16 DUX4-induced double-stranded RNAs are enriched for non-coding intergenic RNAs. (A) Histogram of consensus “peaks” from DiffBind analysis of dsRIP-seq experiment (grey) with constitutive peaks (blue, |log2 fold change| < 1.0 and FDR-adjusted p-value > 0.05), and DUX4-increased (red, log2 fold change > 4.0 and FDR-adjusted p-value < 1x10-10). (B) Genomic feature distribution overlap of each dsRNA, defined as in Fig 16A. See Methods for feature definitions and prioritizations. (C-D) Browser screenshots in the indicated hg38 regions of normalized stranded ribominus MB135-iDUX4 RNA-seq data in two intergenic locations. These tracks represent merged read counts for the indicated duplicate datasets.

4.3.3 *DUX4-induced dsRNAs are enriched for repeat sequences including Alu, LINE-1, HERVL/MaLR and the pericentric HSATII repeat*

Endogenous transcripts that contain repetitive sequences are known to form dsRNAs in humans. For example, *Alu* SINE repeats are vastly overrepresented in human dsRNAs [116] where they often form inverted, hybridized pairs within a single RNA transcript [118,122]. LTR-containing ERVs have also been shown to form dsRNAs in tumor cells treated with 5-azacytidine. Because DUX4 activates transcription of repetitive elements, we therefore hypothesized that DUX4-induced dsRNAs were likely comprised of repetitive sequences.

To test this possibility, we first identified the set of all DUX4-induced repeat sequences using the Dfam database of repetitive elements [123] in our stranded total RNA sequencing dataset (Fig 17A and Supplementary file 2). This analysis confirmed the previously published observation that DUX4 activates expression of HSATII pericentric satellite repeats, MaLRs, ERVs and some LINE-1 repeats [15,16,29,31,92].

Next, we determined the set of all repeat subfamilies that were enriched by the K1 or J2 antibodies compared to the IgG control in DUX4 expressing myoblasts (Supplementary file 3). Intersecting all dsRNA enriched repeats with the DUX4-induced repeats identified the set of repeats that were both strongly induced by DUX4 and were also enriched for dsRNAs (red dots, Fig 17B-C and Fig S16A-B). These repeats therefore constitute repeat classes which can form dsRNA and that were highly upregulated by DUX4. As expected, *Alu* repeat subfamilies were greatly enriched in the dsRIP-seq datasets compared to the IgG control, though they were not specifically upregulated following DUX4 expression (Fig S17A). We observed a modest enrichment for LINE-1, ERVL-MaLR and HERVL repeats in the K1 and J2 RIPs (Fig S17B-C and Supplementary file 3), which also showed some increased expression by DUX4 (Fig S17A), indicating that these repeat transcripts may constitute a sizeable fraction of DUX4-induced dsRNAs.

The most highly induced repeat by DUX4 which was also enriched for dsRNA was the pericentric satellite repeat class HSATII (Fig 17A-C and Fig S16A-B). This suggested that HSATII-derived transcripts might form a considerable DUX4-specific fraction of dsRNAs. Compared to non-DUX4 expressing cells where 0.00-0.01% of reads mapped to HSATII, approximately 4.1% of K1 dsRIP-seq reads from DUX4 expressing myoblasts aligned to HSATII repeats. For unknown reasons, the J2 antibody showed more modest affinity than K1 for HSATII, though these two antibodies do have noted differences in dsRNA epitope preferences (compare Fig 17B and Fig S16A). It should also be noted that the dsRIP-seq with the control IgG also showed a high proportion of HSATII reads in the DUX4-induced samples (2.1% of the IgG RIP-seq reads), likely representing the high level of HSATII expression, where it can represent over 2% of the total ribosome depleted RNAs in DUX4-expressing cells. Nonetheless, the overall enrichment in the dsRIP-seq indicated that a large fraction of this highly abundant RNA likely represents DUX4-induced dsRNAs.

4.3.4 *DUX4 binds to and activates HSATII transcription*

Satellite repeats can be transcribed during certain cellular stresses such as heat shock [124]. However, our previously published ChIP-seq indicated that DUX4 can likely directly bind HSATII repeats [16,29], making it unlikely that HSATII activation is merely an indirect consequence of DUX4 expression. Furthermore, the consensus HSATII sequence contains close matches to the known DUX4 binding motif (Fig S18A) and variation within unmapped HSATII repeat sequences [16,125] creates the potential for large regions of arrayed near-perfect matches for DUX4 binding sites within pericentric regions. Indeed, Sanger sequencing of cloned HSATII PCR amplicons verified that these repeats often contain multiple copies of the consensus DUX4 binding motif (Fig S18A), suggesting that DUX4 directly binds to and activates transcription of HSATII.

To further test that DUX4 can directly activate HSATII transcription, we cloned a multicopy HSATII repeat that contained DUX4 binding motifs into the promoterless pGL3 luciferase reporter vector in both the forward and reverse orientation (Fig S18B). Co-transfection of HEK293T cells with this HSATII-luciferase reporter and a DUX4 expression vector or control GFP expression vector showed robust activation of luciferase RNA expression by DUX4 compared to co-transfection of the control vector (Fig 17D). Activation was dependent on the presence of the HSATII sequence as the parental reporter vector showed minimal activity in the presence of DUX4 (Fig 17D). This result provides additional evidence that the binding of DUX4 within HSATII repeats that was previously identified in the ChIP-seq studies correlates with its direct initiation of HSATII transcription.

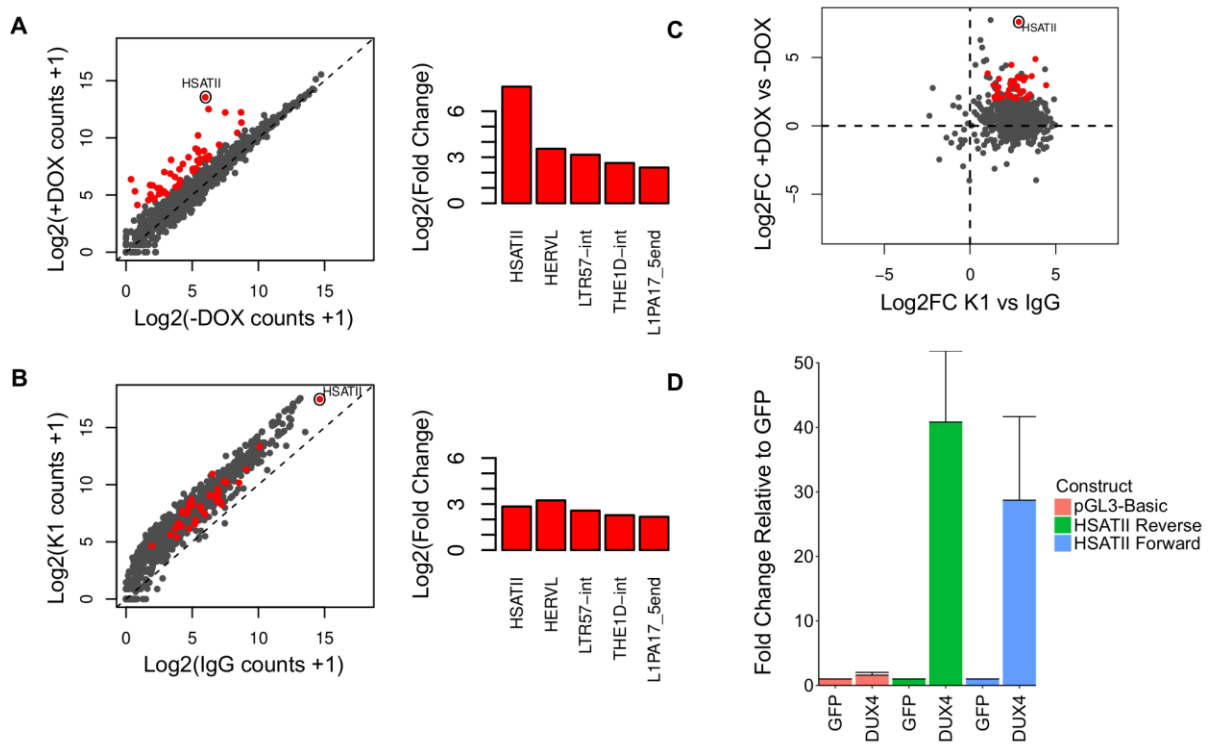


Figure 17 DUX4-induced dsRNAs are enriched for repeat sequences including HSATII (A) Scatterplot depicting log2 normalized read counts of Dfam predicted repeat class subfamilies within stranded RNA-seq dataset of cells -/+ DOX is shown on the left. Repeat subfamilies are

highlighted as red when |log2 fold change| > 2.0 and FDR-adjusted p-value < 0.01. DESeq2 moderated log2 fold change values of select DUX4-induced repeat subfamilies are shown on the right. (B) Scatterplot depicting log2 normalized read counts of Dfam predicted repeat class subfamilies within K1 vs IgG dsRIP-seq datasets in the +DOX condition. Significantly differentially expressed repeats (red dots) from (A) are considered enriched and highlighted as red in this plot if they meet the criteria of |log2 fold change| > 1.0 and FDR-adjusted p-value < 0.01. DESeq2 moderated log2 fold change values of select repeat subfamilies are shown on the right. (C) Scatterplot of Dfam predicted repeat class subfamilies showing DESeq2 moderated log2 fold change values (Log2FC) in +DOX vs -DOX stranded RNA-seq (y-axis) compared to moderated log2 fold change values in K1 vs IgG RNA immunoprecipitations in the +DOX condition (x-axis). Red points are highlighted as in panel (B). (D) RT-qPCR showing levels of luciferase expression normalized to RPL27 and depicted relative to GFP transfection control for the indicated pGL3 Basic luciferase construct. 293T cells were transfected with either GFP or DUX4 expression vectors and co-transfected with pGL3 Basic (no insert), HSATII Reverse or HSATII Forward vector. Here “forward” means that the HSATII amplicon cloned into the pGL3 Basic vector is in the consensus orientation relative to the luciferase sense strand. Error bars represent the standard deviation of the mean for three independent transfections except in the case of the pGL3-basic DUX4 transfection where one sample was omitted due to a technical failure which led to non-amplification.

4.3.5 *DUX4-induced nuclear dsRNA aggregates correlate with the accumulation of temporally controlled bidirectional HSATII transcripts*

Because HSATII repeats can be bidirectionally transcribed in the context of certain tumor cells [49], we considered that bidirectional transcription of HSATII may lead to dsRNA formation following DUX4 expression. Analysis of a DUX4-induced dsRNA locus which overlapped Dfam predicted HSATII repeats demonstrated that, although reads mapped predominantly to the reverse complement of the consensus HSATII sequence (hereafter referred to as “reverse” HSATII transcripts), there was some indication that bidirectional “forward” transcription also occurred (Fig 18A). For these reasons, we postulated that

bidirectional HSATII transcription might lead to the accumulation of intermolecularly formed HSATII dsRNA foci in DUX4 expressing cells.

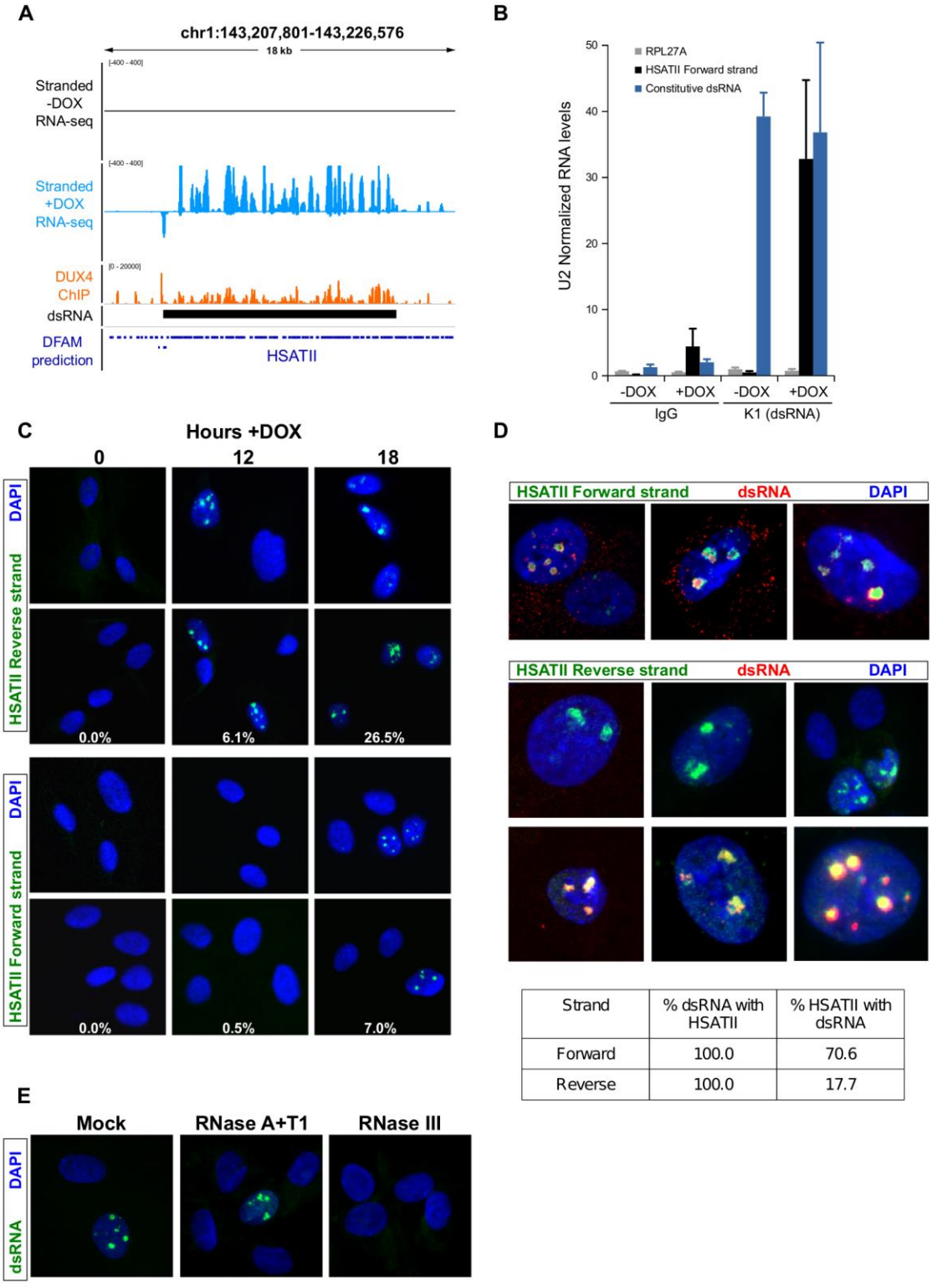
The somewhat modest enrichment of HSATII RNA using the K1 and J2 antibodies in our dsRIP-seq dataset could be a result of the less abundant forward strand acting as the limiting substrate for HSATII dsRNA formation. In agreement, strand-specific RT-qPCR following K1 dsRIP confirmed that the forward HSATII transcripts were highly enriched as dsRNA (Fig 18B), whereas the enrichment of HSATII in a non-strand-specific RT-qPCR was more modest (Fig S19), similar to our unstranded dsRIP-seq enrichment. We therefore conclude that DUX4-induced HSATII transcripts likely form dsRNA predominantly through pairing of their bidirectional transcripts, rather than through intramolecular interactions and that the less abundant forward strand acts as a limiting factor for dsRNA formation.

We next sought to investigate the dynamics of bidirectional HSATII transcription in DUX4 expressing cells. Thus, we performed RNA fluorescence in situ hybridization (RNA-FISH) with probes designed to detect the forward or reverse HSATII sequence [126] in a time course following treatment with doxycycline. This experiment revealed the appearance of intranuclear HSATII foci in DUX4 expressing cells with either forward or reverse probes by the 18-hour time point (Fig 18C), confirming the bidirectional expression of HSATII following DUX4 induction. Interestingly, the probe targeting the more predominant reverse HSATII transcript identified foci in a substantial percentage of nuclei at both 12 and 18 hours following DUX4-induction, whereas the probe to the less abundant forward transcript identified foci in a smaller number of nuclei and only at the 18-hour time point (Fig 18C).

Next, we tested whether DUX4-induced dsRNA aggregates form via hybridization of forward and reverse HSATII strands by performing immunofluorescence with the K1 antibody combined with RNA-FISH using HSATII probes at 18 hours following DUX4 expression. Consistent with the ability for bidirectional HSATII transcripts to form dsRNA, this experiment revealed that DUX4-induced nuclear dsRNA aggregates nearly perfectly coincided with both

forward and reverse HSATII RNA in DUX4 expressing cells (Fig 18D). Indeed, 100% of the observed K1 stained dsRNA foci were associated with forward and reverse HSATII FISH foci. These HSATII overlapping K1 dsRNA foci were confirmed as double stranded because they were not evident following treatment with the double-stranded ribonuclease, RNase III, but were still evident following treatment with the single stranded RNases A and T1 (Fig 18E). As reverse HSATII foci formation preceded forward HSATII foci formation, and because we hypothesized that both HSATII strands were required for forming dsRNA foci, we predicted that reverse HSATII foci would not always be coincident with dsRNA. Indeed, a minority (~17.7%) of nuclei with reverse HSATII foci coincided with dsRNA aggregates whereas a majority (~70.6%) of nuclei with forward HSATII foci coincided with dsRNA.

Combined, our results indicate that DUX4 expression appears to directly activate the transcription of the reverse HSATII strand at an early time point and the forward strand at a later time point. Upon the activation of the late-forming forward HSATII strand, DUX4 expressing cells form dsRNA aggregates. Because all K1 dsRNA foci detectably contained both forward and reverse HSATII RNA strands, our data show that DUX4-induced nuclear dsRNA aggregates were formed via bidirectional transcription of HSATII repeats, with the less abundant forward HSATII strand acting as a limiting factor for dsRNA formation.



Strand	% dsRNA with HSATII	% HSATII with dsRNA
Forward	100.0	70.6
Reverse	100.0	17.7

Figure 18 DUX4-induced HSATII transcripts are bidirectionally transcribed and form dsRNA. (A) Browser screenshot in the indicated hg38 region of normalized stranded ribominus

MB135-iDUX4 RNA-seq data. Shown at the bottom are Dfam predicted HSATII repeat locations. (B) RT-qPCR of RNAs immunoprecipitations using K1 or IgG as control in MB135-iDUX4 cells -/+ DOX, as indicated. Primers were designed to detect a constitutive dsRNA from our dsRIP-seq dataset or forward HSATII after strand-specific RT-qPCR. IP values were calculated as percentage of +DOX input and normalized to background U2 RNA levels to account for differences in total -/+DOX RNA abundances. Error bars represent the standard deviation of the mean for three independent immunoprecipitations. (C) Confocal images of RNA fluorescence in situ hybridization of MB135-iDUX4 cells treated with doxycycline for the indicated times and hybridized with probes that detect the indicated HSATII strands, where forward and reverse is relative to the consensus sequence. Images are representative from two independent time-course experiments conducted on separate days. Estimates for percent of foci positive nuclei are indicated for each time point and are from > 150 random nuclei counted per time point. (D) Combined immunofluorescence in MB135-iDUX4 cells at 18 hours +DOX, using the K1 dsRNA antibody and RNA-FISH with probe detecting forward (upper) or reverse (lower) HSATII transcripts. Images are representative from two independent, combined IF RNA-FISH experiments conducted on separate days. Note, with the reverse probe a larger subset of DUX4 expressing cells contained HSATII foci, but no dsRNA foci, whereas another class contained HSATII foci which co-localized with dsRNA foci. Estimates for percentages of nuclear foci co-occurrence of the indicated HSATII strand and K1 antibody is shown in table below and are from counts of > 250 random nuclei per time point. (E) K1 immunofluorescence in in MB135-iDUX4 cells at 18 hours +DOX treated with the indicated RNase enzyme or mock treated prior to the immunofluorescence. We did not detect any dsRNA foci after treatment with RNase III, a double-stranded RNase. Experiment was performed twice.

4.3.6 *HSATII transcription is associated with EIF4A3 and ADAR1 aggregation and correlates with the formation of intranuclear BMI1 foci*

We next sought to determine which nuclear proteins may aggregate as a consequence of HSATII-containing dsRNA foci formation in DUX4 expressing cells. We previously showed that the exon junction complex factor EIF4A3 accumulates in intranuclear foci following DUX4 induction [113]. In this prior study, EIF4A3 aggregation appeared to precede dsRNA nuclear foci formation, but also nearly perfectly co-localized with DUX4-induced dsRNA upon their

appearance. We therefore performed combined immunofluorescence and RNA-FISH of DUX4-induced cells with an antibody to EIF4A3 and the FISH probe to the HSATII reverse strand. This showed strong co-localization of EIF4A3 and HSATII foci (Fig S20), indicating that EIF4A3 accumulates with reverse HSATII transcripts prior to the formation of dsRNA.

Although we cannot directly assess ADAR editing of HSATII repeats due to polymorphic variations from the consensus sequence [16], FISH combined with an antibody to ADAR1 indicated a strong redistribution of ADAR1, which binds dsRNA, to HSATII foci in DUX4 expressing cells (Fig S21A). Nearly 100% of these ADAR1 foci also co-localized with K1 dsRNA nuclear staining (Fig S21B), suggesting that DUX4-induced HSATII dsRNA is bound by ADAR1 and leads ADAR1 accumulation as intranuclear foci.

We also wondered about other consequences of HSATII dsRNA formation in DUX4 expressing cells. Derepression of HSATII repeats in cancer cells has been associated with the formation of polycomb bodies [126,127]. In our model system, DUX4 induction of HSATII repeats was similarly associated with the formation of polycomb bodies as determined by immunodetection of the PRC1 complex protein BMI1 (Fig S22A). These BMI1 foci did not tend to co-localize with induced HSATII RNA foci (Fig S22B), although it is possible that they might be associated with demethylated HSATII DNA regions in the genome as has been suggested [126].

4.3.7 *Endogenous DUX4 in FSHD muscle cells induces dsRNA-forming transcripts including HSATII*

In cultures of FSHD skeletal muscle cells, *DUX4* is expressed in only a minor fraction of the muscle nuclei [14]. To determine whether the expression of endogenous *DUX4* in FSHD muscle resulted in transcriptional upregulation of the identified dsRNA regions, we used RT-qPCR and found expression of genic as well as intergenic RNAs in the DUX4-induced dsRNA regions in both FSHD1 (MB073) and FSHD2 (MB200) cultured muscle cells, but not in control (MB135)

muscle cells (Fig S23). Sequencing cDNA amplicons from differentiated FSHD2 cells identified A-G mismatches compared to the gDNA, indicating ADAR editing and dsRNA formation of these transcripts in FSHD myotubes (Fig S24).

We also found elevated HSATII expression in differentiated FSHD myotubes compared to a control non-FSHD cell line (Fig 19A). Consistent with our previous demonstration that a small proportion of FSHD muscle cells formed intranuclear dsRNA foci [113], RNA-FISH revealed rare intranuclear foci of HSATII using the probe to the reverse strand in FSHD myotubes, but not in control cells (Fig 19B). Therefore, as with ectopic expression of DUX4 in MB135-iDUX4 cells, endogenous DUX4 expression in FSHD myotubes also led to dsRNA expression and intranuclear foci of HSATII RNA.

4.3.8 *DUX4-induced dsRNA-forming RNAs including HSATII are expressed in pre-implantation embryos*

Recent studies indicated that DUX4 and its mouse ortholog Dux activate a subset of the early cleavage stage zygotic gene transcription program in human and mouse embryos [29–31]. In humans ZGA involves the expression of HSATII and HERVL retrotransposons and in mice involves the expression of MERVL and GSAT major pericentric satellite repeats. To determine whether cleavage stage expression of *DUX4* correlated with dsRNA formation, we analyzed published early embryo RNA-seq data [29] across several different stages of development. As previously reported, we identified *DUX4* expression at the cleavage stage of embryonic development and this correlated with increased expression of many of the DUX4-induced dsRNA-forming regions, including HSATII (Fig 19C and Supplementary file 4). Interestingly, both forward and reverse strand transcripts of HSATII were at near equal levels in the cleavage stage (Fig S25), perhaps because the samples contained a heterogeneous pool of embryos at different cleavage cell numbers. Analysis of base mismatches between the RNA-seq and the reference genome sequence revealed a dramatic enrichment of A-G mismatches over other

mismatch types across DUX4-induced dsRNA regions (Fig 19D). Thus, our analysis verified that many of the DUX-induced dsRNA-containing transcripts identified in DUX4 expressing myoblasts, including HSATII, are also present in the early embryo.

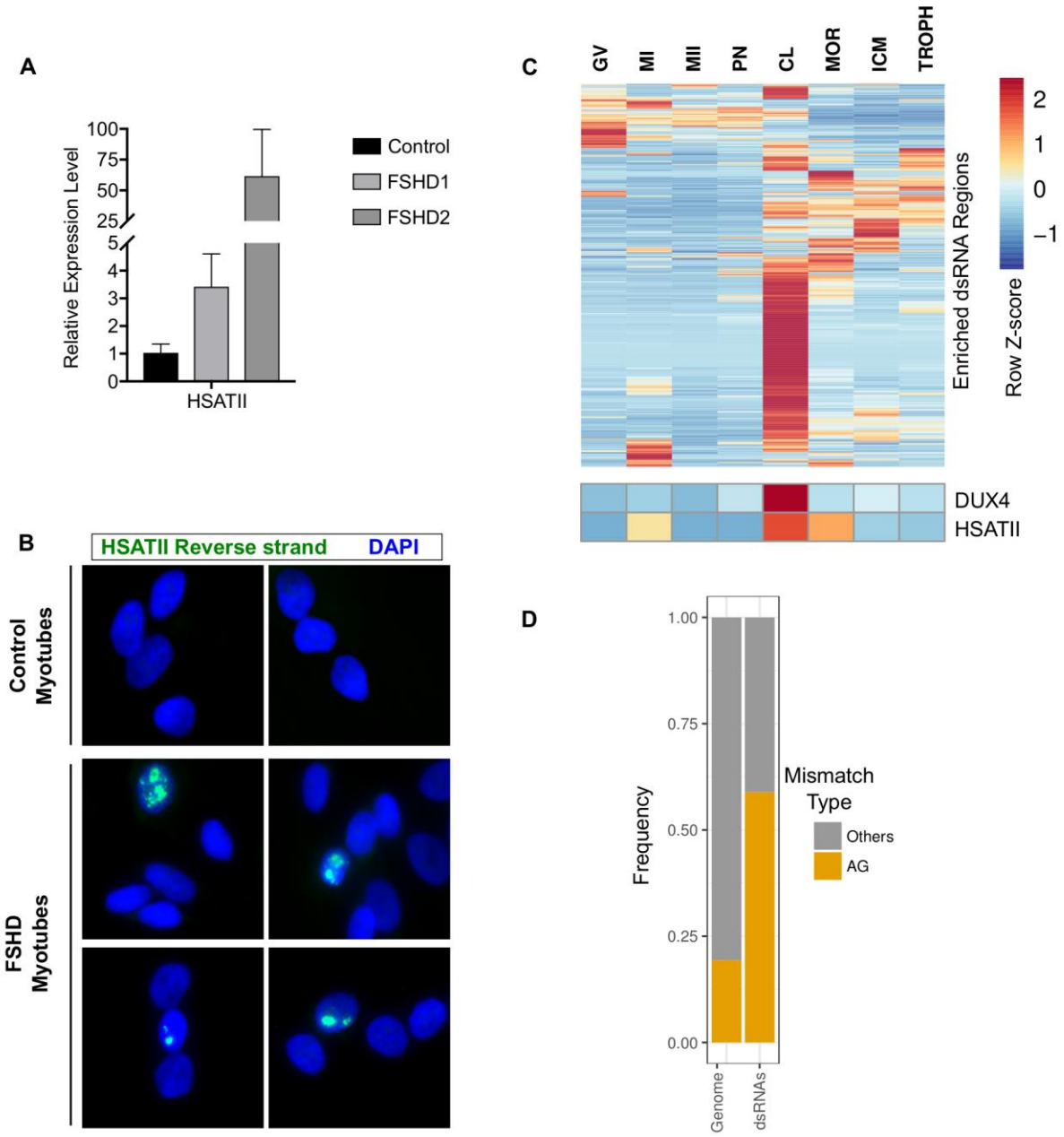


Figure 19 Endogenously expressed DUX4 induces dsRNA forming transcripts, including HSATII. (A) RT-qPCR data showing transcript expression levels of HSATII in control (MB135), FSHD1 (MB073) or FSHD2 (MB200) cells grown in differentiation medium. Data are normalized

to RPL27 levels and shown relative to the control cell line. Data are depicted as the mean values of three experiments performed on independent days. Error bars represent the standard deviation of the mean. (B) RNA-FISH using probes targeting the reverse HSATII transcript in control (MB135) or FSHD (MB073) differentiated myotubes. Consistent with rare DUX4 expression in FSHD muscle cells, we observed a small subset of HSATII-positive nuclei in the FSHD cells, but not control cells. Images are representative from two independent experiments conducted on separate days. (C) Heatmap of DUX4-induced dsRNA, DUX4 and HSATII expression across various stages of human embryo development. Read counts were normalized and depicted as the within row Z-score. Data from Hendrickson, et al 2017. GV (germinal vesicle), MI (metaphase I), MII (metaphase II), PN (pronuclear stage), CL (cleavage stage), MOR (morula) and ICM (inner cell mass), TROPH (trophectoderm). (D) Frequency plot of AG compared to all other possible mismatches in cleavage stage samples from Hendrickson, et al 2017. Whole genome refers to all mismatches called, regardless of location and dsRNAs are mismatches called within DUX4-induced dsRNA regions.

4.4 *Discussion*

When DUX4 is expressed in induced pluripotent stem cells or skeletal muscle cells, it activates genes characteristic of a totipotent developmental expression program [29–31]. For example, DUX4 activates the expression of *ZSCAN4* which has roles both in telomere elongation [128] and DNA demethylation through the degradation of UHRF1 and DNMT1 [129], as well as *KDM4E*, which is a lysine demethylase that can relieve chromatin repression [130]. DUX4 also activates the expression of repetitive elements, including LINE-1, HERVL and MaLR endogenous retroviruses, and HSATII satellite repeats [15,16,29,31,92]. In this study, we provide evidence that DUX4 creates intergenically derived dsRNAs from a subset of these repeats and that intergenic HSATII transcripts coincide with nuclear dsRNA aggregates. Although we used forced expression of DUX4 in skeletal muscle cells to identify these dsRNA-forming transcripts, analysis of RNA and RNA-seq showed that these dsRNA-forming regions were also transcribed and ADAR-edited in the endogenous, DUX4 expressing FSHD muscle cells and cleavage stage embryos. These findings solidify a mechanism of dsRNA formation in

FSHD that contributes to cell toxicity through the activation of the PKR innate immune response pathway, and further raise the possibility that DUX4-induced dsRNA-forming repeat transcripts, such as HSATII, might have a functional role in cleavage stage or stem cell biology.

In mice, during the cleavage stage of early embryogenesis, pericentric “major satellite” repeats are expressed first in the forward, then reverse direction (relative to the consensus sequence) where they have been hypothesized to form dsRNAs that direct repressive heterochromatin modifications to the pericentromeres [35,50]. However, the mechanism of regulating these transcripts remained unknown. Gapmer-mediated depletion of these repeat transcripts blocked chromocenter formation and interrupted development at the 4-cell stage indicating a necessary role for satellite RNAs in establishing pericentric heterochromatin [35]. While transcription of satellite repeats appears to be necessary for heterochromatin establishment, whether dsRNA is required is less clear [131]. As with DUX4 in human embryogenesis, mouse Dux is expressed in the cleavage stage and, interestingly, coincident with transcription of GSAT pericentric repeats [29,31]. In the case of HSATII, our study shows that DUX4 appears to first initiate HSATII transcription in predominantly one direction, producing an RNA encoding the reverse complement of the consensus HSATII sequence, likely through direct targeting, and then later induces HSATII transcription in the opposite direction coincident with the formation of multiple distinct intranuclear dsRNA foci. Given that this bidirectional expression pattern of HSATII by DUX4 parallels the temporal dynamics of mouse expression of GSAT repeats, we postulate that DUX4, and by analogy mouse Dux, has a primary role in regulating the bidirectional transcription of these major satellite repeats in early human and mouse embryogenesis.

In contrast to its potential function in the early embryo, the expression of the HSATII RNA in FSHD could have both direct and indirect biological consequences that could alter RNA processing and chromatin repression. Previously we reported that DUX4 expression induces intranuclear focal aggregates of EIF4A3 [113], and here we showed that EIF4A3 accumulated

with nuclear HSATII RNA foci. Although it remains unclear whether the association of EIF4A3 with HSATII RNA is due to splicing of the HSATII transcripts or another recruitment mechanism, the formation of EIF4A3 foci correlates with the inhibition of nonsense mediated RNA decay (NMD). This suggests that the sequestration of EIF4A3 might contribute to DUX4-mediated inhibition of NMD, and could act in concert with the DUX4-mediated depletion of UPF1 [42,113]. Similar to the association of EIF4A3 with the predominantly expressed reverse HSATII strand, the association of ADAR1 foci with HSATII dsRNA could also contribute to toxicity in DUX4-expressing cells. For example, ADAR1 aggregation might limit the enzyme's activity on non-HSATII containing endogenous dsRNAs, which would be expected to lead to activation of cellular antiviral pathways. In accordance with this, *ADAR1* knockout leads to dsRNA-associated apoptosis in a human cell line that can be rescued by simultaneous knockout of *RNASEL* [132], paralleling our previous finding that DUX4 toxicity can be alleviated by knockdown of *RNASEL* [113].

DUX4-mediated induction of HSATII RNA might have other consequences as well. HSATII transcription has been associated with the formation of intranuclear foci of components of the PRC1 complex (polycomb bodies) at the HSATII loci, which can potentially lead to de-repression of LINE-1 elements or other repeats [126]. The DUX4-induced formation of polycomb bodies, as indicated by the BMI1 foci, are consistent with a model where HSATII transcription in DUX4 expressing cells might have a similar role in the formation of polycomb bodies and activation of LINE-1 repeats. In mice, expression of pericentric major satellite repeats can lead to genome instability, likely by destabilizing DNA replication forks [52]. This causes activation of the DNA damage response and cell death which can be alleviated by knockout of *p53*. Whether HSATII expression causes similar genome instability effects is currently unknown, but its demonstrated activation in tumor cell lines [49,133] suggests that its expression may similarly trigger genome instability.

A possible function for non-HSATII intergenic dsRNAs induced by DUX4 remains more speculative. It is interesting to note the relative enrichment of LINE-1 and ERV repeats within many of the intergenic DUX4-induced dsRNAs, classes of repeats that are expressed in the cleavage stage embryo, but generally repressed in most tissues [134]. Similar to the proposed role of HSATII repeats, the double-stranded secondary structure of these transcripts might also facilitate heterochromatin formation. This has been postulated as a function of dsRNA-forming lncRNAs, which can recruit methyltransferase activity to DNA [135]. Interestingly as well, the DUX4-induced long intergenic transcripts often contain a region of dsRNA enrichment and are reminiscent of the previously characterized “very long intergenic non-coding” (vlinc) RNAs observed in a multitude of pluripotent cells and also cancer lines [136,137], though the nature of the transcriptional regulation of these vlincRNAs is presently not well understood.

The enrichment of retroelement-derived dsRNAs in a subset of DUX4 target genes might also serve a function. Although retroelements are typically subject to DNA methylation and other silencing mechanisms [138], their presence can have a positive effect on gene expression at early stages of development [33,139,140]. Silencing of these retroelements at later stages might then provide precise temporal control of DUX4 targets whose continued expression would counteract later differentiation processes. It is possible, therefore, that the dsRNAs in DUX4-induced genes might contribute to the subsequent silencing of these genes through regional dsRNA-mediated heterochromatin formation.

In summary, our work confirms the induction of dsRNA containing transcripts by DUX4. Similar to constitutively expressed dsRNAs, many of the DUX4-induced dsRNAs are enriched for inverted *Alu* repeats in genic transcripts. In contrast to constitutive dsRNAs, a large fraction of DUX4-induced dsRNAs are intergenic and are modestly enriched for LTR-containing ERVs and LINE-1 elements, but highly enriched for HSATII satellite repeats. DUX4 induction of HSATII repeats in a bidirectional, temporally controlled pattern mirrors the dynamics of mouse GSAT repeat expression in the early embryo and suggests that human DUX4 activation of

HSATII repeat transcripts in the cleavage stage embryo plays a similar role in establishing pericentric heterochromatin via dsRNA formation. Based on our previous work showing DUX4-induced dsRNA toxicity in FSHD skeletal muscle cells (Shadle, et al 2017), it is possible that HSATII dsRNAs are a major contributor to DUX4 toxicity in somatic cells, and that preventing HSATII dsRNA production might be a new therapeutic avenue for FSHD. The expression of DUX4 and bidirectional HSATII transcripts in the early embryo suggests a role for these dsRNAS in early development that merits further study.

4.5 *Materials and Methods*

Accession codes

MB135-iDUX4 dsRIP-seq and total stranded RNA-seq data were deposited into the Gene Expression Omnibus (GEO) under the SuperSeries accession number GSE114940. DUX4 ChIP-seq data used in this study are available from GEO under accession number GSE33838. Early embryo RNA-seq data were obtained from GEO accession number GSE72379.

Cell culture

De-identified human primary myoblast cell lines from the Fields Center for FSHD Neuromuscular Research at the University of Rochester Medical Center were immortalized by retroviral transduction of CDK4 and hTERT [141]. All immortalized human myoblasts were cultured in F10 medium (Gibco/ThermoFisher Scientific) supplemented with 20% FBS (GE Healthcare Life Sciences) and 1% penicillin/streptomycin (Gibco/ThermoFisher Scientific), 10.0 ng/ml recombinant human FGF (Promega) and 1.0 μM dexamethasone (Sigma-Aldrich). Myoblasts were differentiated by culturing in DMEM (Gibco/ThermoFisher Scientific) containing 1% horse serum (Gibco/ThermoFisher Scientific), 1% penicillin/streptomycin (Gibco/ThermoFisher Scientific), 10 μg/ml insulin and 10 μg/ml transferrin for 48-72 hours. All

"+DOX" labels mean that cells were grown for approximately 18 hours in the presence of 1.0 µg/ml of doxycycline hyclate (Sigma-Aldrich), unless otherwise noted.

dsRNA Immunoprecipitation

Human myoblast MB135-iDUX4 cells were treated -/+DOX and washed in PBS, trypsinized and counted prior to lysis. Approximately $1.2x10^6$ cells were used for each IP. Lysis was performed in 1.0 ml total volume by sonication using a Diagenode Bioruptor on light setting (5 minutes total, 30s on/off at 4°C) in a buffer composed of 15 mM Tris pH 7.5, 0.1 M NaCl, 5 mM $MgCl_2$, 0.5% Triton X-100, 1 mM dithiothreitol, and 40 U/ml RNase inhibitor (ThermoFisher Scientific). Lysates were precleared using 40.0 µl of protein G Dynabeads (ThermoFisher Scientific) for 1 hour prior to an overnight incubation at 4°C with either J2, K1 or an isotype-matched anti-GFP (IgG) control antibody. 40.0 µl of protein G Dynabeads were added the following morning for 1 hour to bind the antibody, and beads were subsequently washed four times with 1.0 ml of cold lysis buffer. After the final wash, 1.0 ml of TRIzol (ThermoFisher Scientific) was added directly to the beads for RNA extraction.

RNA extraction and library preparation

TRIzol RNA extractions were performed as per manufacturer's recommendations. Ribosomal RNA depletion was performed using the NEBNext rRNA depletion kit (New England Biolabs). First strand synthesis was achieved using SuperScript III reverse transcriptase (ThermoFisher Scientific) with random hexamers and using the following thermocycler conditions: 25°C 10min, 50°C 30min, 55°C 30min and 85°C 5min. Second strand synthesis was achieved using the NEB second strand cDNA synthesis kit, following manufacturer's recommendations. The double-stranded cDNA was fragmented to approximately 250 bp average size using a Diagenode Bioruptor sonicator in 120 µl total volume for three 10-minute intervals on medium intensity at 4°C, 20s on/off, replenishing with pre-cooled water between each interval. Libraries were made

from the AmpureXP (Beckman Coulter) purified cDNA using the Ovation Ultralow system V2 kit and following the manufacturer's instructions. Stranded total RNA-seq libraries were made from 1.0 µg of MB135-iDUX4 RNA using the KAPA stranded RNA-seq kit with RiboErase (HMR) and following the manufacturer's instructions. Libraries were sequenced using 100 bp single-end sequencing on the Illumina HiSeq 2500 platform by the FHCRC Genomics facility.

dsRIP sequence alignment and differential peak calling

We essentially treated our dsRIP-seq data as diffuse (e.g. histone mark) ChIP-seq to find locations of enrichment. First, dsRNA-IP reads were aligned to hg38 using BWA version 0.7.12 with the option -n 6. Next, peaks were called by comparison to the IgG background control condition using MACS2 version 2.1.0 using the options: --broad, --broad-cutoff 0.01, -q 0.01. We then used Diffbind version 2.3.8 to differentially call dsRNA enrichment between +/- DOX conditions with "minOverlap" set at 0.5, which treated J2 and K1 antibodies as replicates for comparison purposes. The significance threshold was set at 1e-10. We extracted the consensus peakset from the Diffbind dba report, setting the significance threshold as 1 to capture all dsRNA-enriched loci. We filtered our list of enriched dsRNAs to eliminate dsRNAs which overlapped known rRNAs, tRNAs and snRNAs (obtained from UCSC table browser repeatMasker track filter repclass: "snRNA OR rRNA OR tRNA OR snoRNA") using bedtools "intersect". We used the Integrative Genomics Viewer (IGV) version 2.3.98 for data visualization. We used the "kpPlotRegions" function from the karyoploteR package to plot location of DUX4-induced dsRNA-enriched regions.

Feature annotation

We used bedtools "annotate" and GenCode v24 feature annotations downloaded from the UCSC genome browser to classify per-base feature overlaps of dsRNAs with "intergenic" being defined as the absence of a feature overlap and "downstream" and "upstream" being defined as

5 kb 3-prime or 5-prime of an annotated gene feature, respectively. Overlapping features were prioritized as: coding exons > 5 UTR > 3 UTR > intron > upstream 5 kb > downstream 5 kb > intergenic.

Stranded RNA-seq

Stranded RNA-seq libraries were aligned using TopHat version 2.1.1 with Bowtie version 2.2.9 and the iGenomes UCSC gene transfer format file with the following options: --read-mismatches 8 --read-edit-dist 8.

RNA editing detection

We used the python script, REDItools *denovo* version 1.0.4 (Picardi and Pesole, 2013) to call editing sites on our aligned samples. Replicate BAM files were merged using samtools before analysis. The following REDItools parameters were used to limit false positives: -E (Exclude positions with multiple changes), -a t (two-tailed Fisher's exact test), -c 10 (minimum read coverage of 10), -T 6-6 (6 bases were trimmed from each end of the reads), -q 25 (minimum quality score of 25), -n 0.1 (minimum editing frequency of 0.1) and -v 3 (minimum number of reads supporting the variation set at 3). The SNP147 GTF file was downloaded from the UCSC genome table browser to exclude known polymorphic sites from consideration. We used a p-value threshold value of < 0.1.

Repeat analysis

For repeat counting in our stranded, ribosome-depleted RNA-seq dataset we used the TEtranscripts script from TEtoolkit package [142] with the multi-mode option and a custom GTF file which was made from the current (07-Nov-2016) release of Dfam non-redundant repeat matches. This approach enabled counting within predicted HSATII repeats (which are not available in the hg38 repeatMasker annotation used). We performed differential expression

analysis on the resulting count table with DESeq2 using default options. For differential expression of repeats in our dsRIP-Seq data, we performed the above analysis in K1 or J2 versus IgG in doxycycline-treated samples.

Confirmation of A-I editing

We amplified dsRNA regions containing putative A-I RNA editing sites using Phusion High Fidelity DNA polymerase (ThermoFisher Scientific) from genomic DNA or cDNA of doxycycline-induced MB135-iDUX4 cells. Purified amplicons were Sanger sequenced by the FHCRC genomics facility on a 3730xl DNA Analyzer and visualized using Geneious Pro software, version 5.0.4.

RNA isolation and real time qPCR

For immunoprecipitation samples, RNA was isolated using TRIzol reagent (ThermoFisher Scientific) according to the manufacturer's protocol and using GlycoBlue (ThermoFisher Scientific) as a co-precipitant. For all other samples, RNA was isolated with the RNeasy kit (QIAGEN), according to the manufacturer's protocol. Purified RNA was treated with DNaseI (ThermoFisher Scientific), heat inactivated, and reverse transcribed into cDNA using Superscript III (ThermoFisher Scientific) following the manufacturer's instructions. Quantitative PCR was performed with SYBR green reagent (ThermoFisher Scientific). For RT-qPCR of RNA immunoprecipitation, random hexamers (ThermoFisher Scientific) were used for non-strand specific detection. Data were calculated relative to +DOX input and normalized to U2 background RNA values for each independent immunoprecipitation. For HSATII forward strand specific RT-qPCR, we used the 'SATII_pericentric_1L' oligo to prime cDNA synthesis with a 5-prime linker ATGGATCACGAGAACACTGA sequence. For qPCR we used the 'SATII_pericentric_1R' and linker sequence as the primers. Samples were subsequently normalized to U2 signal in the respective random hexamer primed reactions, as above.

HSATII RNA-FISH and combined immunofluorescence RNA-FISH

Locked nucleic acid, FITC conjugated HSATII probes were purchased from QIAGEN and are based on the sequence used in previous publications [126,143,144]. Probe 1: 5' FAM-ATTCCATTCAGATTCCATTCGATC detects the reverse HSATII transcript. Probe 2, which detects the forward HSATII transcript is the reverse complement of probe 1 with 5' FAM modification. We performed HSATII RNA-FISH essentially as described previously [126] with modifications. Briefly, cells grown on coverslips were rinsed in cold CSK buffer (100 mM NaCl, 300 mM Sucrose, 3mM $MgCl_2$, 10mM Pipes pH 6.8) and then permeabilized in ice cold extraction buffer (CSK, 0.1% TX-100 and 10 mM Ribonucleoside Vanadyl Complex (NEB)) for 5 minutes. After a second CSK rinse, cells were fixed in 4% paraformaldehyde (Electron Microscopy Sciences) in PBS for 10 minutes at room temperature. Cells were dehydrated in series of 70% ethanol, 90% ethanol and 100% ethanol and then air dried. Probes were diluted to 5.0 pmol/ml and denatured at 75°C for 5 minutes in WCP buffer (50% formamide, 2XSSC, 10% dextran sulfate) and hybridized overnight at 37°C. Coverslips were washed with 50% formamide/2xSSC at 37°C and then twice with 2xSSC/0.1% TX100 for 5 minutes before mounting using prolong gold antifade with DAPI (ThermoFisher Scientific).

For combined immunofluorescence and RNA-FISH we first fixed cells on coverslips using 4% paraformaldehyde for 7 minutes at room temperature. Cells were permeabilized with 0.5% TX100 in PBS for 5 minutes prior to overnight incubation with primary antibodies at 4°C. Samples were incubated with appropriate secondary antibody for 1 hour at room temperature prior to fixation of the antibody interaction via a second crosslinking step using 4% paraformaldehyde for 10 minutes at room temperature. Cells were incubated with probes which were denatured as above for 2 hours at 37°C prior to sequential washes with 15% formamide 2xSSC for 20 minutes at 37°C, 2xSSC for 20 minutes at 37°C and 2xSSC for 5 minutes at room temperature. Cells were imaged using a Leica TCS SP5 II confocal microscope, where

indicated, or a Zeiss AxioPhot. Image channel merging and processing was performed using ImageJ software.

Analysis of human embryo RNA-seq data

We aligned raw reads to genome build hg38 using TopHat2, allowing two mismatches per read and maximal twenty multiple alignments. The counts for enriched dsRNA regions were computed by summarizeOverlaps() from the GenomicAlignments Bioconductor package with "IntersectStrict" mode. Next, we applied the countRepeats R package (https://github.com/TapscottLab/countRepeats), which is designed specifically for counting reads of repetitive elements originating from forward/reverse transcripts. To obtain DUX4 counts, we aligned the raw reads to the customized D4Z4 “genome” comprised of DUX4 exon sequences.

Cloning of HSATII and testing DUX4 activation

A ~1.5 kb PCR product was cloned from MB135 cell genomic DNA using the “SATII pericentric” primers which were synthesized with flanking NheI and HindIII restriction sites such that the amplicon could be readily ligated into the pGL3 basic vector. The opposite orientation clone was made using the original vector as template by PCR with the flanking restriction sites swapped to reverse the orientation of the HSATII insert in the same position. We tested for luciferase expression by RT-qPCR using SYBR green reagent and the comparative C_T method.

Statistical analysis

No statistical methods were used to predetermine sample sizes. Masking or blinding was not used during group allocation, data collection or data analysis.

Antibodies

K1 anti-dsRNA, SCICONS English & Scientific Consulting; K1 anti-dsRNA, SCICONS English & Scientific Consulting; ab180573 anti-EIF4A3, Abcam; anti-hADAR, polyclonal rabbit antibody obtained from B. Bass; ab126783 anti-BMI1, Abcam.

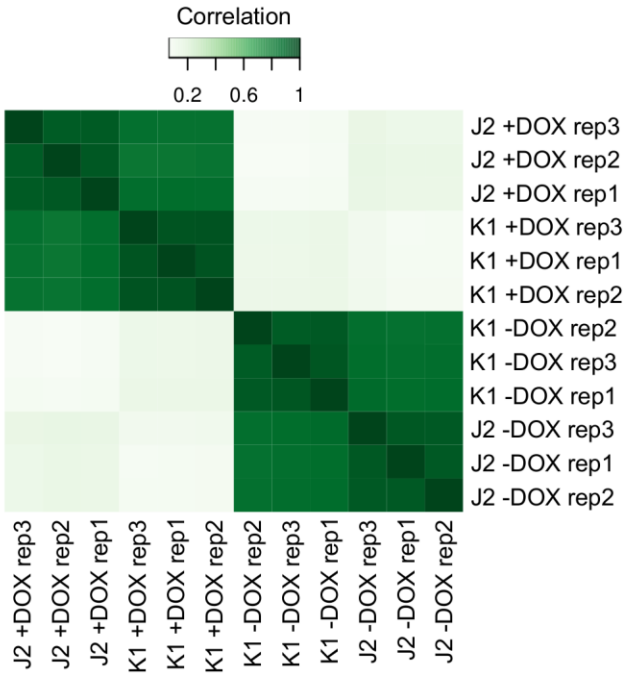
Primers

A-G_conf1_chr19_1_L	AACCAACCTTCATCCCAGTG
A-G_conf1_chr19_1_R	CACAAAAACTGCAGCCACAT
A-G_conf1_chr7_1_L	TCACACCTTATGGCACAAGAA
A-G_conf1_chr7_1_R	GCATCTGCTTCAGGGTTTTC
A-G_conf1_chr1_2_L	TCTGGCAGCTGCTTCTAGTTC
A-G_conf1_chr1_2_R	ATGGATGGTCCACATCACCT
chr18_1_dsRNA_L	GTGATGGAATGTGGGGAAAG
chr18_1_dsRNA_R	ATCACTCGGGAGTCTCTTCG
ZSCAN4_1_dsRNA_L	CAGGCTCAAGTGATCCTTCC
ZSCAN4_1_dsRNA_R	GCACTTTTGGAGGCTGAAAG
RPL27-1L	GCAAGAAGAAGATCGCCAAG
RPL27-1R	TCCAAGGGGATATCCACAGA
DUX4_ex2-3 F2	CGGAGAACTGCCATTCTTTC
DUX4_ex2-3 R2	CAGCCAGAATTTCACGGAAG
Luciferase_F	CCAGGTATCAGGCAAGGATATG
Luciferase_R	GTTCGTCTTCGTCCCAGTAAG
SATII_pericentric_1L	TGAATGGAATCGTCATCGAA
SATII_pericentric_1R	CCATTCGATAATTCCGCTTG
Constitutive_dsRNA1L	AAAATTGCTCAGCCTTGTGC
Constitutive _dsRNA1R	TGTCCTGCTAAATCCCCTTG

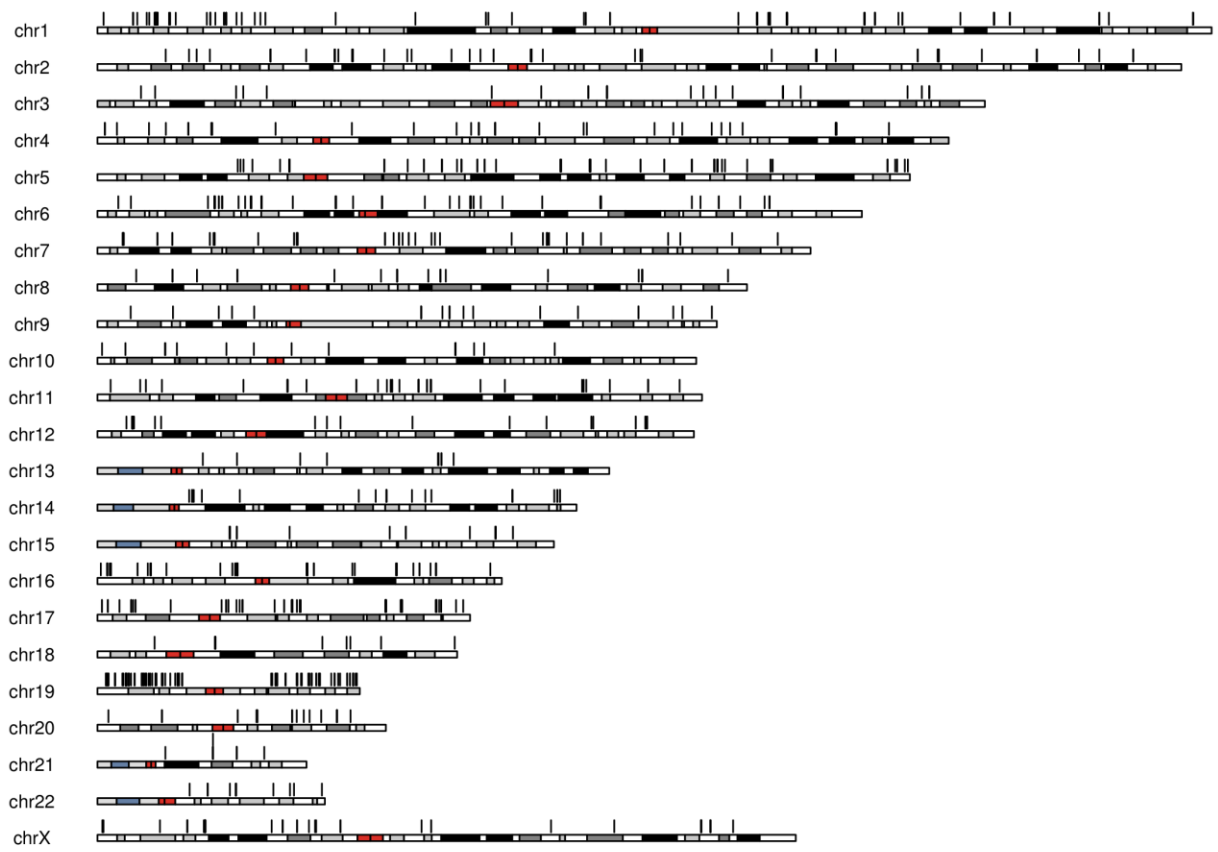
Supplementary Material

Supplemental Files for this chapter are available through the NCBI GEO via accession number GSE114940.

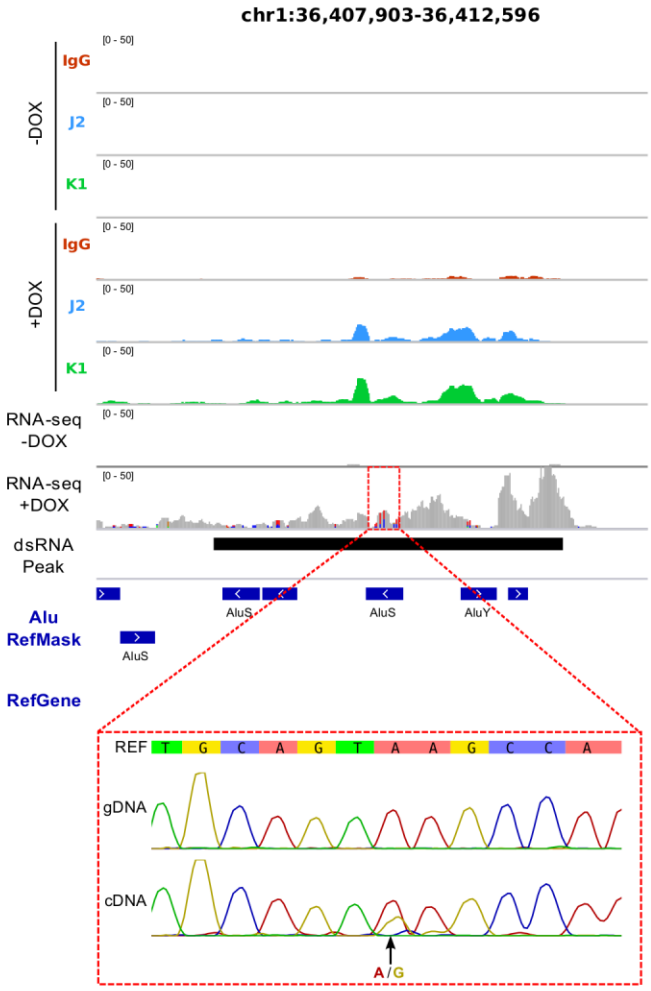
4.6 *Supplemental Figures*



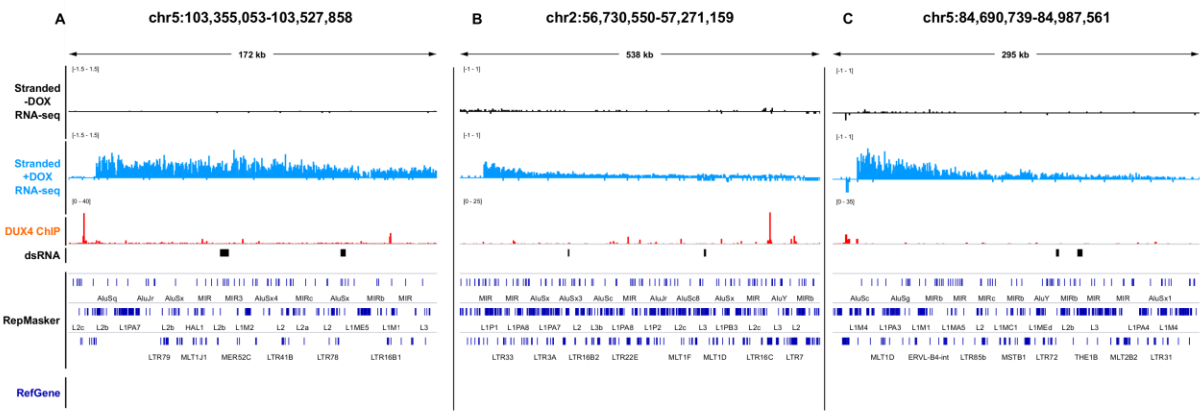
Supplemental Figure 11 Correlation heatmap based on affinity scores across all consensus peaks. Generated using the DiffBind Bioconductor package. There is a very high correlation between K1 and J2 antibody binding within each condition, which formed the basis for merging the two separate datasets in subsequent analyses.



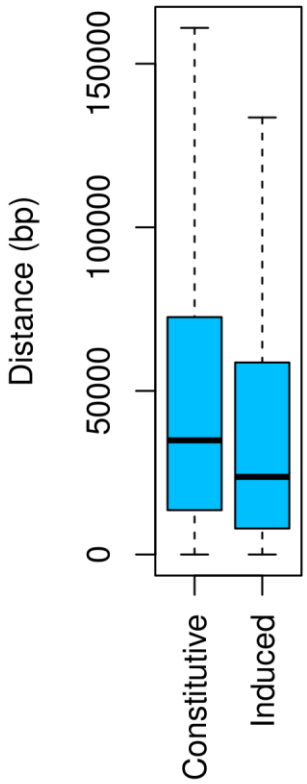
Supplemental Figure 12 Chromosomal ideogram of DUX4-induced dsRNA locations. Representation of all DUX4-induced dsRNA locations in MB135-iDUX4 cells (black bars above chromosomes).



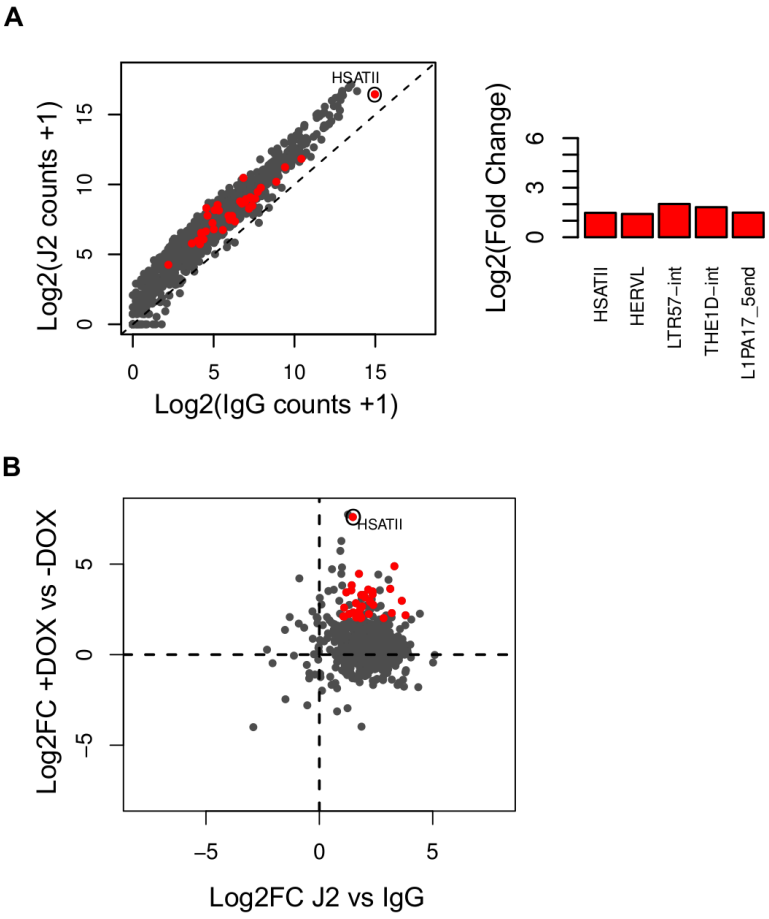
Supplemental Figure 13 Sanger sequencing confirmation of an additional RNA-editing event in a DUX4-induced dsRNA. See Fig 15D-E for description



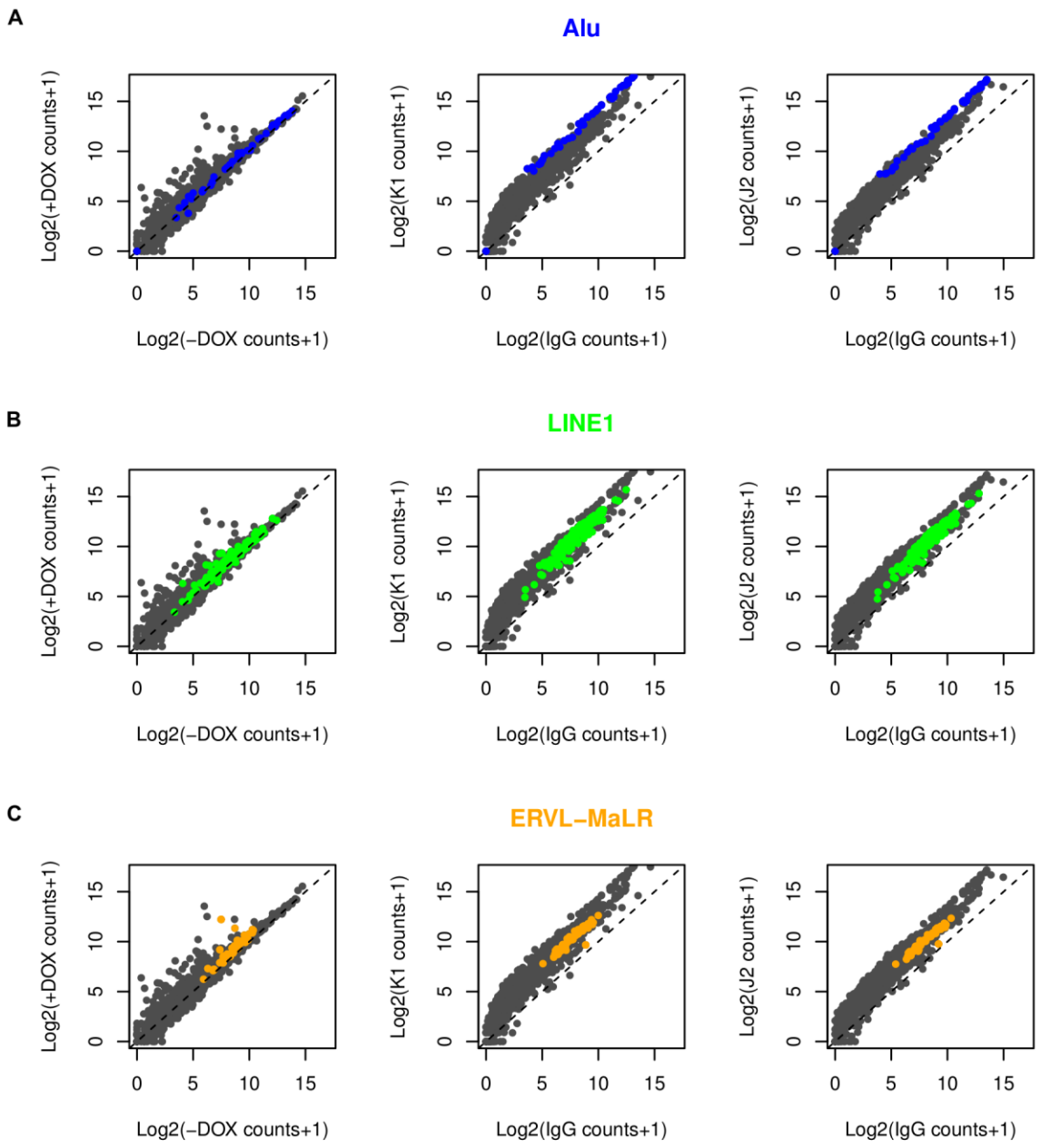
Supplemental Figure 14 Additional examples of DUX4-induced intergenic dsRNAs. (A-C) As in Figure 16C-D, for the indicated locations.



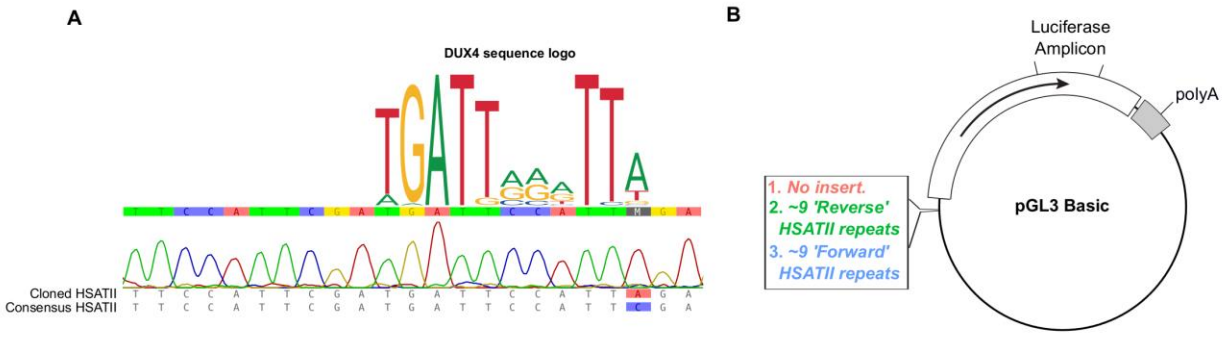
Supplemental Figure 15 DUX4-induced dsRNAs are more proximal to DUX4 ChIP-seq peaks than constitutive dsRNAs. Boxplot of base pair distance to the nearest DUX4 ChIP-seq peak of the indicated dsRNA types. P = 6.83e-10 based on the Wilcoxon rank sum test. For plotting purposes only, extreme outliers (points beyond whiskers) were omitted.



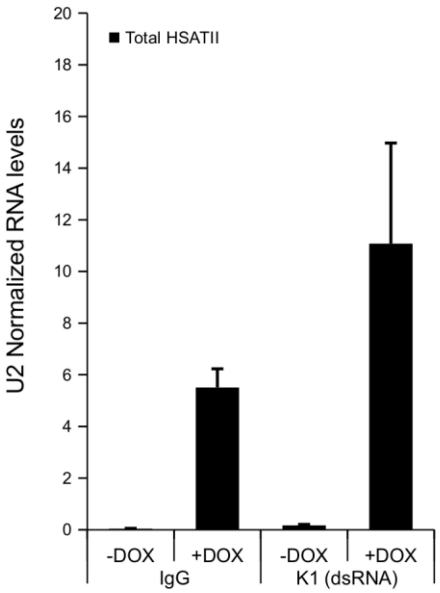
Supplemental Figure 16 DUX4-induced HSATII repeat enrichment in J2 immunoprecipitations (A) As in Fig 17B except using J2 antibody data. (B) As in Fig 17C except comparing to J2 antibody moderated log2 fold change values.



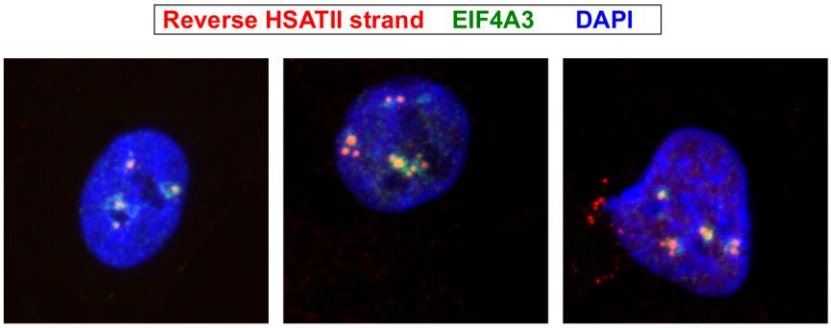
Supplemental Figure 17 dsRNA enrichment of DUX4 activated repeat families (A) Scatterplot depicting log2 normalized read counts of Dfam predicted subfamilies within stranded RNA-seq dataset of MB135-iDUX4 cells -/+ DOX (left), +DOX K1 vs IgG dsRIP samples or +DOX J2 vs IgG samples, with Alu repeats highlighted in blue (without expression or significance thresholding). (B) As in (A) except highlighting LINE1 repeat subfamilies in green. (C) As in (A) except highlighting ERV-MaLR repeat subfamilies in orange.



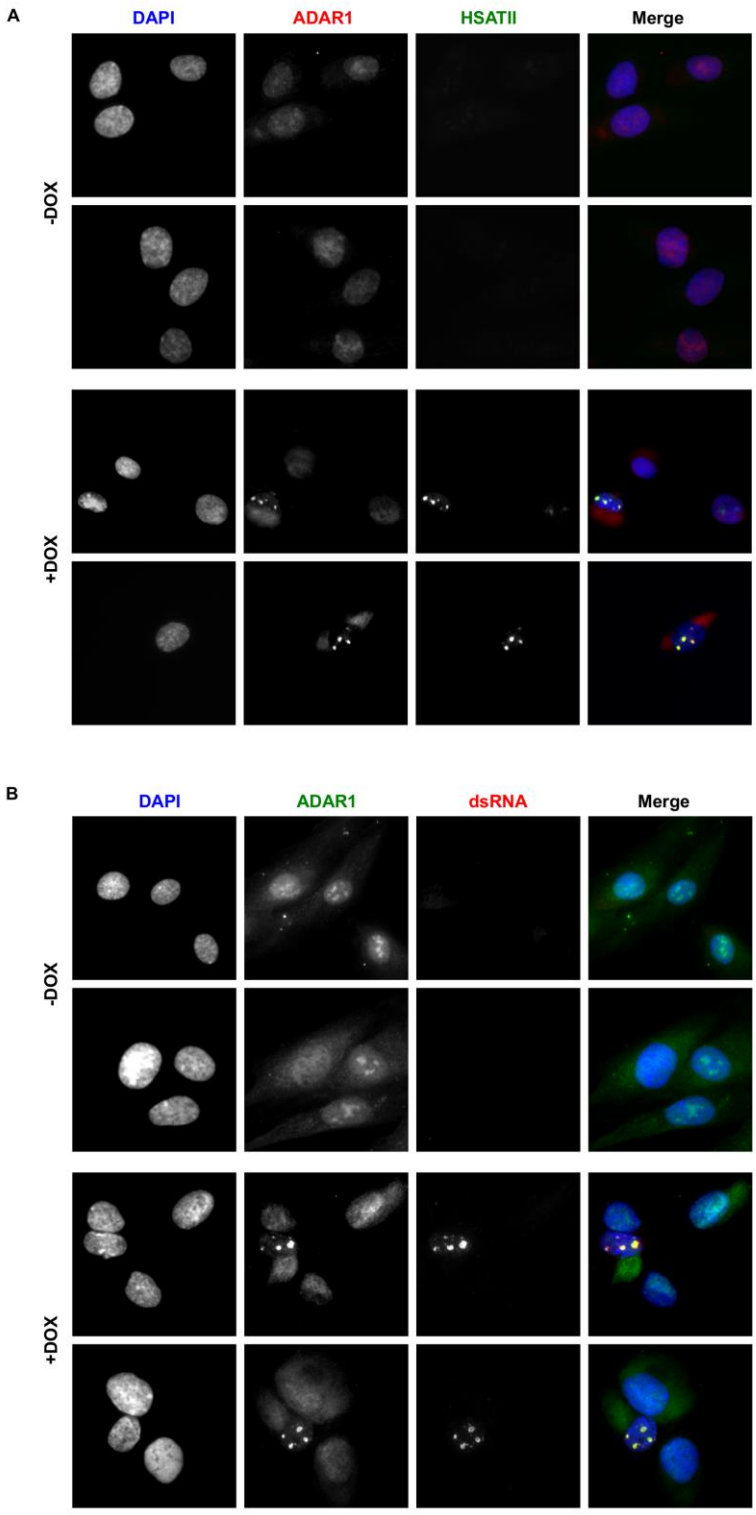
Supplemental Figure 18 DUX4 directly activates HSATII transcription. (A) Example Sanger sequencing from cloned ~1.5 kb HSATII amplicon showing divergence from consensus HSATII sequence that creates a good match for DUX4 binding. Sequence logo is based on the JASPAR database DUX4 motif. (B) Schematic showing location of ~1.5 kb HSATII cloning site in pGL3 Basic vector, upstream of luciferase gene and amplicon site for primers used to detect luciferase expression.



Supplemental Figure 19 Non-strand specific RT-qPCR for shows modest enrichment for dsRNA. As in Fig 18B except a non-strand specific RT-qPCR was performed to detect all (forward and reverse) HSATII transcripts.

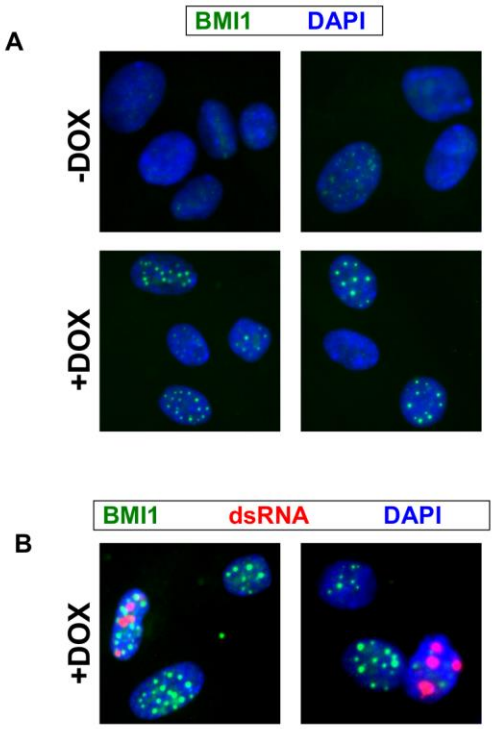


Supplemental Figure 20 EIF4A3 aggregates with HSATII transcripts in DUX4 expressing cells. Immunofluorescence in MB135-iDUX4 cells using antibody for EIF4A3 combined with RNA-FISH using probes targeting the reverse HSATII transcript. All cells were induced with doxycycline. Images are representative from two independent, combined IF RNA-FISH experiments conducted on separate days.

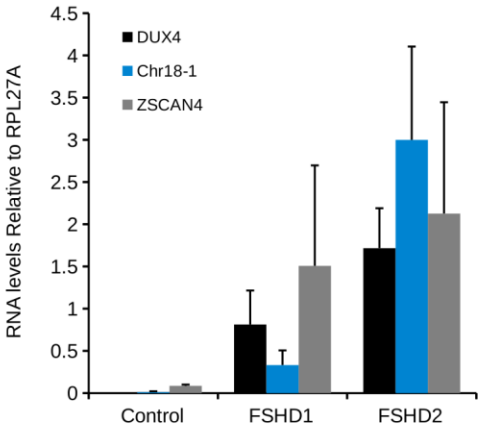


Supplemental Figure 21 ADAR1 co-localizes with DUX4-induced HSATII dsRNA nuclear foci. (A) Combined immunofluorescence with ADAR1 antibody and RNA-FISH with probes that detect the reverse HSATII strand transcripts in MB135-iDUX4 cells -/+ DOX. Images are

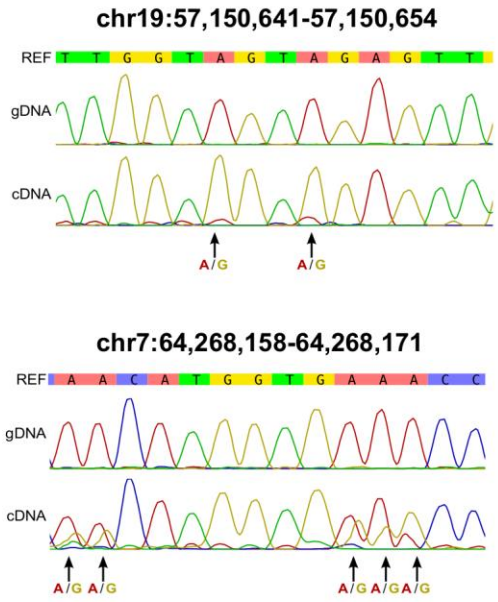
representative from one experiment. (B) Co-immunofluorescence with K1 anti dsRNA and ADAR1 antibodies in MB135-iDUX4 cells -/+ DOX. Images are representative from three independent co-IF experiments conducted on separate days.



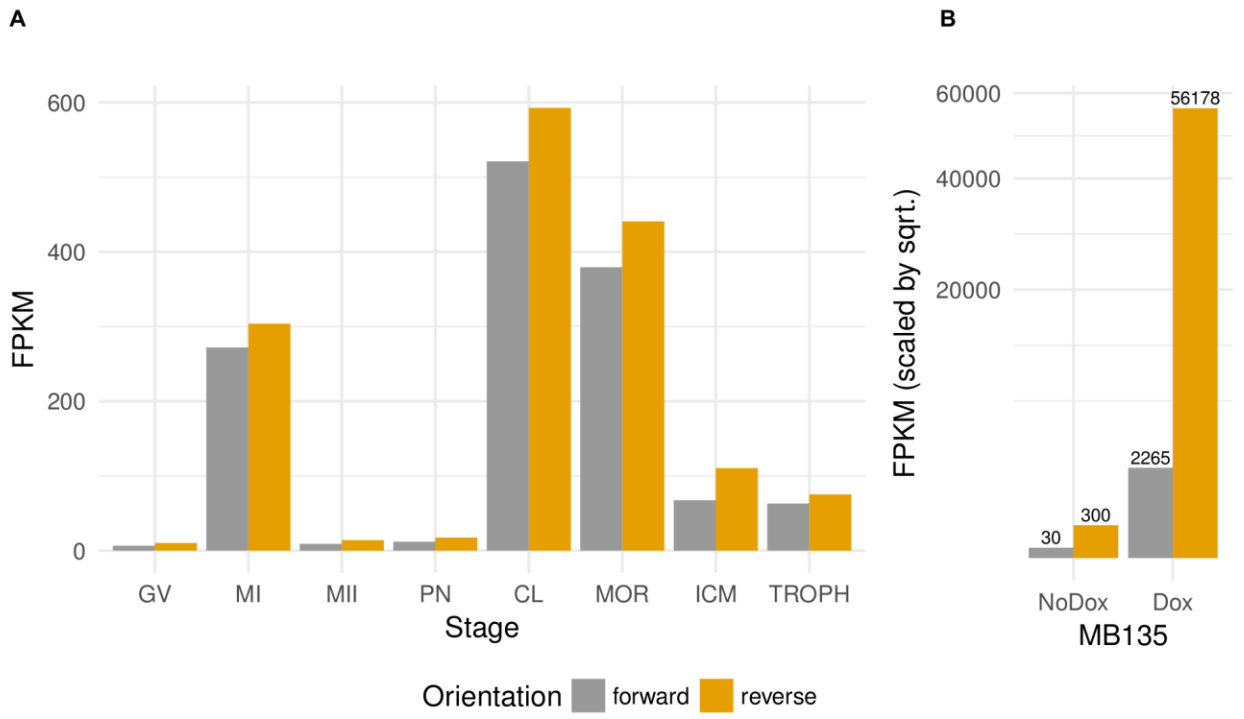
Supplemental Figure 22 BMI1 aggregates in DUX4 expressing cells do not colocalize with DUX4-induced dsRNA aggregates. (A) BMI1 immunofluorescence in -/+ DOX treated MB135-iDUX4 cells. Images are representative from two independent experiments conducted on separate days. (B) BMI1 and K1 dsRNA antibody co-immunofluorescence in doxycycline treated MB135-iDUX4 cells. Images are representative from one experiment.



Supplemental Figure 23 RT-qPCR of additional identified DUX4-induced dsRNA enriched locations. RT-qPCR data showing transcript expression levels using indicated primers in control (MB135), FSHD1 (MB073) or FSHD2 (MB200) cells grown in differentiation medium. Data are normalized to RPL27 levels and depicted as the mean values of three experiments performed on independent days. Error bars represent the standard deviation of the mean.



Supplemental Figure 24 Confirmation of dsRNA editing in FSHD cells. Sanger sequencing results of amplicons designed across the shown regions (hg38) in either genomic DNA (gDNA) or complementary DNA (cDNA) of differentiated FSHD2 (MB200) cells, with AG mismatches highlighted.



Supplemental Figure 25 HSATII transcript strand expression. (A) FPKM from RNA-seq data in the indicated developmental stages of reads mapping to HSATII repeat sequences on either forward or reverse strand (relative to the consensus HSATII sequence). (B) as in (A) except in MB135-iDUX4 cells and scaled by the square root, with unadjusted FPKM values depicted above bars.

Chapter 5. *Discussion*

A major objective of my thesis work was to characterize the molecular pathways responsible for DUX4 toxicity. To attain this goal, I developed a model cell culture system to study the effects of DUX4 expression and showed that this system faithfully recapitulated known transcriptional responses to DUX4 which occur in FSHD cells. This includes the transcriptional activation of repeat elements which are enriched during early embryonic development [29] and reactivated in some tumor cell lines [126] but suppressed in somatic tissues [46]. Presumably, repeat sequences are normally silenced to avoid the deleterious effects of their expression, though the consequences of their misexpression in DUX4 toxicity and/or pathology is less explored.

My thesis work revealed two primary pathways responsible for DUX4 mediated cell death. First, DUX4 expression stabilized mRNA of the known oncogene, c-MYC through a still unidentified mechanism. Overexpression of c-MYC can lead to apoptosis in non-tumor cells [85], presumably to avoid tumorigenesis. It is plausible that upregulation of c-MYC protein contributes to toxicity in DUX4 expressing cells as knockdown of MYC diminished DUX4 toxicity. Second, I discovered that DUX4 expressing cells formed aggregates of dsRNA and activated the antiviral, dsRNA sensing, pro-apoptotic PKR protein. Knockdown of either PKR or RNASEL, another antiviral, dsRNA responding protein, mitigated at least some of the toxicity caused by expression of DUX4. Since DUX4 activates endogenous retroviral sequences and other repeats [16] that may be inappropriately sensed as exogenous viral sequences, I hypothesized that the massively increased expression of these repeats could form dsRNAs leading to activation of dsRNA response pathways in DUX4 expressing cells.

Indeed, DUX4 activated repeat sequences likely form double stranded RNA structures. Unlike “constitutive” dsRNAs, which are mainly formed from inverted repeats of short *Alu* SINEs residing within introns of protein coding genes [118], DUX4-induced dsRNAs can be formed

largely via repeats embedded within intergenic noncoding sequences. These sequences were relatively enriched for HERVL, MaLR, LINE-1 and HSATII pericentric repeats. Because of the massive upregulation of HSATII repeats following DUX4 expression, I focused on this class of repeats in more detail. I found that, like with early embryo expression of mouse pericentric repeats [35], HSATII transcripts were bidirectionally expressed with a predominant “reverse” strand preceding the expression of the less predominant “forward” strand.

In sum, my results provide evidence for new molecular mechanisms of DUX4 toxicity which merit further study. Below, I discuss unresolved areas and future avenues of research regarding the roles for MYC and dsRNA defense pathways in mediating DUX4 toxicity.

5.1 *DUX4 and enhanced c-MYC expression*

The identification of c-MYC (MYC) as a potential mediator of DUX4 toxicity raises several outstanding questions. One question that remains unanswered is how MYC mRNA is stabilized by DUX4. Under most cellular conditions, MYC mRNA is turned over relatively quickly [145]. As DUX4 suppresses RNA quality control pathways such as NMD [42], I considered that MYC mRNA might consequently be stabilized by diminished NMD. However, knockdown of the NMD component EIF4A3 and/or combined knockdown with exon junction component Y14 did not alter MYC mRNA levels, suggesting an alternative mechanism.

MYC mRNA levels are co-translationally controlled via binding of the Coding Region Determinant (CRD) sequence within the *MYC* open reading frame by IGF2BP1 [146]. Binding of the CRD by IGF2BP1 stabilizes MYC mRNA by preventing endonucleolytic cleavage, but translation uncovers the CRD making it available to endunucleases and yielding tight co-translational control of MYC protein levels [146]. It remains to be determined whether the CRD plays a meaningful role in DUX4-mediated stabilization of MYC mRNA. Interestingly, several dsRNA-binding proteins such as Staufen1 and the RNA helicase DHX9 also appear to play a positive role in MYC RNA stability via the CRD [146]. Perhaps as a homeostatic consequence of

dsRNA, DHX9 RNA levels increase in DUX4 expressing cells. This increase could provide at least a partial mechanism for MYC stabilization. In whole, future studies will be required to elucidate the importance of the CRD in DUX4-mediated MYC mRNA stabilization.

An unanswered question is whether DUX4 increased MYC plays a role in early stem cell biology. MYC was included as one of the original factors for reprogramming fibroblasts to induced pluripotent stem cells [147] and has been shown to be important for both the establishment and maintenance of pluripotency. MYC levels increase during mouse embryonic genome activation [148], when DUX4 is normally expressed [29]. MYC associates with co-activators and chromatin modifiers to alter the expression landscape and could conceivably cooperate with DUX4 to achieve full activation of its early transcriptional program.

Another possible consequence for elevated MYC during early development might be in “cell competition”. Cell competition is a process whereby winner (high MYC) cells provoke apoptosis in loser (low MYC) cells as a means of establishing cellular fitness benefit for the organism [149]. Cell competition is particularly evident in epiblast cells of the mouse embryo [149]. Controlling MYC levels via RNA stabilization may be an important component of DUX4 mediated zygotic genome activation, DUX4 target gene activation and cell-non-autonomous effects in FSHD muscle. Therefore, it would be interesting to further investigate whether DUX4 expression can influence cell death in the non-DUX4-expressing cells in FSHD muscle cell culture because of increased MYC levels.

Enhanced MYC levels may also lead to a shift in metabolism that resembles a pluripotent cell state including upregulation of glycolysis as a major source of energy production. DUX4 expressing cells do appear to shift to a more glycolytic form of metabolism based on measurements of extracellular acidification and oxygen consumption rates (unpublished observations). The role for MYC in this potential metabolic shift in DUX4 expressing cells will require future study, but it is tempting to speculate that MYC might be a major component of this shift as MYC maintains transcription required for elevated glycolysis in human pluripotent stem

cells [150]. Future experiments should elucidate the importance for MYC in activation of the transcriptional network of DUX4 and in alteration of metabolic pathways.

5.2 *Activation of repeats including HSATII by DUX4*

In human and mouse, many transposable elements, including endogenous retroviruses, are highly expressed during early embryogenesis. Recently, functional roles for repeat transcripts have been explored in greater depth and it is now appreciated that the expression of certain repeat elements is even necessary for developmental progression. For example, depleting LINE-1 transcription in the early mouse embryo arrests cells at preimplantation. Fascinatingly, it was recently found that LINE-1 appears to shape the transcriptome in the early mouse embryo directly through transcription-directed chromatin reorganization, independent of LINE-1 retrotransposition [151]. LINE-1 transcript levels are notably elevated following DUX4 expression (see Chapter 4 and [16]), mimicking the high levels of LINE-1 in the early embryo. While it has been shown that DUX4 can bind to and directly activate other repeats such as ERVs, it is currently unclear how DUX4 would activate LINE-1. Based on the lack of convincing ChIP enrichment and no identified DUX4 binding motif within LINE-1, it is likely that activation by DUX4 is indirect. A mechanistic understanding of DUX4 triggered LINE-1 upregulation may be gained from studies of aberrant repeat transcription in cancer cells. Of note, some tumor cells, as in FSHD cells, have both elevated HSATII and LINE-1 repeat expression [126]. In these tumor cells, an apparent loss of DNA methylation at HSATII-containing 1q12 DNA leads to redistribution of the polycomb repressive complex 1 (PRC1) to the demethylated pericentric 1q12 locus, leading to a loss of normal PRC1 repression elsewhere [126]. This PRC1 sequestration might dampen the normal repression of LINE-1 repeats thus facilitating their reactivation. Because of HSATII activation and PRC1 reorganization in DUX4 expressing cells, it is possible that DUX4 induced LINE-1 repeat reactivation parallels the proposed mechanism observed in HSATII positive cancer cells. An important next step would be to determine whether

the PRC1 component BMI1 is indeed sequestered to HSATII DNA loci in DUX4 expressing cells.

Interestingly, a recent study showed that *Dux* expression itself may be controlled by LINE-1 ncRNA recruitment of repressor proteins to the *Dux* locus [152]. If LINE-1 ncRNA also represses *DUX4* expression as it can in mice, it could provide an intriguing mechanism of negative auto-regulation in the embryo in order to fine tune DUX4 expression. That is, if DUX4 indirectly activates LINE-1 ncRNAs, then they could in turn feedback to turn off DUX4 expression. Determining precisely how DUX4 controls LINE-1 expression and whether LINE-1 indeed feeds back to influence DUX4 expression in humans could then enhance our understanding of embryonic genome activation in humans.

What importance might HSATII expression have in early human development? Pericentric satellite repeats are thought to be repressed in somatic tissues because their expression is associated with genome instability and the activation of innate immune pathways [52,153]. Mouse pericentric major satellite expression appears to peak in the two-cell stage and, like LINE-1, is necessary for developmental progression [35] likely because their transcription is, paradoxically, needed for heterochromatin formation at pericentromeres [50]. Mechanistically this heterochromatinization probably occurs via recruitment of repressor proteins which would include the SUV39H H3K9 methyltransferases and HP1. As detailed in Chapter 4, it has been speculated that double stranded RNAs could play a role in mammalian pericentromere heterochromatin formation [50] as occurs in fission yeast [112], though this idea is somewhat controversial. Future studies should clarify the role, if any, for dsRNA and/or RNA interference mechanisms relying on dicer in establishing pericentric heterochromatin in mammals.

An alternative mechanism for RNA-mediated HSATII silencing could be via single stranded RNA recruitment of repressive complexes. Single stranded pericentric RNAs can interact with and recruit Suv39h in mice [154], which likely maintains pericentric heterochromatin. The functional role for bidirectional transcription is less clear under this model,

though dsRNAs could conceivably be more important for establishment [50] rather than maintenance of heterochromatin. One possibility is that bidirectional transcription can abruptly halt RNA polymerase progression on pericentromeres due to collision, which is a known mechanism for gene regulation [155]. Once suitable satellite transcript levels are reached, the transcription of the opposite orientation could act as the mechanism to quickly stop transcription as heterochromatin formation is completed. If the “factor” responsible for forward (antisense) HSATII transcription was known, then we could test for collision by depleting this factor. Currently, efforts to identify this factor are under way and will make use of our previously published enChIP technique [22] for co-immunoprecipitation of DNA-protein interactions using a catalytically dead CAS9 protein targeted to HSATII repeats.

5.3 *Conclusion*

There is a gap in the collective understanding of how DUX4 mis-expression can lead to FSHD. In my dissertation work I developed a cell culture system for studying FSHD in order to explore pathways involved in FSHD pathogenesis. My work raises the possibility that stabilization of labile transcripts such as MYC and activation of repeats, including the activation of pericentric satellite repeats, are at least partially responsible for DUX4 toxicity. Future studies will be needed to crystalize the role for repeat activation in DUX4 toxicity associated with FSHD and in heterochromatin formation in the early embryo. My work suggests that integrating our growing knowledge of the pathways activated by DUX with the recent idea that DUX4 plays a fundamental role in development of the early embryo can further our ability to understand this complex disease.

References

1. Hilbert JE, Kissel JT, Luebbe EA, Martens WB, Mcdermott MP, Sanders DB, et al. If you build a rare disease registry, will they enroll and will they use it? Methods and data from the National Registry of Myotonic Dystrophy (DM) and Facioscapulohumeral Muscular Dystrophy (FSHD). Contemp Clin Trials. 2012;33(2):302–11.
2. Richards M, Coppée F, Thomas N, Belayew A, Upadhyaya M. Facioscapulohumeral muscular dystrophy (FSHD): An enigma unravelled? Hum Genet. 2012;131(3):325–40.
3. Wijmenga C, Brouwer OF, Moerer P, Padberg GW, Frants RR, Weber JL. Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. Lancet. 1990;336(8716):651–3.
4. Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, et al. Chromosome 4 DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet. 1992;2:26–30.
5. Deutekom JCT van, Wljmenga C, Tlenhoven EAE van, Gruter AM, Hewitt JE, Padberg GW, 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):2037–42.
6. Hewitt JE, Lyle R, Clark LN, Valleley EM, Wright TJ, Wijmenga C, et al. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystropothhy. Hum Mol Genet. 1994;3(8):1287–95.
7. Gabellini D, Green MR, Tupler R. Inappropriate gene activation in FSHD: A repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell. 2002;110(3):339–48.
8. Gabellini D, D'Antona G, Moggio M, Prelle A, Zecca C, Adami R, et al. Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. Nature. 2005;439(7079):973.

9. Klooster R, Straasheijm K, Shah B, Sowden J, Frants R, Thornton C, et al. Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. Eur J Hum Genet. 2009;17(12):1615–24.

10. Tupler R, Berardinelli A. Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. J Med Genet. 1996;33:366–70.

11. Gabriëls J, Beckers M-C, Ding H, De Vriese A, Plaisance S, van der Maarel S., et al. Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene. 1999 Aug;236(1):25–32.

12. Beckers MC, Gabriëls J, Van Der Maarel S, De Vriese A, Frants RR, Collen D, et al. Active genes in junk DNA? Characterization of DUX genes embedded within 3.3 kb repeated elements. Gene. 2001;264(1):51–7.

13. Lemmers RJLF, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, Dauwerse JG, et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science. 2010 Sep 24;329(5999):1650–3.

14. Snider L, Geng LN, Lemmers RJLF, Kyba M, Ware CB, Nelson AM, et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet. 2010 Oct;6(10):e1001181.

15. Geng LN, Yao Z, Snider L, Fong AP, Cech JN, Young JM, et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Dev Cell. 2012 Jan 17;22(1):38–51.

16. Young JM, Whiddon JL, Yao Z, Kasinathan B, Snider L, Geng LN, et al. DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. PLoS Genet. 2013 Nov;9(11):e1003947.

17. Gilbert JR, Stajich JM, Wall S, Carter SC, Qiu H, Vance JM, et al. Evidence for heterogeneity in facioscapulohumeral muscular dystrophy (FSHD). Am J Hum Genet. 1993;53(2):401.

18. Lemmers RJLF, Tawil R, Petek LM, Balog J, Block GJ, Santen GWE, et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. Nat Genet. 2012;44:1370–4.

19. Blewitt ME, Gendrel A-V, Pang Z, Sparrow DB, Whitelaw N, Craig JM, et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. Nat Genet. 2008 May;40(5):663–9.

20. Balog J, Thijssen PE, Shadle S, Straasheijm KR, Van Der Vliet PJ, Krom YD, et al. Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4. Epigenetics. 2015 Nov 17;2294(November).

21. Van Den Boogaard ML, Lemmers RJLF, Balog J, Wohlgemuth M, Auranen M, Mitsuhashi S, 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):1020–9.

22. Campbell AE, Shadle SC, Jagannathan S, Lim JW, Resnick R, Tawil R, et al. NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. Elife. 2018;7:e31023.

23. Zeng W, de Greef JC, Chen Y-Y, Chien R, Kong X, Gregson HC, et al. Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS Genet. 2009 Jul;5(7):e1000559.

24. Ottaviani A, Rival-Gervier S, Boussouar A, Foerster AM, Rondier D, Sacconi S, et al. The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in Facio-Scapulo-Humeral dystrophy. PLoS Genet. 2009;5(2):e1000394.

25. Lim J-W, Snider L, Yao Z, Tawil R, Van Der Maarel SM, Rigo F, 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):4817–28.

26. Barr H, Hermann A, Berger J, Tsai H-H, Adie K, Prokhortchouk A, et al. Mbd2 contributes

to DNA methylation-directed repression of the Xist gene. Mol Cell Biol. 2007 May;27(10):3750–7.

27. Sacconi S, Lemmers RJLF, Balog J, Van Der Vliet PJ, Lahaut P, Van Nieuwenhuizen MP, et al. The FSHD2 gene SMCHD1 Is a modifier of disease severity in families affected by FSHD1. Am J Hum Genet. 2013;93(4):744–51.
28. 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;
29. Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet. 2017;49(6):925–34.
30. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet. 2017;49(6):941–5.
31. Whiddon JL, Langford AT, Wong CJ, Zhong JW, Tapscott SJ. Conservation and innovation in the DUX4-family gene network. Nat Genet. 2017;49(6):935–40.
32. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature. 2012;487(7405):57.
33. Peaston AE, Evsikov A V, Graber JH, Vries WN De, Holbrook AE, Solter D, et al. Retrotransposons Regulate Host Genes in Mouse Oocytes and Preimplantation Embryos. Dev Cell. 2004;7(4):597–606.
34. Ishiuchi T, Enriquez-Gasca R, Mizutani E, Bošković A, Ziegler-Birling C, Rodriguez-Terrones D, et al. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat Struct Mol Biol. 2015;22(9):662–71.
35. Probst A V, Okamoto I, Casanova M, El Marjou F, Le Baccon P, Almouzni G. A Strand-specific burst in transcription of pericentric satellites is required for chromocenter

formation and early mouse development. Dev Cell. 2010;19(4):625–38.

36. Bosnakovski D, Xu Z, Gang EJ, Galindo CL, Liu M, Simsek T, et al. An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. EMBO J. 2008 Oct 22;27(20):2766–79.

37. Bosnakovski D, Choi SH, Strasser JM, Toso E a, Walters M a, Kyba M. High-throughput screening identifies inhibitors of DUX4-induced myoblast toxicity. Skelet Muscle. 2014 Jan;4(1):4.

38. Banerji CRS, Panamarova M, Hebaishi H, White RB, Relaix F, Severini S, et al. PAX7 target genes are globally repressed in facioscapulohumeral muscular dystrophy skeletal muscle. Nat Commun. 2017;8(1):2152.

39. Homma S, Beermann M Lou, Boyce FM, Miller JB. Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. Ann Clin Transl Neurol. 2015 Feb;2(2):151–66.

40. Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, Vande Velde C, et al. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat Genet. 2008;40(5):572–4.

41. Tollervey JR, Curk T, Rogelj B, Briese M, Cereda M, Kayikci M, et al. Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat Neurosci. 2011;14(4):452–8.

42. Feng Q, Snider L, Jagannathan S, Tawil R, van der Maarel SM, Tapscott SJ, et al. A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. Elife. 2015 Jan 7;4:1–13.

43. Saldi TK, Ash PE, Wilson G, Gonzales P, Garrido-Lecca A, Roberts CM, et al. TDP-1, the Caenorhabditis elegans ortholog of TDP-43, limits the accumulation of double-stranded RNA. EMBO J. 2014 Dec 17;33(24):2947–66.

44. Shelkovnikova TA, Kukharsky MS, An H, Dimasi P, Alexeeva S, Shabir O, et al.

Protective paraspeckle hyper-assembly downstream of TDP-43 loss of function in amyotrophic lateral sclerosis. Mol Neurodegener. 2018;13(1):30.

45. Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, et al. Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol. 2004;22(8):1001.

46. Zamudio N, Bourc'His D. Transposable elements in the mammalian germline: A comfortable niche or a deadly trap. Heredity (Edinb). 2010;105(1):92–104.

47. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. Cell. 2015 Aug;162(5):961–73.

48. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. Cell. 2015 Aug;162(5):974–86.

49. Bersani F, Lee E, Kharchenko P V., Xu AW, Liu M, Xega K, et al. Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. Proc Natl Acad Sci. 2015;112(49):15148–53.

50. Santenard A, Ziegler-Birling C, Koch M, Tora L, Bannister AJ, Torres-Padilla ME. Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. Nat Cell Biol. 2010;12(9):853–62.

51. Leonova KI, Brodsky L, Lipchick B, Pal M, Novototskaya L, Chenchik AA, et al. p53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs. Proc Natl Acad Sci. 2013;110(1):E89–98.

52. Zhu Q, Hoong N, Aslanian A, Hara T, Benner C, Heinz S, et al. Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. Mol Cell. 2018;

53. Kowaljow V, Marcowycz A, Ansseau E, Conde CB, Sauvage S, Mattéotti C, et al. The

DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord. 2007 Aug;17(8):611–23.

54. Vanderplanck C, Ansseau E, Charron S, Stricwant N, Tassin A, Laoudj-Chenivesse D, et al. The FSHD atrophic myotube phenotype is caused by DUX4 expression. PLoS One. 2011 Jan;6(10):e26820.

55. Wallace LM, Garwick SE, Mei W, Belayew A, Coppee F, Ladner KJ, et al. DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. Ann Neurol. 2011 Mar;69(3):540–52.

56. Dandapat A, Bosnakovski D, Hartweck LM, Arpke RW, Baltgalvis KA, Vang D, et al. Dominant Lethal Pathologies in Male Mice Engineered to Contain an X-Linked DUX4 Transgene. Cell Rep. 2014;8(5):1484–96.

57. Rickard AM, Petek LM, Miller DG. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Hum Mol Genet. 2015 Oct 15;24(20):5901–14.

58. Block GJ, Narayanan D, Amell AM, Petek LM, Davidson KC, Bird TD, et al. Wnt/β-catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. Hum Mol Genet. 2013 Dec 1;22(23):4661–72.

59. Banerji CRS, Knopp P, Moyle LA, Severini S, Orrell RW, Teschendorff AE, et al. β-catenin is central to DUX4 -driven network rewiring in facioscapulohumeral muscular dystrophy. J R Soc Interface. 2015;12(102):20140797.

60. Lek A, Rahimov F, Jones PL, Kunkel LM. Emerging preclinical animal models for FSHD. Trends Mol Med. 2015;21(5):295–306.

61. Yao Z, Snider L, Balog J, Lemmers RJLF, Van Der Maarel SM, Tawil R, et al. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. Hum Mol Genet. 2014 Oct 15;23(20):5342–52.

62. Soneson C, Delorenzi M. A comparison of methods for differential expression analysis of

RNA-seq data. BMC Bioinformatics. 2013;14(1):91.

63. Chen Y, McCarthy D, Robinson M, Smyth GK. edgeR: differential expression analysis of digital gene expression data User's Guide. Bioconductor; 2014.

64. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. Genome Biol. 2016;17(1):13.

65. Snider L, Asawachaicharn A, Tyler AE, Geng LN, Petek LM, Maves L, et al. RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: New candidates for the pathophysiology of facioscapulohumeral dystrophy. Hum Mol Genet. 2009;18(13):2414–30.

66. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

67. Li B, Dewey CN. RSEM : accurate transcript quantification from RNA - Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(323):1–16.

68. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.

69. Dvinge H, Ries RE, Ilagan JO, Stirewalt DL, Meshinchi S, Bradley RK. Sample processing obscures cancer-specific alterations in leukemic transcriptomes. Proc Natl Acad Sci. 2014;111(47):16802–7.

70. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11(3):R25.

71. Wickham H. ggplot2: elegant graphics for data analysis (Use R!). Springer New York, NY; 2010.

72. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. 2015;44(D1):D336–42.

73. Tawil R, van der Maarel SM, Tapscott SJ. Facioscapulohumeral dystrophy: the path to

consensus on pathophysiology. Skelet Muscle. 2014;4(1):12.

74. Das S, Chadwick BP. Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. PLoS One. 2016;11(7):e0160022.

75. Meier P, Finch A, Evan G. Apoptosis in development. Nature. 2000 Oct 12;407(6805):796–801.

76. Eguchi K. Apoptosis in autoimmune diseases. Intern Med. 2001;40(4):1–7.

77. Mattson MP. Apoptosis in Neurodegenerative Disorders. Nat Rev Mol Cell Biol. 2000 Nov 1;1(2):120–30.

78. Shaha C, Tripathi R, Mishra DP. Male germ cell apoptosis: regulation and biology. Philos Trans R Soc Lond B Biol Sci. 2010 May 27;365(1546):1501–15.

79. Turki A, Hayot M, Carnac G, Pillard F, Passerieux E, Bommart S, et al. Functional muscle impairment in facioscapulohumeral muscular dystrophy is correlated with oxidative stress and mitochondrial dysfunction. Free Radic Biol Med. 2012 Sep 1;53(5):1068–79.

80. Felix CA, Kappel CC, Mitsudomi T, Nau MM, Tsokos M, Crouch GD, et al. Frequency and diversity of p53 mutations in childhood rhabdomyosarcoma. Cancer Res. 1992 Apr 15;52(8):2243–7.

81. Otten AD, Firpo EJ, Gerber AN, Brody LL, Roberts JM, Tapscott SJ. Inactivation of MyoD-mediated expression of p21 in tumor cell lines. Cell Growth Differ. 1997 Nov;8(11):1151–60.

82. Pelengaris S, Khan M, Evan G. Suppression of Myc-induced apoptosis in β cells exposes multiple oncogenic properties of Myc and triggers carcinogenic progression. Cell. 2002;109:321–34.

83. Colau, G., Thiry, M., Leduc, V., Bordonné, R., & Lafontaine DL. The Small Nucle(ol)ar RNA Cap Trimethyltransferase Is Required for Ribosome Synthesis and Intact Nucleolar Morphology. Mol Cell Biol. 2004;24(18):7976–86.

84. Koutnikova H, Campuzano V, Foury F, Dollé P, Cazzalini O, Koenig M. Studies of

human, mouse and yeast homologues indicate a mitochondrial function for frataxin. Nat Genet. 1997 Aug;16(4):345–51.

85. Wirth M, Stojanovic N, Christian J, Paul MC, Stauber RH, Schmid RM, et al. MYC and EGR1 synergize to trigger tumor cell death by controlling NOXA and BIM transcription upon treatment with the proteasome inhibitor bortezomib. Nucleic Acids Res. 2014 Jan;42(16):10433–47.

86. Liu J, Chandra D, Tang S, Chopra D, Tang D. Identification and characterization of Bimγ, a novel proapoptotic BH3-only splice variant of Bim. Cancer Res. 2002;(62):2976–81.

87. Hoffman B, Liebermann D a. Apoptotic signaling by c-MYC. Oncogene. 2008 Oct 27;27(50):6462–72.

88. Schonborn J, Oberstraβ J, Breyel E, Tittgen J, Schumacher J, Lukacs N. Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts. Nucleic Acids Res. 1991;19(11):2993–3000.

89. Srivastava SP, Kumar KU, Kaufman RJ. Phosphorylation of Eukaryotic Translation Initiation Factor 2 Mediates Apoptosis in Response to Activation of the Double-stranded RNA-dependent Protein Kinase. J Biol Chem. 1998 Jan 23;273(4):2416–23.

90. Li T, Shi Y, Wang P, Guachalla L. Smg6/Est1 licenses embryonic stem cell differentiation via nonsense-mediated mRNA decay. EMBO J. 2015;34(12):1630–47.

91. Wery M, Descrimes M, Vogt N, Dallongeville A-S, Gautheret D, Morillon A. Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. Mol Cell. 2016 Feb 4;61(3):379–92.

92. Jagannathan S, Shadle S, Resnick R, Snider L, Tawil RN, van der Maarel SM, et al. Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. Hum Mol Genet. 2016 Aug 15;(415):ddw271.

93. Miller JW, Urbinati CR, Teng-Umnuay P, Stenberg MG, Byrne BJ, Thornton CA, et al. Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with

myotonic dystrophy. EMBO J. 2000;19(17):4439–48.

94. Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS, et al. Three proteins, MBNL, MBLL and MBXL, co-localize in vivo with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. Hum Mol Genet. 2002;11(7):805–14.

95. Prendergast GC. Mechanisms of apoptosis by c-Myc. Oncogene. 1999 May 13;18(19):2967–87.

96. Chan CC, Dostie J, Diem MD, Feng W, Mann M, Rappsilber J, et al. eIF4A3 is a novel component of the exon junction complex. Rna. 2004;10(2):200–9.

97. Wang D, Wengrod J, Gardner LB. Overexpression of the c-myc oncogene inhibits nonsense-mediated RNA decay in B lymphocytes. J Biol Chem. 2011 Nov 18;286(46):40038–43.

98. Zhang F, Romano PR, Nagamura-inoue T, Tian B, Dever TE, Mathews MB, et al. Binding of Double-stranded RNA to Protein Kinase PKR Is Required for Dimerization and Promotes Critical Autophosphorylation Events in the Activation Loop. J Biol Chem. 2001;276(27):24946–58.

99. Meyer SN, Amoyel M, Bergantinos C, de la Cova C, Schertel C, Basler K, et al. An ancient defense system eliminates unfit cells from developing tissues during cell competition. Science (80-). 2014 Dec 5;346(6214):1258236–1258236.

100. Gallant P. Myc, cell competition, and compensatory proliferation. Cancer Res. 2005 Aug 1;65(15):6485–7.

101. Yasuda T, Tsuzuki S, Kawazu M, Hayakawa F, Kojima S, Ueno T, et al. Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. Nat Genet. 2016 Mar 28;48(5).

102. Liu Y-F, Wang B-Y, Zhang W-N, Huang J-Y, Li B-S, Zhang M, et al. Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia. EBioMedicine. 2016 Jun;8:173–83.

103. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods. 2014;11(8):783–4.

104. Boutros M, Brás LP, Huber W. Analysis of cell-based RNAi screens. Genome Biol. 2006;7(7):R66.

105. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB--a database for integrating human functional interaction networks. Nucleic Acids Res. 2009 Jan;37(Database issue):D623-8.

106. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. Prlic A, editor. PLoS Comput Biol. 2013 Aug 8;9(8):e1003118.

107. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

108. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6.

109. Schneider C a, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9(7):671–5.

110. Skene PJ, Henikoff S. A simple method for generating high-resolution maps of genome-wide protein binding. Elife. 2015;4:e09225.

111. Geng LN, Tyler AE, Tapscott SJ. Immunodetection of human double homeobox 4. Hybridoma. 2011;30(2):125–30.

112. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science (80-). 2002;297(5588):1833–7.

113. Shadle SC, Zhong JW, Campbell AE, Conerly ML, Jagannathan S, Wong CJ, et al. DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. PLoS Genet. 2017 Mar

8;13(3):e1006658.

114. Schlee M, Hartmann G. Discriminating self from non-self in nucleic acid sensing. Nat Rev Immunol. 2016;16(9):566–80.

115. Bass BL. RNA editing by adenosine deaminates that act on RNA. Annu Rev Biochem. 2002;71(I):817–46.

116. Athanasiadis A, Rich A, Maas S. Widespread A-to-I RNA Editing of Alu-Containing mRNAs in the Human Transcriptome. PLOS Biol. 2004 Nov 9;2(12):e391.

117. Blango MG, Bass BL. Identification of the long, edited dsRNAome of LPS-stimulated immune cells. Genome Res. 2016;26(6):852–62.

118. Blow M, Futreal PA, Wooster R, Stratton MR. A survey of RNA editing in human brain. Genome Res . 2004 Dec 1;14(12):2379–87.

119. Lybecker M, Zimmermann B, Bilusic I, Tukhtubaeva N, Schroeder R. The double-stranded transcriptome of Escherichia coli. Proc Natl Acad Sci. 2014 Feb 25;111(8):3134–9.

120. Whipple JM, Youssef OA, Aruscavage PJ, Nix DA, Hong C, Johnson WE, et al. Genome-wide profiling of the C. elegans dsRNAome. RNA. 2015 Mar 24;21(5):786–800.

121. Bahn JH, Lee JH, Li G, Greer C, Peng G, Xiao X. Accurate Identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res. 2011 Jan 1;22(1):142–50.

122. Kim DDY, Kim TTY, Walsh T, Kobayashi Y, Matise TC, Buyske S, et al. Widespread RNA editing of embedded Alu elements in the human transcriptome. Genome Res. 2004;14(9):1719–25.

123. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, et al. The Dfam database of repetitive DNA families. Nucleic Acids Res. 2016 Jan 4;44(D1):D81–9.

124. Jolly C, Metz A, Govin J, Vigneron M, Turner BM, Khochbin S, et al. Stress-induced transcription of satellite III repeats. J Cell Biol. 2004 Jan 5;164(1):25–33.

125. Altemose N, Miga KH, Maggioni M, Willard HF. Genomic characterization of large

heterochromatic gaps in the human genome assembly. PLoS Comput Biol. 2014 May 15;10(5):e1003628.

126. Hall LL, Byron M, Carone DM, Whitfield TW, Pouliot GP, Fischer A, et al. Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies. Cell Rep. 2017;18(12):2943–56.

127. Brueckmann NH, Pedersen CB, Ditzel HJ, Gjerstorff MF. Epigenetic Reprogramming of Pericentromeric Satellite DNA in Premalignant and Malignant Lesions. Mol Cancer Res. 2018 Mar 1;16(3):molcanres.0477.2017.

128. Zalzman M, Falco G, Sharova L V, Nishiyama A, Thomas M, Lee S-L, et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature. 2010 Mar 24;464:858.

129. Dan J, Rousseau P, Hardikar S, Veland N, Wong J, Autexier C, et al. Zscan4 Inhibits Maintenance DNA Methylation to Facilitate Telomere Elongation in Mouse Embryonic Stem Cells. Cell Rep. 2017 Aug 22;20(8):1936–49.

130. Jack APM, Bussemer S, Hahn M, Pünzeler S, Snyder M, Wells M, et al. H3K56me3 Is a Novel, Conserved Heterochromatic Mark That Largely but Not Completely Overlaps with H3K9me3 in Both Regulation and Localization. PLoS One. 2013 Feb 22;8(2):e51765.

131. Casanova M, Pasternak M, El Marjou F, Le Baccon P, Probst A V, Almouzni G. Heterochromatin reorganization during early mouse development requires a single-stranded noncoding transcript. Cell Rep. 2013;4(6):1156–67.

132. Li Y, Banerjee S, Goldstein SA, Dong B, Gaughan C, Rath S, et al. Ribonuclease l mediates the cell-lethal phenotype of double-stranded RNA editing enzyme ADAR1 deficiency in a human cell line. Elife. 2017;6.

133. Ting DT, Lipson D, Paul S, Brannigan BW, Akhavanfard S, Coffman EJ, et al. Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. Science (80-). 2011;331(6017):593–6.

134. Gerdes P, Richardson SR, Mager DL, Faulkner GJ. Transposable elements in the mammalian embryo: Pioneers surviving through stealth and service. Genome Biol. 2016;17(1):1–17.

135. Weinberg MS, Morris K V. Transcriptional gene silencing in humans. Nucleic Acids Res. 2016;44(14):6505–17.

136. Kapranov P, St Laurent G, Raz T, Ozsolak F, Reynolds CP, Sorensen PHB, et al. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is “dark matter” un-annotated RNA. BMC Biol. 2010;8(1):149.

137. St Laurent G, Shtokalo D, Dong B, Tackett MR, Fan X, Lazorthes S, et al. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. Genome Biol. 2013;14(7):R73.

138. Rowe HM, Trono D. Dynamic control of endogenous retroviruses during development. Virology. 2011;411(2):273–87.

139. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 2015;522(7555):221–46.

140. Jachowicz JW, Bing X, Pontabry J, Bošković A, Rando OJ, Torres-Padilla ME. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat Genet. 2017;49(10):1502–10.

141. Stadler G, Chen JCJ, Wagner K, Robin JD, Shay JW, Emerson Jr CP, et al. Establishment of clonal myogenic cell lines from severely affected dystrophic muscles-CDK4 maintains the myogenic population. Skelet Muscle. 2011;1(1):12.

142. Jin Y, Tam OH, Paniagua E, Hammell M. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics. 2015;31(22):3593–9.

143. Silahtaroglu AN, Tommerup N, Vissing H. FISHing with locked nucleic acids (LNA):

Evaluation of different LNA/DNA mixmers. Mol Cell Probes. 2003;17(4):165–9.

144. Swanson EC, Manning B, Zhang H, Lawrence JB. Higher-order unfolding of satellite heterochromatin is a consistent and early event in cell senescence. J Cell Biol. 2013;jcb-201306073.

145. Wisdom R, Lee W. The protein-coding region of c-myc mRNA contains a sequence that specifies rapid mRNA turnover and induction by protein synthesis inhibitors. Genes Dev. 1991 Feb 1;5(2):232–43.

146. Weidensdorfer D, Stöhr N, Baude A, Lederer M, Köhn M, Schierhorn A, et al. Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. RNA. 2009;15(1):104–15.

147. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126(4):663–76.

148. Zeng F, Schultz RM. RNA transcript profiling during zygotic gene activation in the preimplantation mouse embryo. Dev Biol. 2005;283(1):40–57.

149. Clavería C, Giovinazzo G, Sierra R, Torres M. Myc-driven endogenous cell competition in the early mammalian embryo. Nature. 2013;500(7460):39.

150. Cliff TS, Wu T, Boward BR, Yin A, Yin H, Glushka JN, et al. MYC controls human pluripotent stem cell fate decisions through regulation of metabolic flux. Cell Stem Cell. 2017;21(4):502–16.

151. Jachowicz JW, Bing X, Pontabry J, Bošković A, Rando OJ, Torres-Padilla ME. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat Genet. 2017;49(10):1502–10.

152. Percharde M, Lin C-J, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, et al. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell. 2018;

153. Tanne A, Muniz LR, Puzio-Kuter A, Leonova KI, Gudkov A V, Ting DT, et al. Distinguishing the immunostimulatory properties of noncoding RNAs expressed in cancer

cells. Proc Natl Acad Sci. 2015;112(49):15154–9.

154. Camacho OV, Galan C, Swist-Rosowska K, Ching R, Gamalinda M, Karabiber F, et al. Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. Elife. 2017;6:e25293.

155. Hobson DJ, Wei W, Steinmetz LM, Svejstrup JQ. RNA polymerase II collision interrupts convergent transcription. Mol Cell. 2012;48(3):365–74.

Vita

Sean Shadle grew up in Ashland, Oregon. He studied molecular biology as an undergraduate at Portland State University. He was a research technician for two years at Oregon Health Sciences University studying mouse embryonic stem cell biology prior to arriving for graduate school in Seattle, at the University of Washington. He performed his PhD work under the mentorship of Doctor Stephen Justice Tapscott.