DUX4 expression in development versus disease:

DUX4 expression in spermatogenesis, placentation, cancer and FSHD

Andrew A. Smith

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Stephen J. Tapscott, Chair

Robert Bradley

Robert Eisenman

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

DUX4 expression in development versus disease:

DUX4 expression in spermatogenesis, placentation, cancer and FSHD

Andrew A. Smith

Chair of the Supervisory Committee:

Stephen J. Tapscott

Department of Neurology

DUX4 is a pioneer transcription factor that drives zygotic gene activation (ZGA) in the early embryo. DUX4 is also the causative gene of facioscapulohumeral dystrophy (FSHD). Recently, DUX4 has been credited for driving the totipotent program in naïve embryonic stem cells, and for driving immune evasive properties in an assortment of cancers. Thus, DUX4 has its fingers in a lot of pots, but its mode of expression in these different contexts is not well understood and there hasn't been a thorough study of DUX4 expression across different tissue types in development.

The mouse ortholog of DUX4, mouse Dux, largely mirrors the function of its human counterpart and, in these chapters, I discover mouse Dux expression in the adult male germline as well as in the trophoblast giant cells (TGC) of the developing placenta, which I propose is driving maternal immune tolerance during embryo implantation. I then define the transient nature of DUX4 expression in rare cells across an assortment of human cancers and show how

DNA-damaging agents can promote this DUX4+ state, which then leads to an immune evasive phenotype.

The expression of DUX4 in the context of FSHD is relatively well-defined, but these rules don't appear to apply to DUX4 expression in development and cancer. I show that DUX4 RNA transcripts are fundamentally different in development and cancer from those driving DUX4 in FSHD. Even though these transcripts all end up driving expression of the same DUX4 protein, this indicates different molecular mechanisms may be activating transcription of DUX4 in a developmental context. Furthermore, I demonstrate that DUX4 can drive transcription of itself. Although other master regulator genes have been reported to autoregulate, this is novel for DUX4.

I hope these findings aid the field in its pursuit of understanding the modes and consequences of DUX4 expression in wanted and unwanted contexts.

DUX4 expression in development versus disease:
DUX4 expression in spermatogenesis, placentation, cancer and FSHD

Andrew A. Smith

TABLE OF CONTENTS:

Chapter 1: Introduction

1.1 Early Embryo Development and Lineage Commitment

Human oocytes are filled with maternally deposited RNAs and protein which carry out all molecular functions of the early embryo from the zygote through the first two cell divisions. Once these maternally deposited RNAs are degraded, the embryo begins expression of its own genes for the first time, a process called zygotic gene activation (ZGA)[1]. This transition occurs at the 4-cell stage in humans and at the 2-cell stage in mice, also known as the cleavage stage, and is defined by expression of a specific suite of early embryonic genes. As the embryo continues to grow, its cells begin to diverge in gene expression as they commit to specific tissue lineages.

The first lineage commitment happens simultaneously with a morphological change in the embryo where the berry-like arranged cells (morula) transition to a ball-within-a-sphere arrangement (blastocyst). The outer layer of blastocyst consists of cells called trophoblasts, which will give rise to the extraembryonic tissues (placenta and amnion) whereas the cluster of cells inside are called the inner-cell mass, will give rise to all embryonic tissues[2].

While trophoblasts differentiate into extraembryonic tissues to facilitate implantation, maternal immune tolerance, and access to nutrients, the cells of the inner cell mass will eventually make a second lineage commitment to one of the 4 germ layers: ectoderm (skin and nervous system), mesoderm (muscle, blood, bone, and internal organs), or endoderm (respiratory and gastrointestinal systems)[3], and I dub germ cells (sperm or egg) the 4th germ layer. These initial cell fate decisions only represent the beginning, as these cells continue to make lineage and cell fate commitments until they differentiate into all of the specific cell types of the fully formed body.

Despite differentiation and commitment to specific lineages, some cells of the adult body remain partially undifferentiated and reside in tissues as cellular reservoirs needed for growth, tissue regeneration, and injury repair[3,4]. The cells of these reservoirs are called adult stem cells, and typically retain the ability to differentiate into multiple types of cells as needed. Although these cell fate decisions underlie both development as well as regeneration and growth, there remains much we don't understand regarding how stem cells navigate these decisions. The stem cell field currently strives to better understand and capitalize on the power of stem cells in hopes to build, maintain, and repair entire organisms,

1.2 Stem Cells and Potency

Stem cell biologists refer to the ability of a cell to contribute to different types of cells as a grade of potency[5]. Stem cells in the adult body that contribute to tissue repair and regeneration are considered to be multipotent –the ability to give rise to a limited number of different cell types which all lie within a specific cell lineage. Stem cells from the inner cell mass of the early embryo are considered to be pluripotent –the ability to give rise to all cell types of the body. And stem cells from the cleavage stage embryo and earlier are considered to be totipotent –the ability to give rise to all embryonic and extra-embryonic tissues. The totipotent state is of particular interest to stem cell biologists because these cells have unlimited differentiation potential for therapy and may hold new opportunities for reproductive biology. Of note, totipotent stem cells are capable of giving rise to germline stem cells, which may represent a source of continuous production of gametes such as sperm and eggs.

In vitro culture of stem cells in the laboratory has furthered our understanding of these early cell fate decisions and holds promise for downstream cellular based therapies *in vitro* and *in vivo*. Currently, the stem cell field has reliable culture methods for pluripotent stem cells and for many types of multipotent stem cells, whereby we can maintain them in their undifferentiated states while retaining their ability to differentiate towards specific cell types. Despite advances in culture conditions for other such types of stem cells, totipotent and germline stem cells have been two of the most elusive cells to study as we cannot yet reliably culture these while maintaining these cell states. In 2022, Taubernaud-Stowers *et al*.[6] published that within their naïve pluripotent cultures, they identified subpopulations of cells expressing a ZGA and 8-cell-like transcriptional programs, indicating they may have rare totipotent cells in certain culture conditions. The presence of these rare subpopulations expressing a ZGA and 8-cell-like transcriptional program demonstrated that *in vitro* culture of cells retaining totipotency may be possible although further work is required to understand and expand this population.

1.3 Cancer Stem Cells

Many cancers respond well to clinical treatment and enter into complete remission or periods of remission. However, some cancers return, often with new increased chemoresistance, immune evasion, and metastasis capability[7,8]. These cancers frequently consist of pathologically undifferentiated tumors, which almost universally have a worse prognosis[9]; these undifferentiated tumors characteristically express early embryonic genes, suggesting the activation of a gene expression program associated with increased

differentiation potency. This led to the notion that there may be a subpopulation of cells within tumors which function as cancer stem cells, regenerating the cancer if they survive treatment.

This cancer stem cell hypothesis had a short run of popularity. One tenet of this theory is that chemotherapies target the highly proliferative and more differentiated cancer cells, which form the bulk of the tumor, leaving behind the slower-dividing stem cells capable of regenerating the tumor. This provides an explanation and possible therapeutic target for treatment approaches as it explains the reoccurrence after drug-induced bottleneck as well as cancer cell plasticity and potency. This idea would also explain the long-recognized correlation between cancers and the expression of early embryonic gene programs. If cancer cells can re-express genes meant only for the development of the early embryo they can gain access to potent (undifferentiated) cell states with immune evasive phenotypes and generate a broader spectrum of cell types depending on the environmental assaults they face. Should the cancer stem cell hypothesis be true, this would demonstrate an arms race between cancer evolution and cancer treatment. If a subpopulation of cancer cells is both resistant to chemotherapies and highly potent, then it's easy to imagine that they are able to give rise to the next stage of cancer progression after treatment.

The cancer stem cell hypothesis is currently debated because no true cancer stem cell has been found. Furthermore, cancer cells often exhibit phenotypic plasticity[10]; many cancer cells will change their gene expression profile when their environment changes, they are chemically assaulted, or they become stressed (i.e. starved of nutrients). This intrinsic plasticity allows cancers to quickly adapt to molecular challenges. This falls in alignment with known cancer heterogeneity concerns, where cells within specific regions of tumors are to significantly differ from one another. The presence of a cancer stem cell, capable of adapting to ever changing environments within the host body, would easily explain cancer heterogeneity. Despite this, many cancer biologists argue that the majority of cancer cells are stem cell-like in nature compared to cells of adjacent healthy tissue – arguing that intrinsic plasticity is a characteristic of all cancer cells, rather than a rare subpopulation of cancer stem cells.

In pursuit of differentially regulated genes across a broad spectrum of cancers, a recent study identified DUX4 as being uniquely miss-expressed in a broad range of cancers[11]. This is of particular interest because DUX4 is an early embryonic transcription factor proposed to drive the totipotent program in early embryonic stem cells. Moreover, this study reported that DUX4 expression in cancer reduced surface expression of the major histocompatibility complex class I (MHC-I –a foundational molecule in immune surveillance[12]) and correlated with reduced

success of immunotherapies, leading to worse clinical outcomes. In this dissertation I describe my research on the nature of DUX4 expression in cancer cells.

1.4 DUX4

DUX4 belongs to the double-homeodomain gene (*DUX*) family which is unique to placental mammals. The *DUX* family of genes have been categorized into 4 classes (*DUXA, DUXB, DUXB-like*, *DUXC*) and phylogenetic analysis suggests that primate specific *DUX4* and murine *Dux* (*mDux*) originated from retrotransposition of the *DUXC* mRNA[13-15]. The ancestral *DUXC* gene has been lost in primates and rodents but has been retained in other eutherian mammalian genomes. All *DUX* genes encode proteins with N-terminal double homeodomains (that bind to DNA) but only the DUXC branch harbors a highly conserved C-terminal transactivation domain (CTD) required for target gene expression. Both mouse Dux and human DUX4 are potent transcription factors which drive expression of hundreds of genes[16]. Mouse Dux expression in human cells, or human DUX4 in mouse cells, shows a significant overlap in transcriptional targets[16], therefore these genes are considered orthologs and we can derive insights into the function of DUX4 by studying mouse Dux.

The *DUX* family of genes have been found to be transiently expressed in pre-implantation embryos across many species. Both mouse Dux and human DUX4 are briefly expressed in the cleavage stage embryo (2-cell in mouse and 4-cell in human) and drive the major arm of zygotic genome activation (ZGA)[13,17,18]. In embryonic and induced pluripotent stem cell (ESC/iPSC) cultures, transient expression of the ZGA program occurs in a rare population of cells[19]. Expression of DUX4/Dux in this rare population drives ZGA gene expression thought to recapitulate the naïve totipotent state of the pre-implantation embryo[6,17].

DUX4 is subsequently repressed and epigenetically silenced in most somatic tissues, apart from the testis, where DUX4 is expressed in the germ cells[20]. Many genes known to be regulated by DUX4 in the early embryo are also expressed in the testis, suggesting that DUX4 may regulate a similar gene program in the male germline[21]. The exact function of DUX4 in the testis is still not fully understood and whether mouse Dux is also expressed in the testis remains unknown. As we gain insights into the developmental functions of DUXC family members, we are supported by a strong foundation of knowledge of DUX4 from its discovery as the causative gene in Facioscapulohumeral dystrophy (FSHD).

FSHD is the third most common muscular dystrophy, the onset of which is characterized by weakness in muscles of the face, shoulder blades, and upper arms[22]. Misexpression of

DUX4 in muscle cells was proposed to be the driver of this disease in the 1990s, but the specific mechanics of DUX4 expression in FSHD muscle cells was not laid out until 2010.

1.5 The discovery of DUX4 in FSHD

In 1990 the causative gene for Facioscapulohumeral dystrophy (FSHD) was localized to chromosome 4q [23-25], and soon after truncations in the D4Z4 macrosatellite on chromosome 4 q35 (4q35) were linked to FSHD [26-28]. Analysis of the truncated FSHD D4Z4 macrosatellite revealed that it contained an open reading frame (ORF) which encoded for a 391 amino acid protein with two homeodomains, DUX4 [29]. In the early 2000s, two common distal D4Z4 sequence variants were discovered, 4qA and 4qB and studies determined that FSHD was associated with 4qA variant only [30,31], however, the role of DUX4 in FSHD was still speculative.

In 2007, DUX4 was finally detected in FSHD patient myoblasts and overexpression of DUX4 in myoblasts proved to be cytotoxic [32]. Others later corroborated DUX4 expression within FSHD patient muscle cells and in 2010, the molecular basis for DUX4 driving FSHD was laid out as having both a hypomethylated D4Z4 and a permissive 4qA allele, which contains a poly A signal downstream of the terminal D4Z4 unit[33]. Although the D4Z4 contains many DUX4 open reading frames (ORFs), none contain a poly A signal. In FSHD muscle cells, DUX4 transcripts utilize this distal poly A signal to stabilize DUX4 mRNA, however not all 4q alleles contain a known poly A signal.

Since then, the DUX4 transcriptional program has been largely defined in the context of FSHD. And it's this FSHD molecular signature that is now being recognized in the early embryo, stem cell culture and cancer contexts.

1.6 Models of DUX4 expression and the DUX4 molecular signature

While DUX4 has been principally characterized as the disease-causing gene involved in FSHD pathophysiology, expression of DUX4 and its target genes in normal developmental biology[16,18,20] and in a subset of cancer types[11] has recently emerged. Each D4Z4 unit within the *DUX4* retrogene array contains an intact open reading frame (ORF), and in FSHD, aberrant DUX4 expression comes from the terminal repeat carrying a permissive 4qA allele. How DUX4 is endogenously expressed in other contexts when lacking a permissive chromosome 4 allele is still under investigation.

Aberrant expression of DUX4 is associated with several diseases besides FSHD, including B-cell acute lymphoblastic leukemia (B-ALL)[34] and Ewing-like Small Round Sarcomas[35],

although these cancers don't express a full-length DUX4 protein (DUX4-FL); these express a DUX4 fusion protein which doesn't activate the canonical DUX4 transcriptional program. While DUX4-FL expression has recently been associated with an assortment of cancers, currently there are no known cancer models that express endogenous DUX4-FL and thus the full DUX4 transcriptional program.

DUX4 has diverse functions which include regulating both protein-coding genes and non-coding regions of the genome. DUX4 activates genes characteristic of germ cells and the early embryo, including the PRAME, TRIM, MBDL, ZSCAN, RFPL, KHDC, and primate-specific FAM90 gene families[21]. Other DUX4 target genes include transcription factors *DUXA*, *DUXB*, and *LEUTX*, epigenetic regulator *KDM4E,* and histone variant *H3Y*. DUX4 also drives expression of a handful of repetitive elements and retrotransposons including Alu elements, long interspersed nuclear elements (LINE1), pericentromeric human satellite II (HSATII) repeats, mammalian apparent LTR-retrotransposons (MaLR), and endogenous retrovirus elements (ERVs).

In 2016, Eckersley-Maslin et al.[19] reported that a subpopulation of pluripotent mouse embryonic stem cells in culture was cycling through a transient totipotent-like state as marked by expression of *MERVL* and Zscan4, and it is now known that mouse Dux drives expression of *MERVL* and *Zscan4*[21]. In 2017, Hendrickson et al. reported that DUX4 and mouse Dux drive the ZGA program in the cleavage stage embryo. In 2022, Taubenshmid-Stowers *et al*.[6] reported that subpopulations of cells in naïve pluripotent human stem cell cultures were transiently expressing ZGA, 8-cell-like and totipotent signatures. They defined the ZGA signature as expressing a collection of 23 genes (including *MBD3L2, ZSCAN4*, *DUXA*, *LEUTX*, *KDM4E, H3Y, and PRAME and TRIM* family members) all of which are direct transcriptional targets of DUX4.

Although the aberrant expression of DUX4 in muscle cells leads to cell death, DUX4 appears to be transiently expressed in other developmental contexts leading to cellular reprogramming to an early embryonic-like state. The mechanistic differences in DUX4 expression between these developmental contexts, where cells tolerate DUX4, and FSHD contexts, where cells die, begs to be defined. Herein, we 1) explore the developmental expression patterns of the DUX4 orthologue mouse Dux, 2) define the nature of endogenous DUX4-FL expression in an assortment of cancer cell lines, 3) identify differences in DUX4 transcripts between FSHD and other contexts, 4) propose a direct positive regulator of DUX4 and 5) report on the odd biology of DUX4-coding RNAs.

Chapter 2: DUX4 in development: The mouse ortholog of DUX4 is expressed during spermatogenesis and placentation.

This chapter contains unpublished work

2.1 Introduction

To date, there has not been a thorough study of mouse Dux (mDux) expression across tissues through development. To elucidate the functions of mDux in development, a mDux knockout mouse strain was generated and, in 2019, Chen and Zhang reported that mDux was completely dispensable during embryogenesis[36]. Since then, two studies have reported that this mDux knockout strain exhibits a significantly increased rate of late embryonic mortality[37,38], supporting the notion that loss of mDux is partially embryonic lethality. In this last study, Bosnakovski et al.[38] also mentioned that mouse Dux activates trophectoderm genes and that the observed embryonic lethality resulting from loss of mouse Dux most likely happens post-implantation, leading to the notion that Dux could be playing a role in implantation.

In this current study, we generated an antibody to mouse Dux, surveyed tissues of interest for mouse Dux expression, and report that mouse Dux is expressed in the germ cells of the adult male testis as well as trophoblast giant cells (TGC) of the developing placenta. These results will help further our understanding of the roles of the DUX family of genes during development and tissue homeostasis.

2.2 Results

2.2.1 Monoclonal antibodies reliably detect mouse Dux protein

Rabbits were immunized with mouse Dux peptides to generate Dux antisera. To evaluate the specificity of anti-mDux antisera we genetically engineered the C2C12 mouse myoblast cell line[39] to express a doxycycline-inducible *mDux* transgene (C2C12-i*mDux*) and treated these cells with doxycycline for 16hrs. Western blotting and staining with the mDux antisera detected a 100kDa band that was specific to the +dox-condition (Fig 2.2.1A) . Immunofluorescence with this mDux antisera stained the nuclei of C2C12-i*mDux* cells only in

the +dox condition (Fig 2.2.1B). These data support the specificity of our polyclonal antibody to mDux.

Three different monoclonal mDux antibodies (B7, E9, F1) were then generated and validated for immunohistochemical (IHC) detection of mDux in FFPE sections of in C2C12-i*mDux* cell pellets (Fig 2.2.1C). Anti-mDux monoclone E9 was specific for mDux with the least background staining and was selected for all further analyses.

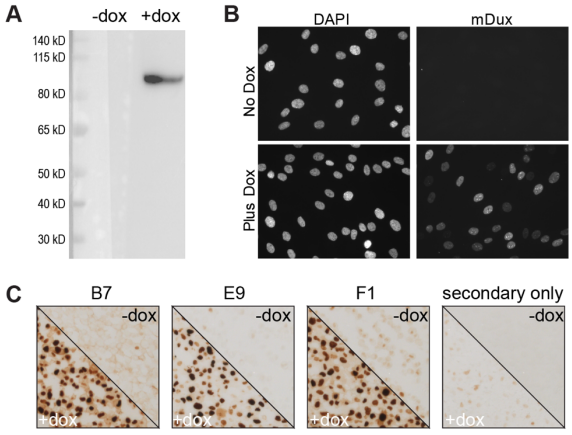


Figure 2.2.1: Generating an antibody to mouse Dux. **(A)** Western blot on C2C12-i*mDux* protein stained with anti-mouse Dux polyclone. **(B)** IF in C2C12-i*mDux* stained anti-mouse Dux polyclone. **(C)** FFPE C2C12-i*mDux* cell pellets stained with anti-mDux monoclones.

2.2.2 mDux is expressed in adult male germ cells

To determine where mDux is expressed in the adult mouse, we used RT-qPCR to survey relative *mDux* RNA abundance in select tissues. Compared to the housekeeping gene *Rpl27*, *mDux* RNA was most abundant in the adult testis (Fig 2.2.2A). Analysis of testis RNA-sequencing data across developmental time[40] demonstrated that *mDux* and mDux target gene expression in mouse testis increases with the onset of sexual maturity ~4wks of age (Fig. 2.2.2B). To confirm this, we harvested testis RNA from 1-, 5-, 8-, and 64-week-old mice and RT-qPCR demonstrated that both *mDux* and its transcriptional target *Zscan4c* was most highly expressed in sexually mature mouse testis (Fig. 2.2.2C). IHC for mDux in adult testis stained germ cell nuclei (Fig 2.2.2D). These data support mDux expression in the germ cells of adult male mice.

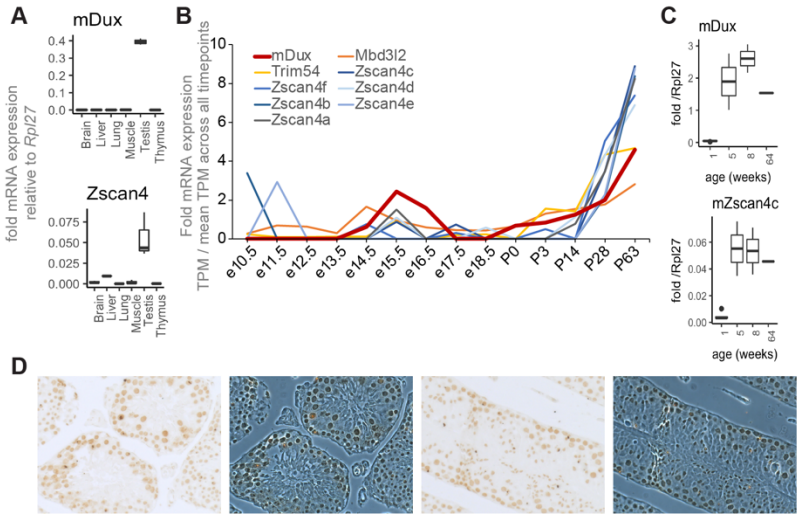


Figure 2.2.2. Mouse Dux is expressed in the germ cells of adult testes. **(A)** RTqPCR tissue survey. **(B)** Testes RNAseq across developmental time IHC mDux in adult testis. **(C)** RTqPCR for mDux and mZscan4c in mouse testis versus age. **(D)** Immunohistochemistry for mDux in FFPE sections of adult mouse testis.

2.2.3 Loss of mDux is lethal post implantation.

While breeding the C57blk/6 mDux knockout mouse strain, we observed reduced litter sizes when we bred heterozygous mDux+/- mice with homozygous mDux-/- mice (Fig 2.2.3A). In particular, we noticed that it was difficult to generate mDux-/- mice, so we plotted all cumulative litter genotypes from each genetic cross as a percentage of the total pups born, and confirmed that, compared to expected ratios of pups (dashed bars) there was a significant reduction in the observed ratio of mDux-/- pups generated (Fig 2B, chi-squared = 0.0001). When we dissected the uteral horn from pregnant mDux+/- females crossed with mDux+/- males, we observed pups being reabsorbed (Fig 2.2.3C, arrow). These data suggest that mDux may be playing a role in post-implantation development (also proposed in[38]).

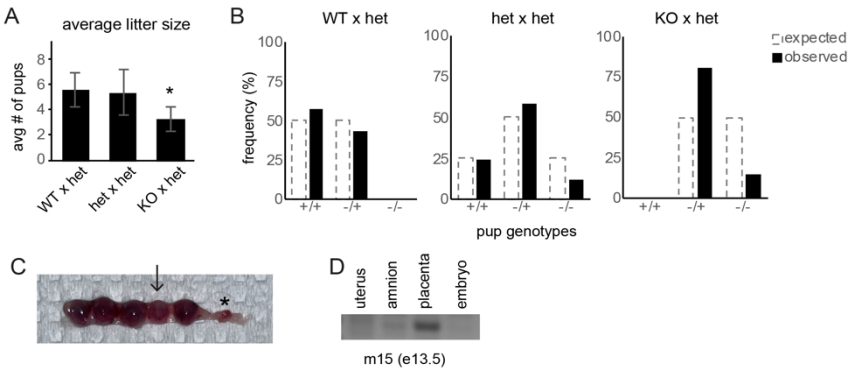


Figure 2.2.3. Loss of mouse Dux is embryonic lethal.
(A) Average litter sizes from crosses of C57blk/6 mDux+/- mice. n >7 litters per parent genotypes. **(B)** Frequency of cumulative pup genotypes from denoted crosses. n >35 pups per parent genotypes. **(C)** Gross morphology of uteral horn from a mDux+/- female crossed with a mDux+/- male. * = ovary. **(D)** Agarose gel of RT-PCR for mDux in embryonic and extra embryonic tissues.

2.2.4 mDux is expressed in the placenta

To evaluate if mDux is expressed during embryogenesis, we micro-dissected embryonic tissue from extra embryonic tissue and from the maternal uterus of e13.5 embryos and extracted the RNA from each. RT-PCR revealed mDux expression in the placenta of e13.5 embryos (Fig 2.2.3D).

2.2.5 mDux is expressed in trophoblast giant cells of the developing placenta

To evaluate which cells in the developing placenta are expressing mDux, we setup timed breeding cages and generated FFPE samples at different timepoints. In e6.5 WT embryos mDux was detected in the nuclei of rare emerging trophoblast giant cells (TGC) apical to the embryo (Fig 2.2.5A). In e9.5 WT embryos mDux was detected the nuclei of TGC throughout the spongiotrophoblast (Fig 2.2.5B).

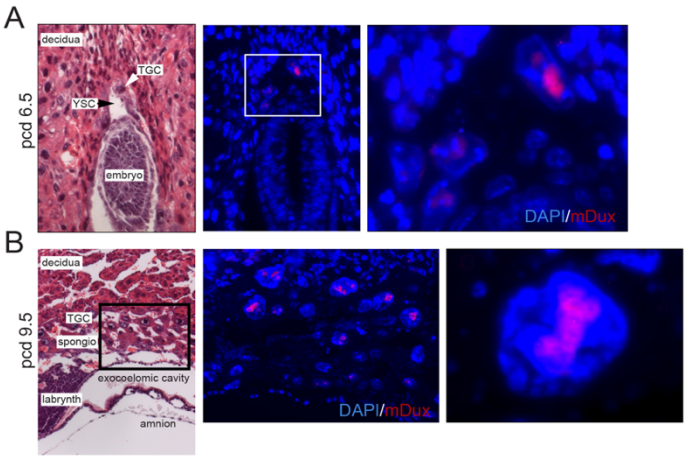


Figure 2.2.5. Mouse Dux is expressed in trophoblast giant cells
(A-B) H&E staining of C57blk/6 embryos paired with immunofluorescence for mDux at **(A)** 6.5 days post-conception and **(B)** 9.5 days post conception, with higher magnification in right panel.

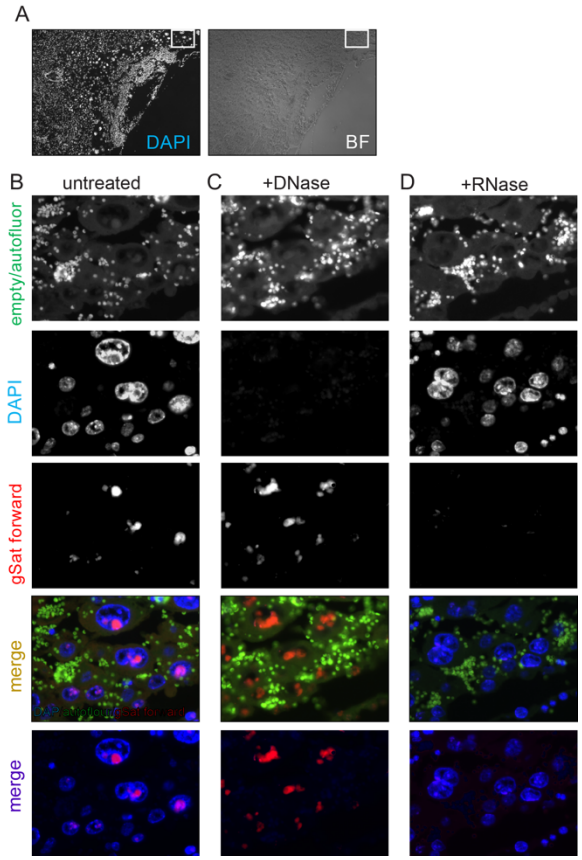


Figure 2.2.6. Major Satellite repeat is expressed in trophoblast giant cells
(A) DAPI stain and bright field 5X images of e9.5 C57blk/6 developing placenta.gSat FISH in e9.5 C57blk/6 developing placenta. **(B)** gSat FISH in e9.5 C57blk/6 developing placenta treated with DNase. **(C)** gSat FISH in e9.5 C57blk/6 developing placenta treated with RNase.

2.2.6 gSat is expressed in TGC

To confirm mDux expression in these TGC, we stained sections of e9.5 embryos for expression of the mouse major satellite repeat (gSat), which is a transcriptional target of mDux. To generate comparable images we focused on a region abundant with TGC, in the spongiotrophoblast zone adjacent to the labyrinth (Fig 2.2.6A). In e9.5 embryo sections, we detected gSat in the nuclei of TGC (Fig 2.2.6B). To confirm that this was staining RNA and not DNA, we treated adjacent sections with either DNase or a cocktail of RNases. In e9.5 embryo sections treated with DNase, we still detected gSat in the nuclei of TGC (Fig 2.2.6C). However, in the in e9.5 embryo sections treated with RNase, we observed loss of gSat staining in the TGC (Fig 2.2.6D). These data confirm that gSat RNA is expressed in the TGC of the developing embryo.

Both human DUX4 and mouse Dux have been reported to reduce MHC-I surface expression and to block responses to certain immune stimuli[41]. If Dux placental mouse Dux expression is negatively regulating maternal immune responses during implantation, we postulated that the syngeneic nature of the C57BL/6 inbred strain might be reducing the necessity of this mechanism and may lead to increased embryonic viability of the Dux -/- mouse.

2.2.7 Crossing background strains increases embryonic lethality of the Dux null mice

To determine the necessity of mouse Dux during embryogenesis in a different strain, we backcrossed the Dux KO genotype in to the 129 strain 5 times and bred these mice with the BL/6 strain.

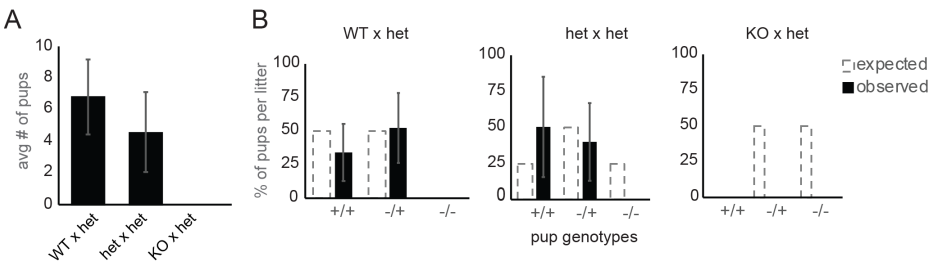


Figure 2.2.7. Backcrossing the Dux knockout line in to 129 strain increases embryonic lethality of Dux-/-
(A) Mean number of pups per litter in denoted crosses. (B) Expected versus observed percentages of each genotype per litter, expressed as mean for denoted crosses. n > 6 litters per cross, except for DUX null crossed with Dux +/-, which never produced a viable litter; error bars = standard deviation.

Crossing 129 Dux+/- mice with BL/6 Dux-/- mice never resulted in a viable litter, regardless male and female background strain (Figure 2.2.7A). This could be interpreted as 100% embryonic lethality, however this could also be interpreted as potential sterility or mating behavioral shortcomings. Breeding Dux+/- with WT mice of the opposite strain did not affect litter sizes and the numbers of Dux+/- offspring agreed with canonical gene inheritance models, however, crossing Dux+/- mice from different background strains, slightly reduced litter sizes and never resulted in a Dux null mouse (Figure 2.2.7B). These data suggest that, while the Dux+/- is tolerated in outbred mice, a Dux null genotype is 100% lethal in outbred mice.

2.3 Discussion

In this chapter, we demonstrated the expression of mouse Dux in the germline of adult male mice and in trophoblast giant cells (TGC) of the developing murine placenta. The expression pattern of mouse Dux in adult testis mirrors the pattern observed in human testis[20], however the functional role of mouse Dux in the testis is not understood and mDux null male mice do appear to be fertile, in agreement with previous studies[13,36,38]. TGC also stained positive for major satellite repeat (a direct transcriptional target of mDux) which further bolsters these

findings and, regarding the necessity of mouse Dux during embryogenesis, we demonstrated that outbreeding BL/6 mice increases embryonic lethality of the Dux knockout strain, potentially to 100% lethally, however, the molecular role of mDux in the TCG and embryo survival is not known.

During embryo implantation, the modulation of the mother's immune system is known to be important for successful development[42]. Because the human orthologue of mDux, DUX4, has been proposed to modulate immune responses[41], it is tempting to postulate that mDux plays a role in maternal immune modulation during implantation. However, the connection between mDux and immune modulation, has yet to be established. Future studies on the role of mDux in development should aim to define the molecular effects of mDux expression in the adult male germ line as well as in TGC.

2.4 Materials and Methods

Mice

C57bl6 and 129 mice were housed, pred and sacrificed according to our approved IACUC protocol

Cell Culture

MB135 and C2C12 myoblast lines were cultured in F-10 10% FBS 1% penicillin/streptomycin. Primary cultures of TGC were cultured in RPMI 20%FBS 1% penicillin/streptomycin. All cultures were maintained at 37C in 5%CO2

Tissue embedding

Mice were anesthetized with 3% isoflurane until unresponsive, then perfused with 1% heprin in PBS followed by 4%PFA in PBS. tissues of interest were then dissected, cubed, and incubated in B+ fixative 16h.

Western Blot

Cells were harvested in RIPA buffer containing protease and phosphatase inhibitors and sonicated on low for 5 minutes. The whole cell lysate was centrifuged at 12,0000 RCF for 10 minutes and the supernatant transferred to a new tube. Protein quantification was performed using the Pierce BCA Protein Assay Kit. Either 5ug or 10ug of whole cell lysate was heated in LDS Sample Buffer with 2.5% β-mercaptoethanol at 70°C for 10 minutes with shaking. For SDS-PAGE, proteins were loaded onto a 4-12% NuPAGE Bis-Tris gel and run at 100V for about 2 hours in NuPAGE MOPS SDS Running Buffer with 300uL NuPAGE antioxidant. Proteins were transferred to a 0.2µm PVDF membrane at 30V for 1 hour in NuPAGE Transfer Buffer with 10% methanol. Membranes were blocked in 5% milk in TBST for 1 hour, then incubated at 4°C overnight with rocking. Blots were washed three times (15 minutes each) with TBST and incubated with HRP-conjugated secondary antibody for 1 hour. Blots were washed three times (15 minutes each) with TBST, and bands were detected via chemiluminescence with either SuperSignal West Pico PLUS Substrate, SuperSignal West Femto Substrate, or SuperSignal West Atto Substrate.

Immunofluorescence and Immunohistochemistry

Cultured cells were fixed in 2% paraformaldehyde 10min, and washed with PBS. Cells were then permeabilized in 0.1% Triton X-100 in PBS for 10 minutes at room temperature. Cells were then incubated overnight at 4°C with DUX4 [E14-3] primary antibody (1:400 dilution) or H3Y1/2 [8H6-2111] (1:333 dilution) and washed with PBS three times for 10 minutes each. Cells were incubated in TRITC-conjugated donkey anti-rabbit secondary antibody (1:666 dilution) and TRITC-conjugated donkey anti-rat secondary antibody (1:666 dilution) for 3 hours at room temperature. Cells were washed with PBS, stained with DAPI for 10 minutes at room temperature, and washed with PBS. Cells were imaged with a fluorescence microscope.

Reverse transcription-quantitative PCR (RT-qPCR)

Tissue was pulverized with a mortar and pestle after snap freezing in liquid nitrogen. RNA from pulverized tissue and cultured cells was harvested with the NucleoSpin RNA Kit (Takara) according to the manufacturer's protocol. RNA quality was verified by NanoDrop 2000 (Thermo Scientific). RNA was treated with DNaseI, Amplification Grade (Invitrogen). Reverse transcription was performed in a 20uL reaction: 200-1000ng whole RNA, 1uL dNTP (10mM), 1uL oligo dT primer (10mM), 4uL 5x SSIV Buffer, 1uL DTT (100mM), 1 uL RNaseOUT, and 1uL SSIV RT enzyme. Thermal cycling conditions for reverse transcription were as follows: 50°C for 40 minutes, 55°C for 30 minutes, and 80°C for 10 minutes. Complementary DNA (cDNA) was treated with 1uL of RNaseH and incubated at 37°C for 20 minutes, then diluted 1:5 or 1:4 with RNase-free H2O. Quantitative real-time PCR (qPCR) was performed on the QuantStudio™ 7 Flex Real-Time PCR System in a 10uL reaction: 2uL cDNA, 5uL 2x iTaq Universal SYBR Green Supermix, 0.3uL forward and reverse primer (10uM), and 2.4uL H2O. qPCR primers were synthesized by Integrated DNA Technologies (IDT) and are listed in Star Methods table. Thermal cycling conditions for qPCR were as follows: 50°C for 2 minutes and 95°C for 10 minutes; 40 cycles of 95°C for 15 seconds and 60°C for 60 seconds. For each Biological replicate, qPCR reactions were run in technical triplicates, including -RT controls. Median CT values of the technical triplicates were used for analysis. Gene expression was normalized to the housekeeping gene RPL27 (ribosomal protein L27). P-values were calculated with an unpaired, one-tailed student's t-test assuming unequal variance (type 3).

Generation of cell lines with dox-inducible codon altered Dux transgene.

Polyclonal C2C12-*imDux*^{CA} and MB135-*imDux*^{CA} cell lines were generated with lentiviral transduction, selected, and maintained in 3µg/mL puromycin as previously described for MB135-*iDUX4*^{CA} cells in Jagannathan et al., 2016.

Chapter 3: DUX4 expression in cancer induces an early embryonic totipotent program with reduced MHC-I expression.

This chapter contains unpublished work.

Summary

The transcription factor DUX4 regulates a portion of the zygotic gene activation (ZGA) program in the early embryo. Many cancers express DUX4 but it is unknown whether this generates cells similar to early embryonic stem cells. Here we identify cancer cell lines that express DUX4 and show that DUX4 is transiently expressed in a small subset of the cells. DUX4 expression activates the ZGA transcriptional program, the subsequent 8C-like program, and markers of early embryonic lineages; while suppressing steady-state and interferon-induced MHC class I expression. Although DUX4 was expressed in a small number of cells under standard culture conditions, DNA damage or changes in growth conditions dramatically increased the fraction of cells expressing DUX4 and its downstream programs. Demonstrating that the transient expression of endogenous DUX4 in cancer cells induces a metastable early embryonic stem cell program and suppresses antigen presentation has implications for cancer growth, progression, and immune evasion.

Keywords: DUX4, Dux, zygotic gene activation, cancer, MHC, HLA, stem cell, totipotent

3.1 Introduction

Many studies have identified intra-tumor cell heterogeneity as being associated with chemotherapy resistance, metastasis and cancer progression[43-46]. However, it has been difficult to identify molecular markers of functional sub-populations that are shared across different cancers, leading to the notion that multiple tumor-specific genetic and environmental interactions combine to generate cancer cell diversity. For example, early embryonic transcriptional programs have long been associated with cancers[9], but specific embryonic transcriptional programs common to multiple types of cancer have not been identified. Our recent identification of DUX4 expression in many different cancers[11] raises the possibility that it might generate a cancer cell population similar to early embryonic stem cells in a diverse set of cancers. The transcription factor DUX4, and its mouse ortholog Dux, are normally expressed in the early embryo, where the brief expression of DUX4/Dux regulates a portion of the initial

zygotic gene activation (ZGA) at the 2-cell (2C) stage in mice and the 4-cell (4C) stage in humans[13,16,18], including the activation of classes of retrotransposons and the HSATII and GSAT pericentromeric repeats in human and mouse, respectively. Although DUX4 and mouse Dux activate a the ZGA program, some groups have reported that mouse Dux is not essential for early embryo development [13,36,38]and one group has suggested the same for human DUX4 ((PMC8990217)). In ESC/iPSC cultures the transient expression of the ZGA program in rare cells establishes a naïve, or totipotent, subpopulation[6,17,19,47], also referred to as 2C-like or 8C-like in mouse and human cells, respectively; and subsequent studies showed that the ZGA program is driven by the transient expression of DUX4/Dux in these cells[6,48]. It remains unknown whether the expression of DUX4 and subsets of its target genes in different cancers represents low level transcriptional noise, or whether these cancers contain cells expressing a coherent ZGA transcriptional network with the emergence of subsequent embryonic programs.

DUX4 was initially characterized through the study of its aberrant expression in skeletal muscle which causes facioscapulohumeral muscular dystrophy (FSHD)[20,33,49]. The DUX4 retrogene is present in each unit of the D4Z4 macrosatellite repeat array located in the subtelomeric regions of both chromosome 4 and chromosome 10[29]. FSHD1, the most common form of FSHD, is caused by deletion of a subset of D4Z4 repeats creating an array of 1-10 D4Z4 units that decreases the epigenetic silencing of the DUX4 locus in somatic cells, as measured by decreased DNA methylation and repressive histone modifications, whereas FSHD2 is caused by mutations in epigenetic modifiers necessary for epigenetic repression of the *DUX4* locus[49-51]. The decreased epigenetic repression of the DUX4 locus results in the occasional burst of DUX4 expression in skeletal muscle occurring in ~0.1-1% of FSHD myoblasts in culture[20,52].

In this study we identified cancer cell lines that express the full-length *DUX4* mRNA (DUX4fl) and show that DUX4 protein and its target gene H3Y are expressed in ~-0.1% of cells under standard tissue culture conditions. The DUX4-positive cells were often in small clusters of two or more cells, consistent with a brief expression of DUX4 followed by cell divisions. Single-cell RNA-sequencing showed that the transient expression of DUX4 in the cancer cells activates the ZGA and 8C-like transcriptional programs, as well as markers of early embryonic lineages; while suppressing MHC Class I (MHC-I) mRNAs. DNA damage resulted in a p53-dependent dramatic increase in the percentage of DUX4 expressing cells and subsequent ZGA gene expression. Together, our study shows that the transient expression of endogenous DUX4 in cell lines from multiple different types of cancers induces a metastable early embryonic stem cell program and suppresses MHC-I expression in a subpopulation of cells.

3.2 Results

3.2.1 Two classes of DUX4-positive cancer cell lines: DUX4-like and DUX4-full-length

We determined read counts for *DUX4* and a panel of 39 genes regulated by DUX4 in the cancer cell lines entered in the Cancer Cell Line Encyclopedia (CCLE)[53]. Comparing *DUX4* mRNA levels to a DUX4-regulated transcriptional signature revealed two different categories of DUX4-expressing cancer cell lines. One category, termed DUX4-like, had relatively high levels of reads mapping to *DUX4* but did not have reads mapping to the set of genes regulated by DUX4. The second category, DUX4-full-length (DUX4-FL), had a low but detectable number of reads mapping to *DUX4* together with reads mapping to multiple DUX4-target genes (Fig. 3.2.1).

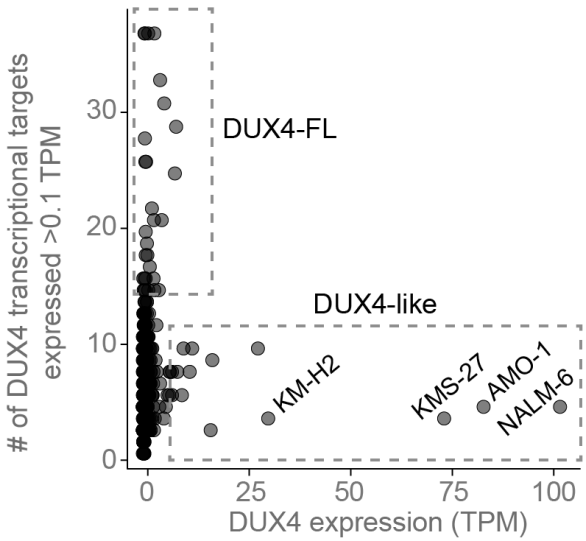


Figure 3.2.1: Two classes of DUX4 expressing cancer cell lines. RNA-seq of 935 CCLE cancer cell lines plotted by *DUX4* expression (TPM) versus number of DUX4 targets expressed (threshold = 0.1TPM).

3.2.2 DUX4-like cancer cell lines do not encode a full-length DUX4 protein.

To determine the origin and structure of the *DUX4*-like RNAs, we focused on the four cell lines with the highest expression: NALM6 (adult B cell acute lymphoblastic lymphoma), AMO1 (plasma cell myeloma), KMS27 (plasma cell myeloma), and KMH2 (Hodgkin lymphoma). Prior studies showed that NALM6 cells have a translocation that inserts DUX4-coding sequences from the D4Z4 repeat into the IgH locus and produce a *DUX4-IgH* fusion transcript that lacks the last 16 amino acids of the DUX4 protein, a region necessary for its transcriptional activity[21,54].

For each of these four cell lines, reverse transcription (RT)-qPCR confirmed both expression of an mRNA amplified by primers to *DUX4* as well as the absence of mRNAs for a target gene of DUX4 (Fig. 3.2.2A-B). 3-prime rapid amplification on cDNA ends (3'RACE) and sequencing of RT-PCR products confirmed the previously described *DUX4-IgH* fusion transcript in the NALM6 cells[55]; whereas the three other cell lines expressed novel *DUX4*-like transcripts

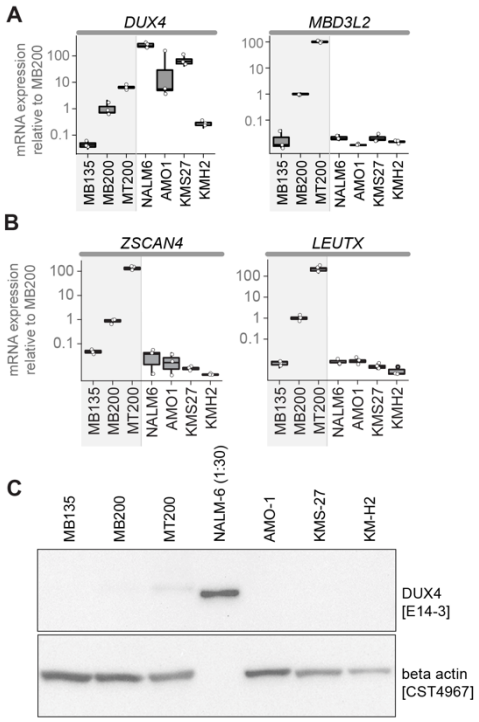


Figure 3.2.2: DUX4-like cancer cell lines do not express full length DUX4

(A-B) RT-qPCR of DUX4-like lines normalized to *RPL27*. Control human myoblasts (MB135) that do not express *DUX4*, FSHD2 myoblasts (MB200) that express low levels of DUX4, and FSHD2 myotubes (MT200) that express higher levels of DUX4 were included for comparison. **(C)** Western blot analysis of with DUX4-like cells lines for DUX4 protein expression with β-actin as control (NALM6 lysate is diluted 1:30).

that mapped to D4Z4-like arrays in the pericentromeric region of chromosome 22 and unmapped genomic regions (Table 1). Based on the sequenced region of these three *DUX4*-like mRNAs and their corresponding genomic sequences, none encoded the full-length DUX4 protein because of frame-shifts and/or stop codons between the homeodomain region and the carboxyterminal domain of DUX4 (Table 1). Western blotting with antibodies to the amino-terminus of DUX4 revealed robust expression in NALM6 but not in the other lines (Fig. 3.2.2C). Together with the absence of DUX4-target gene expression, these data indicate that the DUX4-like group do not express a full-length DUX4 protein that is transcriptionally active.

type	cell line	clone	assay	Primers	alignment	last 60bp	spliced?	3' feature	3' end bp#
DUX4-like	KMS-27	KMS-27-5	3' RACE	exon 1 (+1045 to 3'end)	Homo sapiens DNA, sequence_id: unplaced_0007	AATTCCGGCTGAGGCTCTTCCAACACCTTCCGACGCTCTTTAGTAAAACCCTGCAGAACA	NA	NA	NA
		KMS-27-4-5				TAACTCTTGTGTTGGATCAGCCTATGAGGGCATTGTGACATATATCTGCACTGATCACCC	NA	NA	NA
		KMS-27-4-4			Homo sapiens DNA, sequence_id: unplaced_0076	CTCTGCCTATGGGGGCATTGTGACTTATCTCTACAGTGATCAGCAAGGTGATGAAACCCC	NA	NA	NA
		KMS-27-9-1		exon 1 (+406 to +877)	Homo sapiens isolate CHM13 chromosome 22	ACCCGAGACAGGCTGGCAGGGCGCCCGCCTAGGCAGGCTGCCTGTGCAATGCATTCCCCG	NA	NA	NA
		KMS-27-9-2		exon 1 (+85 to +1230)	Homo sapiens DNA, sequence_id: unplaced_0034	ACCCGGGACAGGCTGACAGGGCGCCTGGGCAAGCAGGCGGCCTGTGCAACATGGCCGCAG	NA	NA	NA
		KMS-27-10-2	PCR	exon 1 (+406 to +670)		AACCCAGGTAAGGCTGGCAGGGAGACCACGCAGTCAGGCGGCCGGTGCAACACAGCCCCC	NA	NA	NA
		KMS-27-11-1		exon 1 (+406 to +1018)	Homo sapiens DNA, sequence_id: unplaced_0011	ACCCAGGTAAGGCTGGCAGGGAGACCACGCANNCNGGCGGCCGGTGCAACACAGCCCCCA	NA	NA	NA
		KMS-27-11-2		exon 1 (+406 to +1230)		AACCCAGGTAAGGCTGGCAGGGAGACCACGCAGTCAGGCGGCCGGTGCAACACAGCCCCC	NA	NA	NA
	AMO-1	AMO-1-25-4	3' RACE	exon 1 (+1045 to 3'end)	Homo sapiens isolate CHM13 chromosome 22	TGGGCTCACCGGCCTGGGATTCCTGCCTTCTAGGTCTAGTCCCGATAAAAGACTCCACAC	NA	NA	NA
		AMO-1-25-5				CGGCCTGGGATTCCTGCCTTCTAGGTCTAGTCCCGATAAAAGACTCCACACAGCTGAAAA	NA	NA	NA
		AMO-1-26-3				GCTCACCGGCCTGGGATTCCTGCCTTCTAGGTCTAGTCCCGATAAAAGACTCCACACAGC	NA	NA	NA
		AMO-1-26-5				GCTCACCGGCCTGGGATTCCTGCCTTCTAGGTCTAGTCCCGATAAAAGACTCCACACAGC	NA	NA	NA
		AMO-1-33-1		exon 1 (+85 to +406)	Homo sapiens isolate CHM13 chromosome 22	GAGCGTGCTTTGAGCGGAACCTGTAACCGGGCATCGCCACCAGAGAACGTCTGTCCCAGA	NA	NA	NA
		AMO-1-34-1		exon 1 (+85 to +1230)	Homo sapiens DNA, sequence_id: unplaced_0011	GAGCGTGCTTTGAGCGGAACCCGTACCCGGGCATCGCCACCAGAGAATGGCTGGCCCAGC	NA	NA	NA
		AMO-1-34-2		exon 1 (+85 to +1230)	Homo sapiens DNA, sequence_id: unplaced_0039	GAGCGTGCTTTGAGCGGAACCCGTACCCGGGCATCACCACCAGAGAACGGCTGGCCCAGG	NA	NA	NA
		AMO-1-34-3		exon 1 (+85 to +1230)		TGCTTTGAGCGGAACCCGTACCCGGGCATCACCNCCAGANAACNGCTGGCCCAGGCCATC	NA	NA	NA
		AMO-1-35-1		exon 1 (+406 to +1230)	Homo sapiens isolate CHM13 chromosome 22	CACCCTGGACAGGCTGGCAGGGCACCCGCGCAGTCAGGCGCCCTGTGCAATGTGGCCCCC	NA	NA	NA
		AMO-1-35-2		exon 1 (+406 to +1230)	Homo sapiens DNA, sequence_id: unplaced_0029	CACGCGGGACAGGCTGGCAGTTTGCCCACGCAGGCAGGCGGCTGGAGCAACGCGGCCCCC	NA	NA	NA
		AMO-1-35-3	PCR	exon 1 (+406 to +1230)	Homo sapiens DNA, sequence_id: unplaced_0007	CACCCGGGACAGGGTGGCAGGGCGCCCGCGCAGGCAGGCGGCCTGTGCAGCGCGGCCCCC	NA	NA	NA
	KM-H2	KMH2-2-1		exon 1 (+406 to +670)		CCACCGAGGAGCCTGAGGGTGAGAGAGCGCCCATTCCGGAGGAGCCGGTGCGGCGTAGGC	NA	NA	NA
		KMH2-3-1		exon 1 (+406 to +1018)	Homo sapiens DUX4 like 14 (pseudogene) (DUX4L	GCGTCCGGGGGCGCGGGCTGGGGAGGTGGAGCTGCCCCGGCTCGGGGTTCCCATGCTGCC	NA	NA	NA
		KMH2-5-1		exon 1 (+406 to +1230)	Homo sapiens isolate CHM13 chromosome 22	CTAGGCACCCGGGACAAGCTGGCACGGCGCAAGCGCAGGCAGGCGGCCTGTGCATCGCGG	NA	NA	NA
		KMH2-5-2		exon 1 (+406 to +1230)		CCAGGCACCCGGGACAGGCTGGCAGGGAGCCCACGCAGGCAGGCGGTCTGTGCAACGTGG	NA	NA	NA
		KMH2-6-1		exon 1 (+406 to +877)		CCAGGCCCATGCGCAGTGGGACAGCCTGGGCCCGCTCAATCGGGGCCACAGGGCCAAGGT	NA	NA	NA
		KMH2-6-2			Homo sapiens DNA, sequence_id: unplaced_0011	AGGGGCTGCGTTGCAAAAACCGCCTACGTGCCCAGGTGCCCTGCCACCCTGTCCTGGGTG	NA	NA	NA
	NALM6	NALM6-3-1	3' RACE	exon 1 (+1018 to 3'end)	DUX4; IgH (matches published fusion transcript)	CTGGAGGCCTCGGAAGAGGCCGCCTCGCTGCGATCCCTCAGGGTTGGCTGTAGTCACCGT	NA	NA	NA

Table 1: 3'RACE targeting DUX4 in DUX4-like cell lines

3.2.3 DUX4-FL cancer cell lines express full length DUX4 and DUX4-target genes in a subset of cells

We ranked the cancer cell lines in the DUX4-FL group based on the number of reads mapping to the panel of DUX4 target genes (Fig. 3.2.3A) and focused on characterizing five of the top cell lines: KLE, HCC38, SKNMC, G401, and SUSA (representing endometrial, mammary epithelial, neural epithelial, rhabdoid, and testicular teratoma cancer cell lines, respectively). As negative controls, we used two cell lines in the CCLE that were negative for DUX4 and its targets, Rh30 and MCF7, as well as the commonly used HeLa cell line that was not included in the CCLE.

Oligo-dT primed RT-qPCR demonstrated DUX4-FL lines expressed DUX4 mRNA at levels comparable to FSHD myoblasts and each expressed subsets of DUX4-target genes at similarly low levels (Fig. 3.2.3B). Furthermore, treating DUX4-FL cell lines with anti-DUX4 MOE

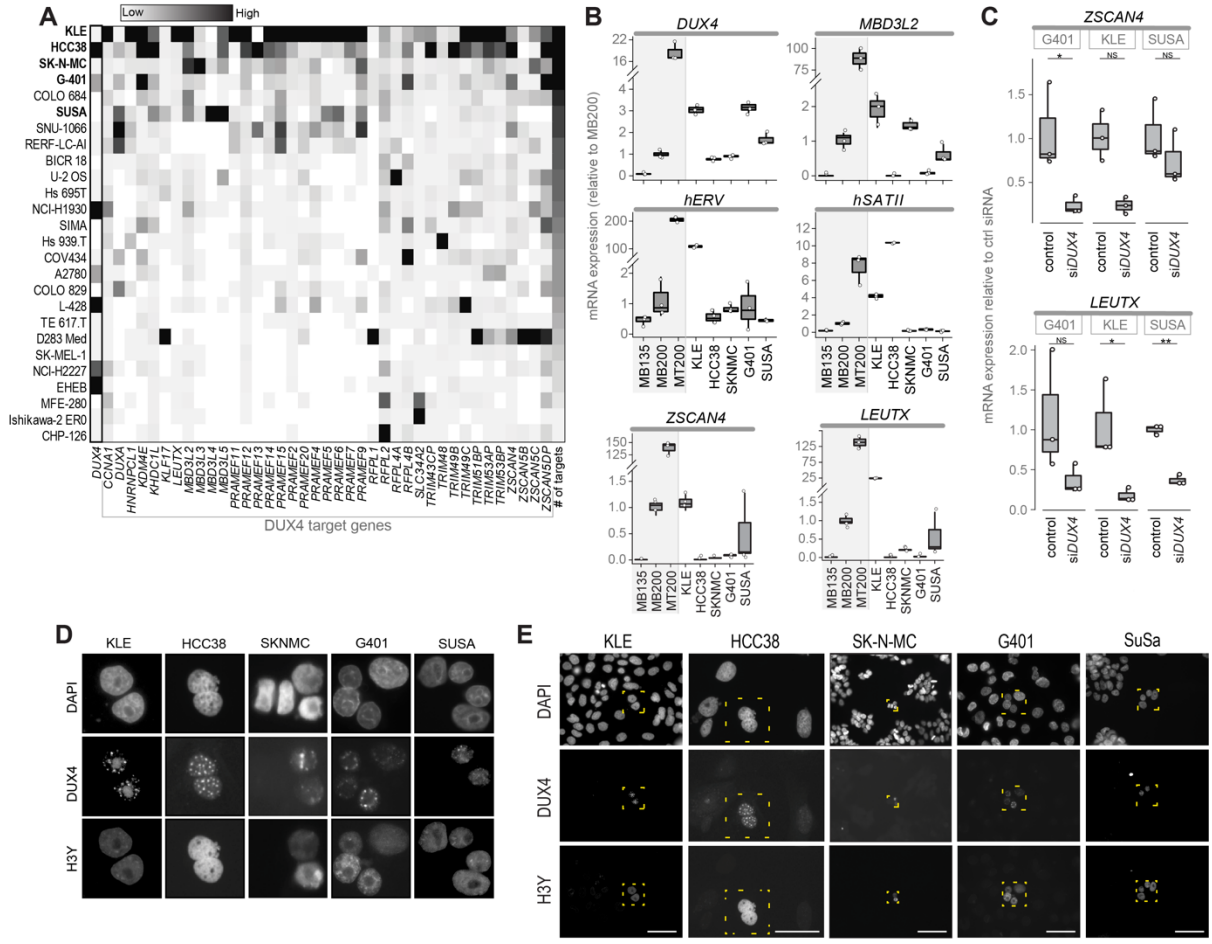


Figure 3.2.3: DUX4-FL cancer cell lines contain rare DUX4 expressing cells

(A) RNA-seq of 935 CCLE cancer cell lines plotted by *DUX4* expression (TPM) versus number of DUX4 targets expressed (threshold = 0.1TPM). **(B)** RT-qPCR of DUX4-like lines normalized to *RPL27*. Control human myoblasts (MB135) that do not express *DUX4*, FSHD2 myoblasts (MB200) that express low levels of DUX4, and FSHD2 myotubes (MT200) that express higher levels of DUX4 were included for comparison. **(C)** RT-qPCR on RNA from DUX4-FL lines normalized to RPL27 and MB200 myoblasts primed with oligo-dT for *DUX4* and *MBD3L2*, or random hexamers for *hERV* and *hSATII*. **(D)** Immunofluorescence for H3Y and DUX4 in DUX4-FL lines (cropped from Fig E). **(E)** Full IF images (cropped in E) . P-values: * < 0.05; **<0.005.

gapmers reduced the expression of the DUX4 targets *ZSCAN4* and *LEUTX* (Fig. 3.2.3C), confirming that DUX4 is driving expression of these genes. Consistent with the low levels of DUX4 and DUX4-target mRNA expression, immunofluorescence identified DUX4 and DUX4-target proteins in a small fraction of the cells (~0.1%), often in pairs or small clusters (Fig. 3.2.3; cropped from 3.2.3E); consistent with a short burst of DUX4 expression, similar to the brief expression of DUX4 or mouse Dux in the early embryo[18] and in ES/iPS cells[6,19]. These results indicate that the low levels of DUX4 and DUX4-target mRNAs in the log-phase cultures might represent relatively high expression in a small population of cells, similar to the case for DUX4 expression in FSHD muscle cells[20].

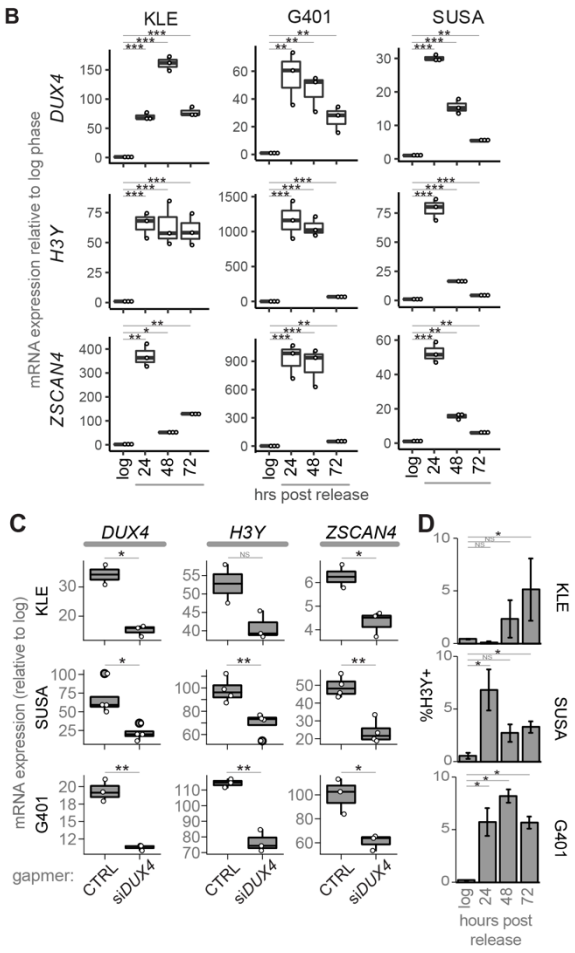
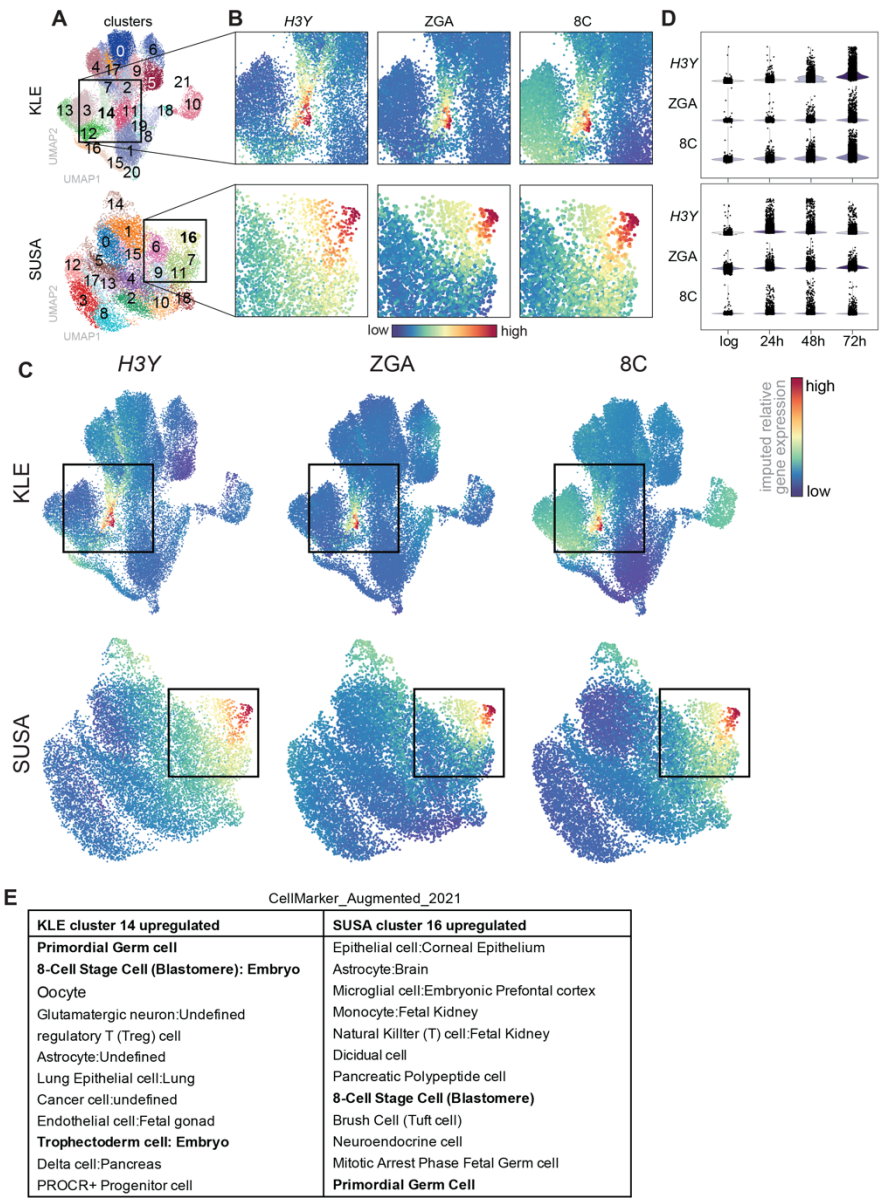


Figure 3.2.4: release from confluence increases transient DUX4 expression.
(B) RT-qPCR for DUX4 and target genes at denoted time points after release from confluence. **(C)** RT-qPCR for DUX4 and DUX4 target genes 48hrs after release from confluence treated with siDUX4 or siCTRL. **(D)** Percentage of *H3Y*-positive nuclei by IF at 24, 48, and 72 hrs post-release form a confluent state; expressed as mean +-SEM. For all RT-qPCR experiments, n=3 biological replicates, except n =2 for KLE gamper control. For IF quantification, n = 3 biological replicates per timepoint (each summed from 5 separate frames of view). P-values: * < 0.05; **<0.005, **<0.0005.

3.2.4 Release from confluence increases transient DUX4 expression

To facilitate the study of the biological consequences of DUX4 expression in these cancer cell lines, we tried multiple culture conditions and determined that there was a substantial and transient increase in the expression of *DUX4* and DUX4-target RNAs following release from confluence (Fig. 3.2.4B). Furthermore, *DUX4* knockdown confirmed that the DUX4-target gene expression was driven by DUX4 (Fig 2.2.4C). Immunodetection of the DUX4-target H3Y showed that the increased mRNA levels correlated with a higher percentage of DUX4-target expressing cells, reaching ~5% of the population, and that the increase in H3Y cells persisted for several days following release from confluence (Fig. 3.2.4D).

3.2.5 Transient DUX4 expression activates DUX4-targets, ZGA genes and early embryonic lineage genes



E

CellMarker_Augmented_2021

KLE cluster 14 upregulated	SUSA cluster 16 upregulated
Primordial Germ cell	Epithelial cell:Corneal Epithelium
8-Cell Stage Cell (Blastomere): Embryo	Astrocyte:Brain
Oocyte	Microglial cell:Embryonic Prefontal cortex
Glutamatergic neuron:Undefined	Monocyte:Fetal Kidney
regulatory T (Treg) cell	Natural Killter (T) cell:Fetal Kidney
Astrocyte:Undefined	Dicidual cell
Lung Epithelial cell:Lung	Pancreatic Polypeptide cell
Cancer cell:undefined	**8-Cell Stage Cell (Blastomere)**
Endothelial cell:Fetal gonad	Brush Cell (Tuft cell)
Trophectoderm cell: Embryo	Neuroendocrine cell
Delta cell:Pancreas	Mitotic Arrest Phase Fetal Germ cell
PROCR+ Progenitor cell	**Primordial Germ Cell**

Figure 3.2.5: DUX4 activates the ZGA/8CLC programs

(A-F) Single-cell RNA sequencing of KLE and SUSA lines in log-phase and 24, 48 and 72hr after release from confluence**. (A)** Phenograph Leiden cell clusters. **(B)** Heat map showing relative expression of *H3Y*, as well as ZGA and 8C gene composite scores cropped from **(C)** full UMAPs. **(D)** Stacked violin plots showing relative expression of *H3Y*, ZGA gene composite, and 8C gene composite in log-phase cells and the indicated time post-release from confluence. **(E)** Most significant EnrichR Cell Marker expression patterns for KLE and SUSA.

To determine whether transient DUX4 expression in the cancer cell lines induced the broader transcriptional program of the cleavage-stage embryo, we performed single-cell RNA sequencing on SUSA and KLE cells in log-phase and at 24, 48, and 72hrs following release from confluence using the 10X Genomics sequencing platform. Data analysis included use of the ScanPy pipeline[56] and regression for cell cycle genes (see Methods), resulting in UMAPs with 18 clusters in the SUSA cells and 21 clusters for KLE cells (Fig. 3.2.5A).

H3Y mRNA was enriched in KLE cluster 14 and SUSA cluster 16 (Fig. 3.2.5B,C), whereas DUX4 expression was not reliably detected with single-cell sequencing, possibly due to its low abundance and

extremely high GC content. EnrichR indicated that these clusters expressed gene sets similar to the 8-cell embryo and primordial germ cells (Fig. 3.2.5E). Consistent with this, composite scores based on 23 DUX4-regulated ZGA genes[6] identified high expressing cells in the same cluster as *H3Y*-expressing cells (Fig 3.2.5B-C). Expression analysis for each ZGA gene showed that all of the 23 genes in this ZGA program were elevated in this cluster (KLE Cluster 14 and SUSA Cluster 16) and, to a lesser extent, in adjacent clusters (KLE Clusters 17 and 12, SUSA Clusters 7 and 11) (Fig. 3.2.7A). In addition to the DUX4-regulated portion of the ZGA program, genes that characterize 8C-like cells (8CLCs) that are not known to be directly regulated by DUX4 (*DPPA3*, *TPRX1*, and *KLF17*) were expressed in cells in these clusters (Fig. 3.2.7A,B), and a composite score based on 717 8CLC genes[6] identified the expression of the broader 8CLC program in these clusters (Fig. 3.2.5C). Because the genes identified as comprising the 8CLC program include the DUX4-regulated ZGA genes and other genes meeting criteria for direct targets of DUX4, we performed a similar analysis of 8CLC genes comparing expression of a basket of 8C genes directly regulated by DUX4 (8C Direct, 64 genes) to the remainder of the 8C gene set (8C Not Direct, 653 genes) (see Methods for assigning gene categories), which showed similar although slightly expanded distribution of expression of putative indirect targets of DUX4, consistent with expression in cells transiting out of the DUX4-expressing population (Fig. 3.2.7C).

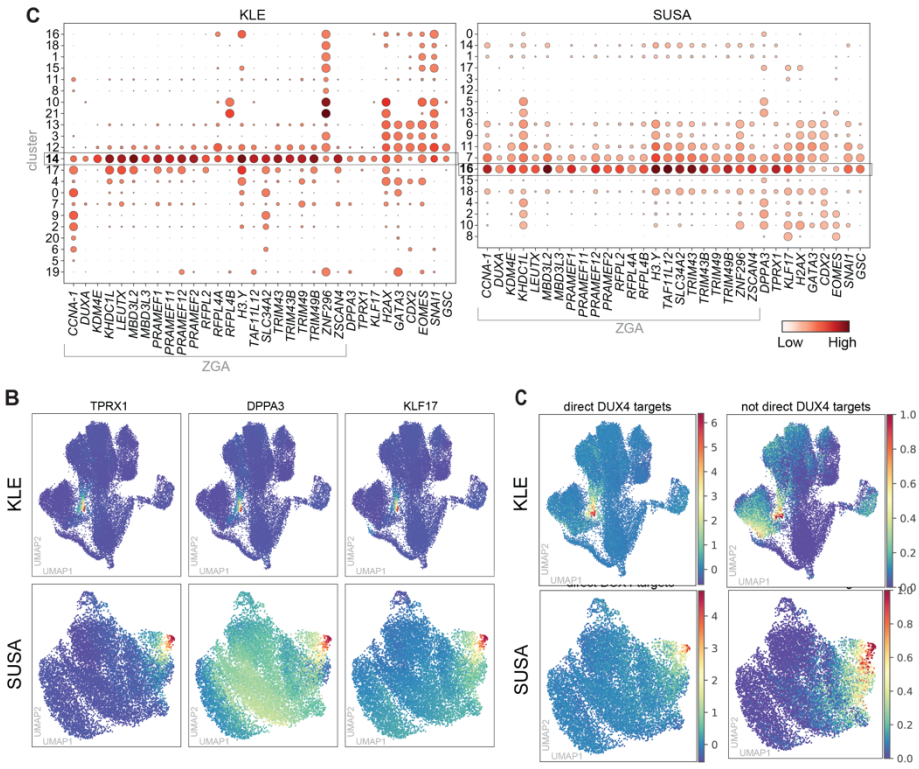


Figure 3.2.7: DUX4 activates the ZGA/8CLC programs.

(A) Dot plot showing the fraction of cells (size) and mean expression (color) of each indicated gene. **(B)** Single-cell RNA-seq as UMAPs demonstrating relative expression of indicated 8C genes. **(C)** Single cell RNA-sequencing UMAPs demonstrating relative expression of 8CLC gene identified as direct targets of DUX4 or as not direct targets of DUX4.

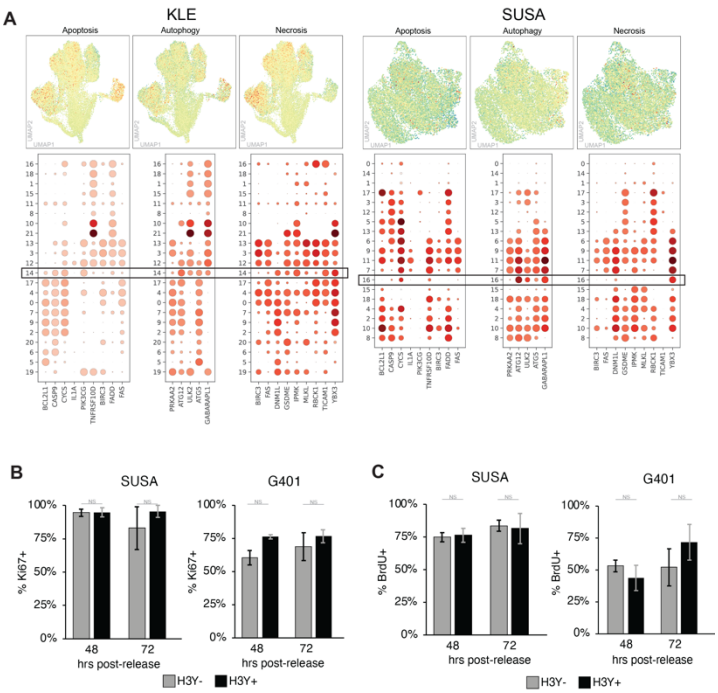


Figure 3.2.8: DUX4 expression does not affect cell death or proliferation.

(A) Single-cell RNA-seq as UMAPs and dot plots of relative expression of composite scores for the indicated cell death pathways. **(B-C)** Immunofluorescent quantification of **(B)** Ki67+ or **(C)** BrdU+ cells after release from confluence, n= 3seperate experiments; >240 cells per timepoint.

Assessing the expression of these genes at the different time points (log-phase and 24, 48, 72 hrs post-confluence) showed that cells expressing the ZGA and 8CLC genes were enriched after release from confluence compared to the log-phase cells (Fig. 3.2.5D), correlating with the RT-qPCR enrichment of the *DUX4* mRNA and DUX4-target genes in these populations. The persistence of the ZGA and 8C populations over 72 hrs indicated that the transient expression of DUX4 in these cancer cells did not result in cell death, and this was supported by the absence of enrichment for gene expression associated with the apoptosis, autophagy, or necrosis pathways in the ZGA/8C enriched cluster (Fig 3.2.8A). In addition, immunostaining for the proliferation marker Ki-67 and BrdU labeling for six hours showed that the H3Y-positive cells had the same high proliferative rate as the H3Y-negative population at both 48 and 72 hrs following release (Fig. 3.2.8B-C). These findings are consistent with the appearance of small clusters of H3Y expressing cells at these time points.

These data indicate that transient DUX4 expression was associated with activation of the DUX4-regulated ZGA transcriptional program and the subsequent broader 8CLC program in these two cancer cell lines. In addition to this early totipotent program, EnrichR also indicated genes associated with trophectoderm and early lineage specification were enriched in the *DUX4*/*H3Y* cluster in KLE cells (Fig. 3.2.5E). This was supported by expression of genes regulating trophectoderm formation (*GATA3*, *CDX2*, *EOMES* (both KLE and SUSA) and *TFAP2C* (KLE) or *TFAP2A* (SUSA)) and genes involved in mesenchymal specification and tumor metastasis (*SNAI1*, *GSC* in both and *FOXC2* in SUSA) in the DUX4-target and adjacent

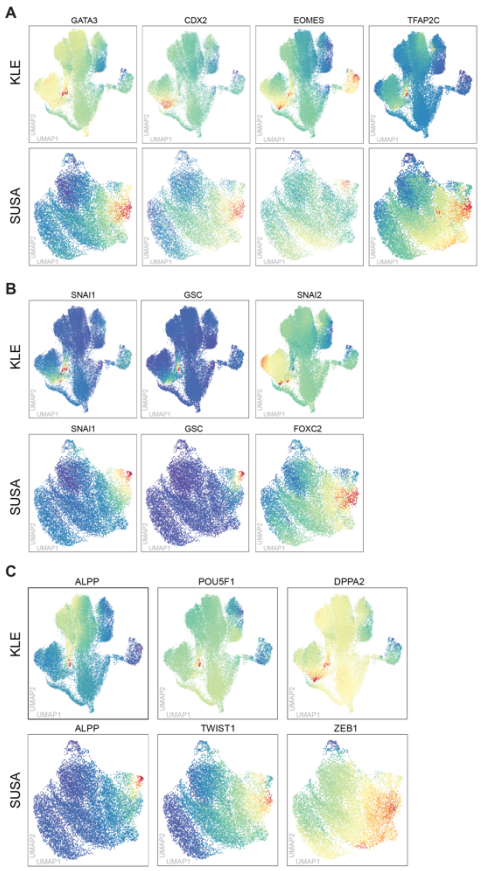


Figure 3.2.9: DUX4 expression Drives early embryonic lineage genes.

(A) Single-cell RNA-seq as UMAPs and dot plots of relative expression of composite scores for the indicated cell death pathways. **(B-C)** Immunofluorescent quantification of **(B)** Ki67+ or **(C)** BrdU+ cells after release from confluence, n= 3seperate experiments; >240 cells per timepoint.

(Fig. 3.2.10E). Similar results were obtained in the cells expressing the ZGA signature (Fig. 3.2.10L). Therefore, the most profound suppression of MHC-I genes in cancers expressing DUX4 is associated with the DUX4-expressing subpopulation and the clusters (Fig. 3.2.7C and 3.2.9 A-B). The expression of the placental alkaline phosphatase (*ALPP*), as well as *TWIST1* and *ZEB1* in SUSA or *POU5F1* and *DPPA2* in KLE further support the expression of trophectoderm, mesenchymal, and early embryonic markers that occur following ZGA (Fig. 3.2.9C), and is consistent with a report that knockout of mouse Dux results in diminished expression of genes involved in trophectoderm development[38]. Together, these results indicate at least a partial induction of early embryonic lineage programs that might have relevance to cancer progression.

Our prior study showed that DUX4-expressing cancers showed reduced MHC-I Class I gene expression and resistance to checkpoint blockade inhibitors[11]. In the current single cell sequencing data, the cells expressing the 8CLC genes showed lower expression of the canonical MHC-I genes (HLA-A, B, and C) and elevated expression of some of the non-canonical MHC genes (HLA-E and F)

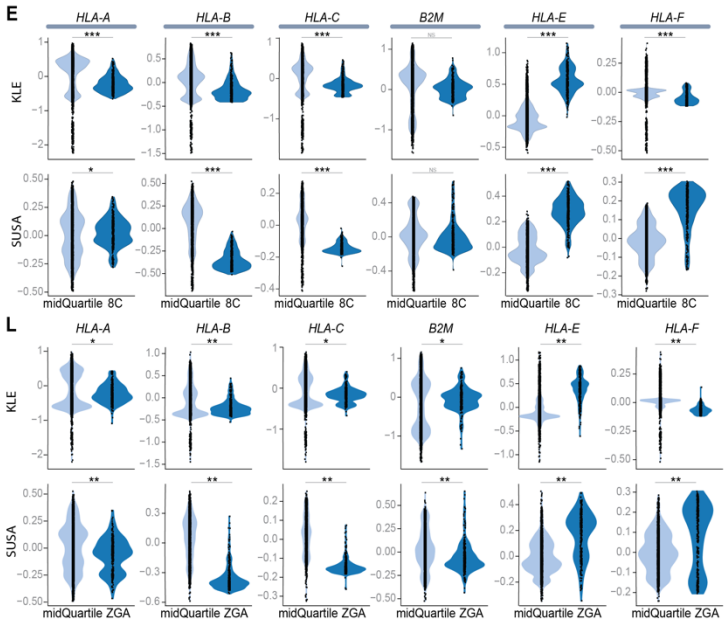


Figure 3.2.10: DUX4 suppresses MHC class I expression.

(E) Violin plots comparing HLA gene expression in the mid quartile or top 350 cells based on imputed 8C gene composite scores. **(L)** Violin plots comparing HLA gene expression in the mid quartile or top 350 cells based on imputed ZGA gene composite scores. P-values: * < 0.05; **<0.005, ***<0.0005, NS = no significance

subsequent ZGA and 8CLC states.

3.2.6 DUX4 expression is sufficient to induce early embryonic lineage genes in cancer cell lines.

RT-qPCR and single-cell sequencing showed a correlation between *DUX4*/*H3Y* expression and the expression of 8CLC and early embryonic lineage genes in the cell populations expressing H3Y and other DUX4-regulated ZGA genes, and the expression of some of these genes was also evident in other clusters (see Fig. 3.2.7C). To determine whether the transient expression of DUX4 is sufficient to induce genes characteristic of these embryonic and extraembryonic programs, the cancer cell lines KLE, SUSA, and G401 were transduced with a doxycycline inducible *DUX4* vector (i*DUX4*^{CA})[57]. Treatment with doxycycline continuously for 24hrs in KLE cells or for a four-hour pulse in SUSA cells and G401 cells resulted in the transient expression of DUX4 and the induction of *ZSCAN4* (Fig. 3.2.11A-C) as well as the induction of early embryonic lineage genes (*GATA3*, *KLF4*, *EOMES*, *CDX2*, *SNA1*, *GSC*) that persisted for several days (Fig. 3.2.11D-F).

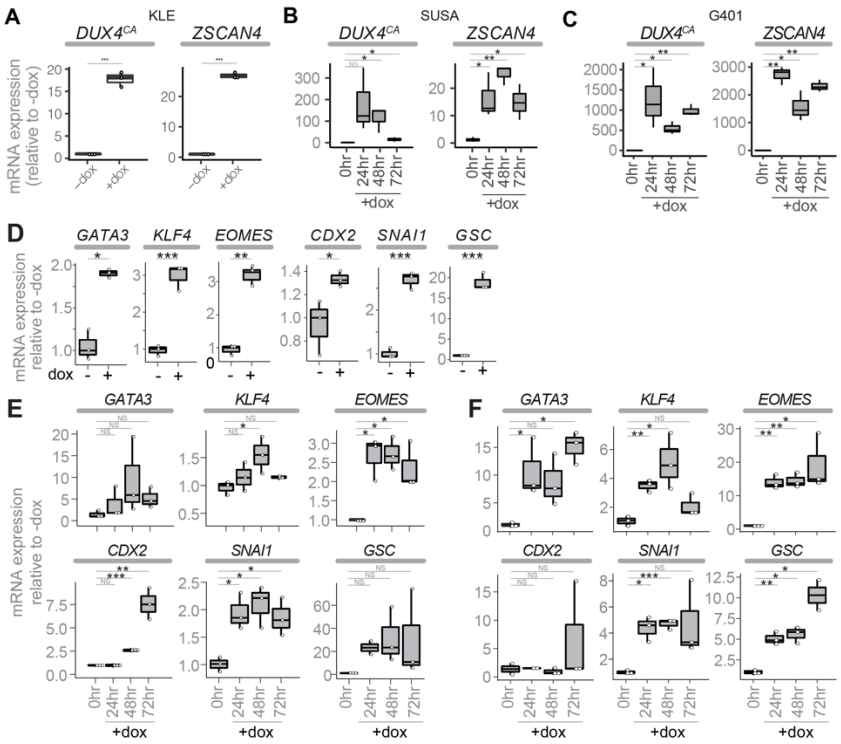


Figure 3.2.11: DUX4 expression is sufficient to induce early embryonic lineage genes in cancer cell lines.

(A-C) RT-qPCR for DUX4 and ZSCAN4 on RNA from: (A) KLE-i*DUX4*^{CA} after 20 hrs of continuous doxycycline (dox); **(B)** SUSA-i*DUX4*^{CA} after following a 4-hour exposure to doxycycline (dox) and harvested 24, 48, and 72 hrs later; **(C)** similar to (B) with RNA from G401-i*DUX4*^{CA}. **(D-F)** RT-qPCR for the indicated trophectoderm and mesenchymal specification genes, after treatment with doxorubicin (dox), in: **(D)** KLE-i*DUX4*^{CA}, **(E)** SUSA-i*DUX4*^{CA}, and **(F)** G401-i*DUX4*^{CA}.

3.2.7 DNA damage enhances DUX4 expression in cancer cells with wild type p53

A recent study[58] of the regulation of DUX4 in early development identified p53 as an inducer of DUX4 expression. Consistent with those findings, treatment of the two cancer cell lines with wild-type p53, SUSA and G401, with etoposide or doxorubicin, two topoisomerase II inhibitors that induce p53 secondary to DNA double-strand breaks, showed strong induction of

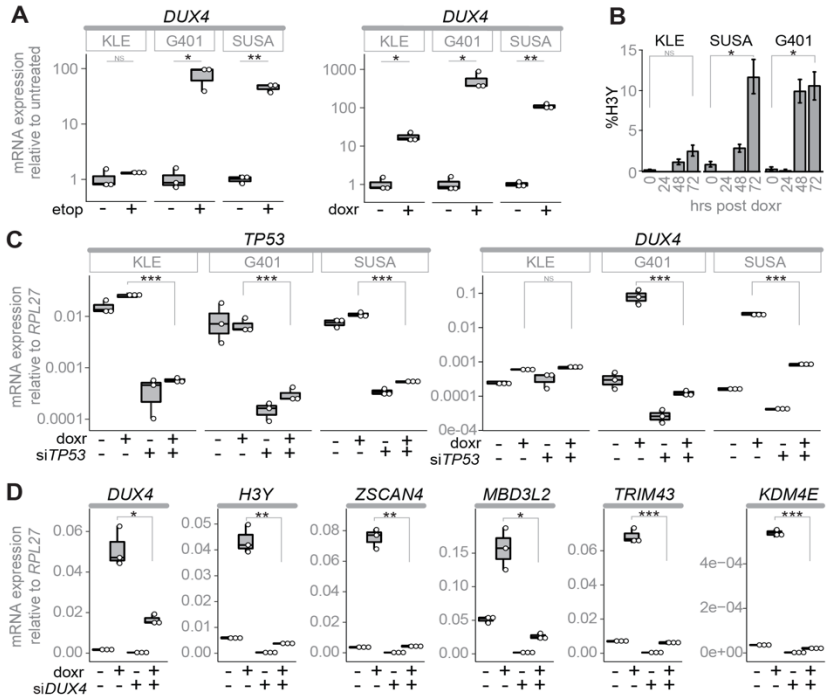


Figure 3.2.12: DUX4 is induced by DNA damage.

(A) RT-qPCR for *DUX4* on 3 DUX4-FL cell lines 48hr post treatment with etoposide (etop) or doxorubicin (doxr) graphed as the fold-increase relative to untreated. **(B)** Percentage of *H3Y*-positive nuclei by IF at the indicated time point after a 6hr treatment with 1uM doxorubicin; expressed as mean +-SEM. **(C)** RT-qPCR on KLE, SUSA, and G401 RNA 24hrs post treatment with doxorubicin in cells pretreated with si*TP53* or siControl. **(D)** RT-qPCR on G401 RNA 24hrs post-treatment with doxorubicin in cells pretreated with anti-*DUX4* MOE gapmer (+) or control (*GFP*) gapmer (-). For all experiments, n=3 biological replicates (individual biological samples displayed as white dots). P-values: * < 0.05; **<0.005, ***<0.0005, NS = no significance.

DUX4 (Fig. 3.2.12A); and this correlated with a substantial increase in the percentage of H3Y positive cells (Fig. 3.2.12B). Knockdown of *TP53* with siRNA substantially reduced *DUX4* induction in these cells following DNA damage (Fig. 3.2.12C). In contrast, the KLE cell line has a mutant *TP53* allele, p.R175H[59], a dominant-negative mutation in the DNA-binding domain that prevents p53 target induction[60-62], and shows more modest induction of DUX4 following DNA damage that is not changed with knockdown of *TP53* (Fig. 3.2.12C). (The DepMap portal reports G401 as having a p.C277F mutation in *TP53*, however the source publication[59] for this reports WT *TP53* in this line and our sequencing did not detect this C277F mutation in G401). Finally, knocking down *DUX4* with MOE gapmers during treatment with doxorubicin largely prevented the induction of ZGA genes (*H3Y*, *ZSCAN4*, *MBD3L2*, *TRIM43*, *KDM4E*) following doxorubicin treatment (Fig. 3.2.12D).

3.2.8 Cancer cell populations expressing DUX4 suppress MHC Class I presentation

The single cell sequencing indicated that the expression of DUX4 suppressed steady-state levels of MHC-I mRNAs (see Fig. 3.2.10). To determine whether DUX4 induction in these cancer cells will suppress interferon-gamma (IFNg) stimulated MHC-I RNAs and cell surface

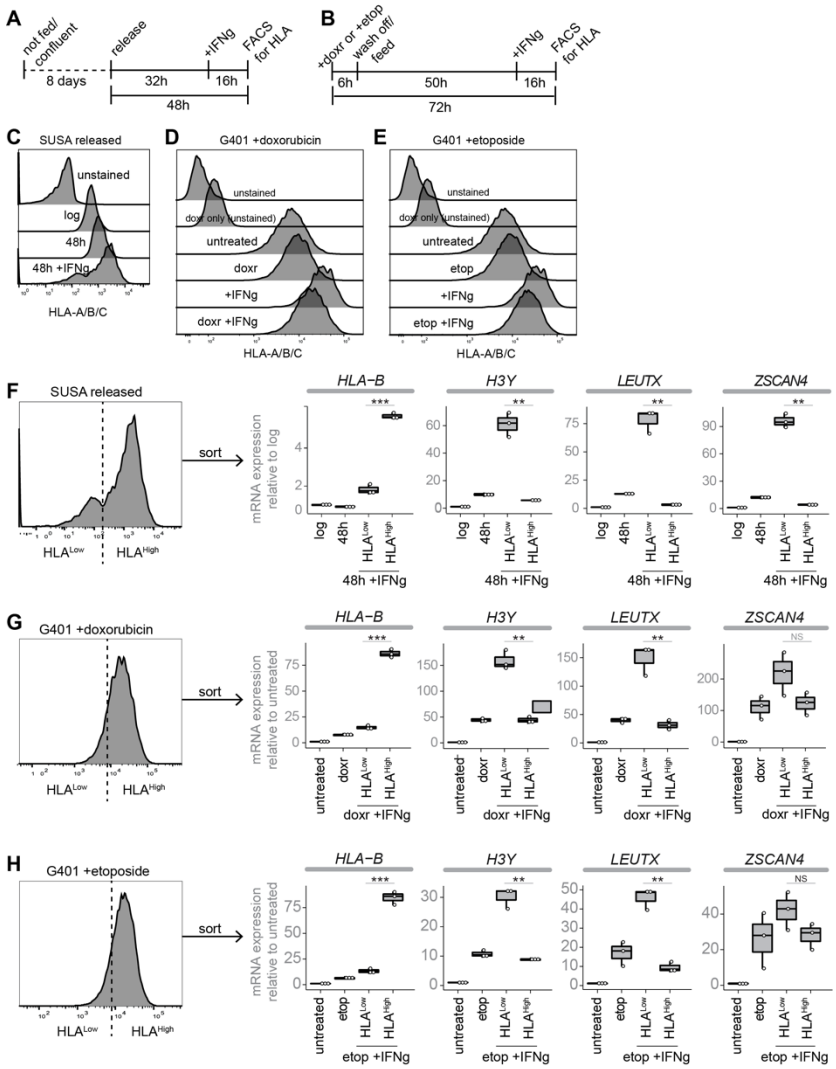


Figure 3.2.13: Low MHC-I surface protein correlates with DUX4 expression
(A-B) Timeline of cell treatments for **(A)** SUSA experiments and **(B)** G401 experiments. **(C-E)** Histograms or HLA-A/B/C expression measured with flow cytometry. **(C)** SUSA 48h post-release from confluence were stained with BV605 anti HLA-A/B/C. **(D-E)** G401 72h post-treatment with **(D)** 1uM doxorubicin (doxr) or **(E)** 10uM etoposide (etop). (D and E) were analyzed at the same time and use the same control samples. **(F-H)** FACS analysis for HLA^{High} vs HLA^{Low} cells post-IFNg treatment followed by RTqPCR for *HLA-B* and 3 DUX4-target genes on unsorted and sorted populations of: **(F)** released SUSA, **(G)** G401 +doxorubicin, or **(H)** G401 +etoposide. For all RT-qPCR experiments, n=3 technical triplicates. P-values: * < 0.05; **<0.005, ***<0.0005, NS = no significance.

presentation, we treated SUSA cells released from confluence and G401 cells pre-treated with doxorubicin or etoposide with 50ng/mL IFNG for 16 hrs (Fig. 3.2.13A and B) and used flow cytometry to measure relative HLA-A/B/C surface expression (Fig. S3A-C). We sorted the treated populations into HLA^{High} versus HLA^{Low} (Fig. 3.2.13C-E) and analyzed mRNA expression of DUX4-regulated genes that also characterize the ZGA and 8CL states. RT-qPCR demonstrated that low HLA-A/B/C surface expression correlated with high expression of these DUX4 regulated genes (Fig. 3.2.13F-H). These data further support that DUX4 expression in cancer cells is associated with suppressed MHC-I presentation, and demonstrate that both metabolic parameters and DNA-damaging compounds might modulate immune evasion in DUX4-expressing cancers.

3.2.9 Rare DUX4+ cells detected in PDX model

To screen solid tumors for DUX4+ cells, we compiled a panel of antibodies suitable for immunohistochemistry on formalin-fixed paraffin embedded (FFPE) sections. The DUX4

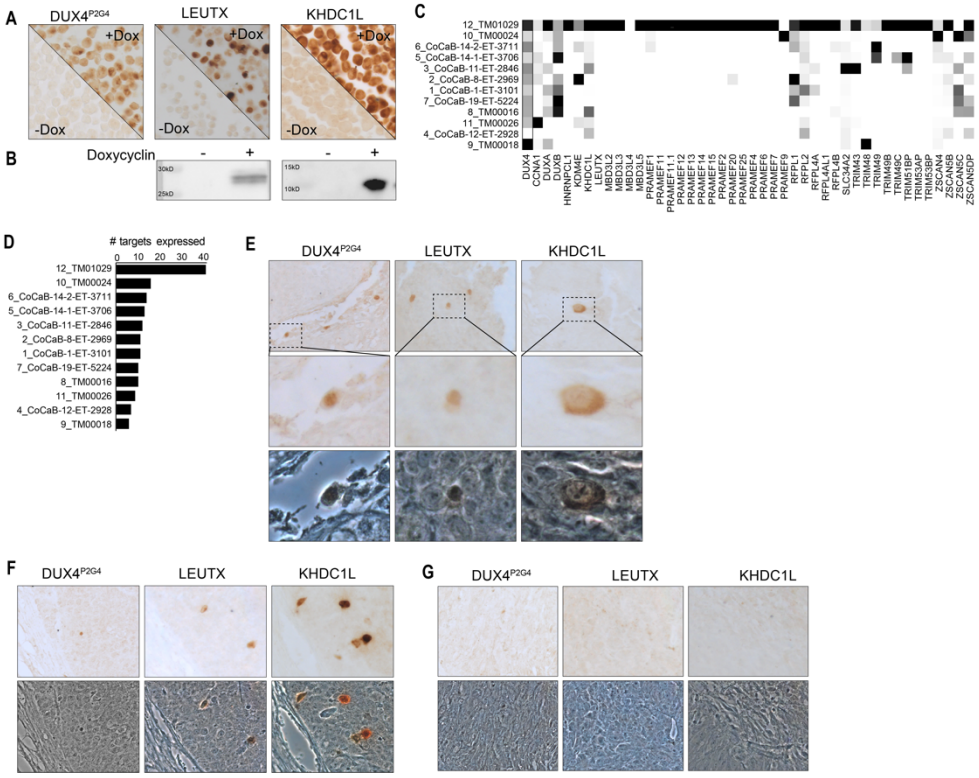


Figure 3.2.14: DUX4+ cells identified in PDX model. **(A)** Immunohistochemistry on PPFE section s of MB135-i*DUX4*^{CA} cell pellets with antibodies indicated. (B) western plots on MB135-i*DUX4*^{CA} protein extract. **(C-D)** bulk RNA-sequenceing of 12 different blader PDX models. **(C)** Heatmap demonstrating relative expression of *DUX4* and DUX4 transcriptional target genes. **(D)** Number of DUX4 target genes expressed over 0.1 TPM. **(E-G)** Immunohistochemistry on PPFE sections of bladder cancer PDX **(E-F)** TM01029. or **(G)** TM00018 for the molecular markers indicated.

antibody P2G4 has been validated for staining PPFE sections, whereas the H3Y antibody does not work on PPFE sections. Therefore, we generated antibodies to DUX4 targets LEUTX and KHDC1L. The specificity of these antibodies was confirmed in PPFE sections of MB135-iDUX4CA +/- doxycycline as well as in western blots (Fig 3.2.14A-B).

To determine if these DUX4+ cells exist in solid tumors, we took advantage of patient derived xenograft (PDX) models with paired RNA-sequencing data. Comparing expression of DUX4 target genes in 12 bladder cancer PDX models demonstrated that PDX TM01029 was a likely candidate to contain DUX4+ cells (Fig 3.2.14C-D). Immunohistochemistry on TM01029 sections detected rare cells staining positive for DUX4 or its transcriptional targets LEUTX and KHDC1L (Fig. 3.2.14E). Staining adjacent sections for these molecular markers demonstrated cells positive for each of these 3 markers in the same region, which could be different planes of the same cells (Fig 3.2.14F), whereas in stained sections of DUX4 target-negative PDX model TM00018 we were unable to detect any positive cells (Figure 3.2.14G). Taken together, these demonstrate the high likelihood that DUX4+ cells exist naturally in solid tumors.

3.3 Discussion

Our prior study showed that the RNAs encoding DUX4 and DUX4-target genes were detected in a wide variety of cancers based on data in The Cancer Genome Atlas (TCGA) and other sequence repositories[11]. However, it remained unknown whether this indicated the presence of a distinct subset of cells in the cancer population that re-activated a broader early embryonic program. In the current chapter, we identified several cancer cell lines that express DUX4 and DUX4-target genes, albeit at relatively low levels. We show that the relatively low levels represent the transient expression of DUX4 in a subset of the cancer cells with subsequent activation of a metastable early embryonic transcriptional program. Single-cell RNA-sequencing showed that these cells activated the early DUX4-regulated ZGA program, as well as the broader 8C-like signature. In addition, the associated expression of genes that characterize early trophectoderm, and epithelial/mesenchymal lineages indicated the progression to components of the early lineage specification pathways, and the suppression of steady-state and IFNg-induced MHC-I expression suggests a contribution to immune evasion. Together, our data suggest that DUX4 expression in cancers can transiently induce a metastable early embryonic program that includes genes characteristic of trophectoderm development and epithelial-to-mesenchymal transition that might contribute to aspects of cancer progression.

The transcriptional program induced by DUX4 was initially described in muscle cells in the context of FSHD where it was shown to regulate an early stem cell program that included induction of *ZSCAN4* and classes of retroposons[21,63]. Subsequent studies showed that transient expression of DUX4 in the 4-cell human embryo, or Dux in the 2-cell mouse embryo, activated the early wave of ZGA gene expression[13,16,18]. Although Dux knockout mice can produce viable embryos, several studies have shown some degree of reduced or delayed ZGA gene expression[36,37] and reduction of genes involved in trophectoderm development[38]. Earlier studies had shown that mouse ES cells transiently cycle through a 2C-like naïve state that includes ZGA gene expression and metabolic and epigenetic reprogramming[19,47]. More recent studies implicate the transient expression of Dux/DUX4 as a regulator of the transient 2C or 8C-like state in mouse or human ES cells, respectively[6,48], and that transient expression of DUX4 in human ES cells can induce the ZGA program and an 8C-like state[64]. Our current study builds on and extends this prior work by using the transcriptional programs identified as characterizing the ZGA and 8CLC human ES cell states[6] to show that the cells in DUX4 expressing cancers initiate the DUX4-regulated portion of the ZGA program and components of the broader 8CLC gene expression program.

It is interesting to consider why some cancer cell lines transiently express DUX4 in a subset of cells. In ES/iPS cells and the early embryo, Dppa2 and Dppa4 have been shown to bind the Dux locus and regulate its initial expression[65] in addition to activation by p53[58], quickly followed by epigenetic suppression by SMCHD1[66,67] resulting in methylation of the locus possibly by antagonizing TET proteins[68]. In FSHD, mutations in chromatin modifiers (most commonly SMCHD1 or DNMT3B) or D4Z4 array sizes of 10 or less result in hypomethylation of the D4Z4 region, inefficient epigenetic repression, and episodic expression of DUX4 in FSHD muscle cells[20,49] indicating that D4Z4 hypomethylation and decreased epigenetic repression is sufficient for episodic DUX4 expression in some somatic cells. In addition, multiple chromatin modifying factors (e.g., components of the NuRD, CAF, bromodomain, and polycomb complexes) (reviewed in [69,70]) and signaling pathways[71,72] can modulate the efficiency of DUX4 suppression. Indeed, our previous study of DUX4 expressing cancers identified loss-of-function mutations in twelve of 23 genes encoding validated or likely repressors of DUX4[11], including SMCHD1 and DNMT3B, as well as high DPPA2 and DPPA4 expression in testicular germ cell tumors. In this chapter study, the relative hypomethylation at the D4Z4 locus in the DUX4-expressing cancer cell lines provides an initial explanation for why these lines are especially permissive for DUX4 expression. We did not detect higher expression of other macrosatellite repeats, suggesting that the de-repression of the D4Z4 does not reflect a genome-wide loss of repeat repression.

The SUSA cell line was isolated from a testicular teratocarcinoma[73] and expresses high levels of DPPA2 and DPPA4[74], factors implicated in the developmental regulation of Dux/DUX4 expression[65]. As a cancer cell line derived from a teratocarcinoma, it might retain an ability to differentiate into early embryonic lineages and our current study implicates DUX4 expression as regulating the expression of some of these early embryonic genes. KLE cells have mutations in multiple chromatin regulators (DNMT1, BAZ3B, MSX1, CBX, PRDM9, ARID1B, and DPF1) which might individually or in concert alter D4Z4 epigenetic repression; and G401 have an inframe deletion in SMARCA5 (E816del), a gene previously shown to regulate DUX4 expression[75]; however, the consequences of this particular mutation remain unknown. Together, these data support a multifactorial model of regulating D4Z4 repression in different cancer cell lines, similar to our prior analysis of DUX4-expressing cancers[11], with the common feature of D4Z4 hypomethylation, a feature causative for episodic DUX4 expression in FSHD muscle cell.

In this model, D4Z4 hypomethylation permits induction of DUX4 expression by factors that would be epigenetically blocked in normal somatic tissues, such as p53 that shows greater penetrance of DUX4 activation in cells with hypomethylated D4Z4. It is likely that the

hypomethylated locus will also be more sensitive to regulation by other factors and signaling pathways. For example, increased DUX4 expression following release from confluence occurs independent of p53 status and might reflect activation of stress pathways, such a p38, or as a consequence of other signaling pathways or the chromatin relaxation associated with S-phase. Future studies are needed to determine whether D4Z4 methylation status is a reliable predictor of DUX4 expression in a broad set of cancers.

The biological significance of DUX4 and ZGA gene expression in cancers might easily be underestimated because of the very low levels of expression in the analysis of bulk tumor RNA. However, our study shows that the low levels in bulk RNA analysis represent sub-populations of cells that broadly activate early developmental programs. Moreover, our demonstration that growth conditions and DNA damage can substantially increase the percentage of DUX4 expressing cells indicates that environmental factors dynamically modulate DUX4 expression and might provide some advantage during specific bottlenecks in cancer progression.

In this regard, it is interesting to speculate on the possible roles of DUX4 in cancer progression. Together with our prior study showing DUX4-mediated suppression of MHC presentation, our current demonstration that DUX4 expressing cancer cells induce components of the 8CLC program and subsequent genes that regulate the specification of embryonic/extra-embryonic tissues and epithelial/mesenchymal specification suggests that DUX4 expression might contribute to many of the major hallmarks of cancer[76,77]. Additional studies will be necessary to determine the specific roles of DUX4 expression in cancer progression, metastasis, immune evasion, and response to therapies.

3.4 Materials and Methods

Haplotyping and Methylation Analysis

Number units in D4Z4 macrosatellites on chromosomes 4 and 10 as well as their haplotype was determined with Pulsed Field Gel electrophoresis, Southern blotting and subsequent hybridization of the blots with probes P13E-11, A and B probes. SSLP length for further haplotyping was determined using an ABI Prism 3100 Genetic analyzer. CpG methylation of D4Z4 macrosatellites on chromosomes 4 and 10 was determined by sensitivity to *Fse*I endonuclease. All as described previously[78].

Cell Culture

MB135 and MB200 myoblast lines were cultured in F-10 media supplemented with 10% fetal bovine serum and 1x penicillin/streptomycin. For differentiation, myoblasts were cultured in DMEM supplemented with 1x penicillin/streptomycin, insulin (10ug/mL), and transferrin (10ug/mL).

Cancer cell lines were cultured in varying media formulation supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. G401 cells were cultured in McCoy's 5A (Modified) Medium; KLE and SCC-9 cells were cultured in DMEM/F-12; SKNMC and MCF7 cells were cultured in EMEM; SUSA, HCC38, NALM6, AMO1, KMS27, KMH2, MOLP-2, and OPM-2 cells were cultured in RPMI. HeLa cells were cultured in DMEM.

Release from confluence

Cells were plated at confluence and maintained without media change for 1-2 weeks. Cells were then exposed to fresh media for 2hrs and then passaged 1:6 for post release time course.

Gapmer and siRNA treatments

MOE Gapmers or siRNAs were mixed with OptiMEM and used at a final concentration of 25nM with 3.75uL/mL RNAiMax Lipofectamine (ThermoFisher, 13778075) overnight at 37C in normal media.

Drug treatments

Cells were treated for 6hrs with 1uM doxorubicin, 10uM etoposide, or 5uL DMSO (vehicle) for 6hrs. RNA was harvested 48hrs post treatment using the Machery-Nagel RNA extraction kit.

Single-cell RNA Sequencing (scRNA Seq)

Cells released from confluence were cryobanked at 24, 48, and 72hrs after release in BamBanker (Fisher, NC9582225) at -80°C. Cells were thawed, washed with cold media +10%FBS, then cold PBS with 1% BSA. Cells were filtered through a 35uM filter and resuspended to 1,000cells/ul and kept on ice until processed according to 10X genomics protocol and sequenced by the FHCC genomics core.

Single-cell RNA Data Analysis

Sequencing data were aligned and quantified using 10X Genomic's Cell Ranger v 6.1.2[79] using the reference human GRCh38 genome and GRCh38-2020-A transcriptome supplied with Cell Ranger. Raw feature matrices were initially analyzed independently using ScanPy v. 1.9.1[56], where genes not expressed in at least 1 cell and cell barcodes without at least 200 unique genes expressed were removed. Cell barcodes were then filtered with the following thresholds: the top 2.5% of cells by UMI counts were removed as potential doublets; cells with log10(UMI) less than 2.8 were removed as low quality; and cells with more than 25% UMIs aligning to mitochondrial reads were removed, also considered poor quality. After filtering, ScanPy was used to normalize and log transform read counts. The top 2500 highly variable genes were calculated and used to perform dimensional reduction via principal component analysis (PCA) with subsequent visualization by PC variance plot to determine optimal number of principal components (PCs) to include in downstream analyses. Next, we calculated the neighborhood graph and performed non-linear dimensional reduction using UMAP and the first 30 PCs. Finally, clusters were calculated using scanpy.tl.leiden() and cluster. Sources of technical bias were plotted on UMAPs to ensure that clustering was not being driven by artifact. After individual samples were processed, they were concatenated using anndata's concatenate function and analyzed in general method as above, including per-cell normalization, log transformation, PCA, neighborhood and UMAP dimensionality reduction using the first 30 PCs. Deviations from the above workflow for the merged dataset include calculating a S and G2M cell cycle gene module score using scanpy.tl.score_genes_cell_cycle, a simple linear regression using scanpy.pp.regress_out() to remove technical noise associated with UMI and mitochondrial counts as well as gene expression associated with S and G2M gene module score, gene expression imputation using scanpy.external.pp.magic(), and clustering using Phenograph ver. 1.5.7[80]. Sources of technical bias were again plotted on the UMAP to ensure that the clustering was driven by biological processes. The count matrix was exported and Seurat (ver. 4.1.1)[81,82] was used to determine cluster marker genes via FindAllMarkers(). Marker genes in EnrichR[83-85] were used to query the CellMarker_Augmented_2021 database[86]. Imputed counts were used for all single gene visualizations, and z-score normalized imputed counts were used for gene module scores calculated via scanpy.tl.score_genes(). Dot plot clustering dendrogram was calculated using ZGA genes alone. Cells were ranked by 8C-like score or ZGA score and MAGIC imputed expression for MHC and B2M genes for the cells in the top 350 and the middle quartile (37.5%-62.5%) of ranked expression. These expression levels were plotted as violin plots using ggplot2[87] and statistical significance was assessed using permutationTest2() from the resample[88] package with 500,000 permutations of the difference between the mean of the top 350 and middle quartile groups.

Western Blot

Cells were harvested in RIPA buffer containing protease and phosphatase inhibitors and sonicated on low for 5 minutes. The whole cell lysate was centrifuged at 12,0000 RCF for 10 minutes and the supernatant transferred to a new tube. Protein quantification was performed using the Pierce BCA Protein Assay Kit. Either 5ug or 10ug of whole cell lysate was heated in LDS Sample Buffer with 2.5% β-mercaptoethanol at 70°C for 10 minutes with shaking. For SDS-PAGE, proteins were loaded onto a 4-12% NuPAGE Bis-Tris gel and run at 100V for about 2 hours in NuPAGE MOPS SDS Running Buffer with 300uL NuPAGE antioxidant. Proteins were transferred to a 0.2µm PVDF membrane at 30V for 1 hour in NuPAGE Transfer Buffer with 10% methanol. Membranes were blocked in 5% milk in TBST for 1 hour, then incubated at 4°C overnight with rocking. Blots were washed three times (15 minutes each) with TBST and incubated with HRP-conjugated secondary antibody for 1 hour. Blots were washed three times (15 minutes each) with TBST, and bands were detected via chemiluminescence with either SuperSignal West Pico PLUS Substrate, SuperSignal West Femto Substrate, or SuperSignal West Atto Substrate.

Immunofluorescence and Immunohistochemistry

Cells were washed with PBS, fixed in 2% paraformaldehyde for 10 minutes, and washed twice with PBS. Cells were then permeabilized in 0.1% Triton X-100 in PBS for 10 minutes at room temperature. Cells were then incubated overnight at 4°C with DUX4 [E14-3] primary antibody (1:400 dilution) or H3Y1/2 [8H6-2111] (1:333 dilution) and washed with PBS three times for 10 minutes each. Cells were incubated in TRITC-conjugated donkey anti-rabbit secondary antibody (1:666 dilution) and TRITC-conjugated donkey anti-rat secondary antibody (1:666 dilution) for 3 hours at room temperature. Cells were washed with PBS, stained with DAPI for 10 minutes at room temperature, and washed with PBS. Cells were imaged with a fluorescence microscope.

Reverse transcription-quantitative PCR (RT-qPCR)

RNA was harvested with the NucleoSpin RNA Kit (Takara) according to the manufacturer's protocol. RNA quality was verified by NanoDrop 2000 (Thermo Scientific). RNA was treated with DNaseI, Amplification Grade (Invitrogen). Reverse transcription was performed in a 20uL reaction: 200-1000ng whole RNA, 1uL dNTP (10mM), 1uL oligo dT primer (10mM), 4uL 5x SSIV Buffer, 1uL DTT (100mM), 1 uL RNaseOUT, and 1uL SSIV RT enzyme. Thermal cycling conditions for reverse transcription were as follows: 50°C for 40 minutes, 55°C for 30 minutes, and 80°C for 10 minutes. Complementary DNA (cDNA) was treated with 1uL of RNaseH and incubated at 37°C for 20 minutes, then diluted 1:5 or 1:4 with RNase-free H2O. Quantitative real-time PCR (qPCR) was performed on the QuantStudio™ 7 Flex Real-Time PCR System in a 10uL reaction: 2uL cDNA, 5uL 2x iTaq Universal SYBR Green Supermix, 0.3uL forward and reverse primer (10uM), and 2.4uL H2O. qPCR primers were synthesized by Integrated DNA Technologies (IDT) and are listed in Star Methods table. Thermal cycling conditions for qPCR were as follows: 50°C for 2 minutes and 95°C for 10 minutes; 40 cycles of 95°C for 15 seconds and 60°C for 60 seconds. For each Biological replicate, qPCR reactions were run in technical triplicates, including -RT controls. Median CT values of the technical triplicates were used for analysis. Gene expression was normalized to the housekeeping gene RPL27 (ribosomal protein L27). P-values were calculated with an unpaired, one-tailed student's t-test assuming unequal variance (type 3).

Generation of cell lines with dox-inducible codon altered DUX4 transgene

A polyclonal KLE-i*DUX4*^{CA} cell line was generated with lentiviral transduction, selected, and maintained in 3µg/mL puromycin as previously described for MB135-i*DUX4*^{CA} cells in Jagannathan et al., 2016. SUSA-i*DUX4*^{CA} and G401-i*DUX4*^{CA} cell lines were generated using CRISPR/Cas9-mediated genomic integration at the AAVS1 safe harbor locus via homology directed repair. The codon-altered *DUX4* transgene was subcloned by restriction enzyme digest into the AfeI and SalI sites of the pMK364 vector backbone (AAVS1-Tet-OsTIR1-AAVS1-Puro; Addgene #158663) containing a dox-inducible promoter and homology arms to the AAVS1 safe harbor locus. SUSA and G401 cells were co-transfected with Cas9/AAVS1-sgRNA expressing plasmid (Addgene #62988) and i*DUX4*^{CA} plasmid described above at a 1:2 ratio using Lipofectamine 3000 (Fisher #L300015) according to the manufacturer's protocol. Cells were incubated for two days post-transfection, and subsequently selected and maintained in 3µg/mL puromycin. Monoclonal cell lines were isolated, and i*DUX4*^{CA} transgene integration at the AAVS1 safe harbor locus was validated with PCR amplification from genomic DNA followed by sanger sequencing.

3' rapid amplification of cDNA ends (3' RACE)

RNA was harvested with the NucleoSpin RNA Kit (Takara) according to the manufacturer's protocol. RNA quality was verified by NanoDrop 2000 (Thermo Scientific). RNA was treated with DNaseI, Amplification Grade (Invitrogen). To enrich for DUX4 transcripts, biotinylated oligonucleotide probes targeting DUX4 exon 1 were denatured for 5 minutes at 95°C, then hybridized to RNA in hybridization buffer (0.05M Tris-HCl pH7.5, 0.75M NaCl, 1mM EDTA, 10mM DTT, RNaseOUT, and 15% formamide). Thermal cycling conditions for hybridation were 75°C for 2 minutes, followed by 15 cycles for 10 minutes each ramping down at -1C per cycle (ranging from 65°C - 50°C). 50uL streptavidin (MyOneC1) beads were washed twice with wash buffer (0.05M Tris-HCl pH7.5, 0.75M NaCl, 1mM EDTA, 10mM DTT, RNaseOUT, and 0.05% Tween20). Beads were resuspended in 25uL binding buffer (0.01M Tris-HCl pH7.5, 2M NaCl, 1mM EDTA, 10mM DTT, RNaseOUT, and 0.05% Tween20), and 100uL hybridized RNA sample was added to beads and incubated overnight at room temperature with rotation. Beads were washed twice with 500uL wash buffer. RNA was eluted by resuspending beads in 100uL dissociation buffer (10mM EDTA, 10mM DTT, RNaseOUT, and 95% formamide) and incubating at 95°C for 2 minutes with shaking. RNA was ethanol precipitated overnight at -80°C, pelleted, air-dried, and resuspended in 20uL of RNase-free H_2O.

RACE reverse transcription was performed with RNA in a 20uL reaction: 7.5uL RNA, 2uL dNTP (10mM), 1uL long dT anchor primer (37.5uM), 4uL 5x SSIV Buffer, 0.5uL DTT (100mM), 0.5uL RNaseOUT, 1uL SSIV RT enzyme. Thermal cycling conditions for reverse transcription were as follows: 50°C for 40 minutes, 55°C for 30 minutes, and 80°C for 10 minutes. Complementary DNA (cDNA) was treated with 1uL of RNaseH and incubated at 37°C for 20 minutes, then purified with the Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit according to the manufacturer's protocol. A nested RACE-PCR was then performed in a 20uL reaction: 1uL template DNA, 1uL forward and reverse primer (10uM), 10uL 2x Phusion Plus PCR Master Mix, 4uL 5x Phusion GC Enhancer, and 3uL ddH_2O. Primary PCR was performed with the 1st set of nested primers the following thermal cycling conditions: initial denaturation at 98°C for 2 minutes; 25 cycles of 98°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds; and final extension at 72°C for 7 minutes. PCR products were purified and a 2nd PCR was performed with the 2nd set of nested primers with the following thermal cycling conditions: initial denaturation at 98°C for 2 minutes; 30 cycles of 98°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds; and final extension at 72°C for 7 minutes. PCR products were run on a 1% agarose TBE gel, and products of interest were gel excised and purified with the Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit according to the manufacturer's protocol. Products were A-tailed in a 20uL reaction: 10uL purified PCR product, 2uL 10X PCR Buffer (-$MgCl_2$), 0.25uL $MgCl_2$ (50mM), 0.25uL dATP (100mM), and 0.25uL Taq polymerase. Reaction was incubated at 72°C for 20 minutes. A-tailed PCR products were then TA TOPO cloned in a 4uL reaction: 2uL A-tailed PCR product, 0.5uL salt solution, and 0.5uL pCR™4-TOPO® TA vector. Reaction was incubated at room temperature for 30 minutes, then transformed in chemically competent DH5α *E. coli* cells. Colonies were grown up in 3mL LB-carbenicillin and Minipreps performed with the PureLink™ Quick Plasmid Miniprep Kit (Invitrogen) according to the manufacturer's protocol. Minipreps were test digested with EcoRI at 37°C for 1 hour and products visualized on a 1% agarose TBE gel. Minipreps were submitted for Sanger sequencing with the M13 Forward primer.

Staining for MHC-I and FACS

Cells were treated with 50ng/mL IFNg (R&D Systems, 285-IF/CF) for 16hrs before lifting with 0.25% Trypsin-EDTA. Cells were pelleted at 300RCF and resuspended in 100uL PBS 1%FBS + 1:50 APC/FIRE or BV605 mouse anti-human HLA-A/B/C (Biolegend) and incubated 20-30min at 4°C in dark. Cells were washed with PBS and filtered through 35uM strainer caps before analyzing and sorting with the BD Symphony 6.

Chapter 4. *DUX4* exon usage in development vs disease

This chapter contains unpublished work.

4.1 Introduction

To develop FSHD, one must have both a permissive D4Z4 haplotype and hypomethylation of the D4Z4. Although the D4Z4 encodes many DUX4 open reading frames (ORFs), it doesn't encode a known polyadenylation signal (PAS)[33]. What defines a permissive D4Z4 allele is having an accessible PAS downstream of the D4Z4, enabling transcripts from the terminal D4Z4 unit to be polyadenylated and DUX4 protein expression in muscle cells[33].

This single PAS is distal to the entire D4Z4 repeat and is transcribed in DUX4 exon 3. To access DUX4 exon 3, transcripts of the terminal D4Z4 unit splice exons 1-2-3. Exons 1 and 2 are encoded in each D4Z4 unit, but exon 3 is encoded in the pLam region – a highly variable non-coding genomic region that's named after the probe that binds to it. To develop FSHD the genome must have a pLam region which encodes this PAS-containing exon 3, these alleles are known as 4qA.

Not all D4Z4 alleles carry this distal exon 3 PAS, for example 4qB alleles completely lack this PAS and certain 4qA and 10qA alleles lack the exon 3 splice acceptors to access this PAS. However, embryos and male germ cells seem to express DUX4 regardless of the D4Z4 haplotypes they harbor. Furthermore, not all DUX4-FL cancer cell lines have a permissive allele. Therefore, there must be differences in DUX4 transcription across these contexts.

In this chapter, we test the ability of permissive D4Z4 alleles to access this distal PAS and define which DUX4-FL cancer cell lines have permissive alleles. We then compare DUX4 sequences and exon usage across different cell types and highlight the differences in DUX4 3'UTR exon usage.

4.2 Results

4.2.1 4A166 variants modulate DUX4 exon 2-3 splicing efficiency

While the D4Z4 haplotype 4qA161 is know to be a permissive allele (containing both the exon 3 PAS and efficient splice acceptors to access it) it is less well known if all 4qA166 alleles can also access this permissive PAS in exon 3. Comparing the 4qA161 sequence of intron 2 and exon 3 to that of 4qA166 revealed 6 SNPs (Fig. 4.2.1A). To evaluate whether these SNPs have the potential to affect exon 3 splicing efficiency we transfected wildtype human myoblasts (MB135) with a vector containing DUX4 exons 1, 2, 3, 6, and 7 together with all, or portions of, the intervening introns driven by the CMV enhancer/promoter. Comparing the relative splicing of exon 2-3, the dominant splice form in FSHD, versus the splicing of exon 2-6, the dominant splice form in the testis[20], allowed an estimation of the relative efficiency of DUX4 exon 2-3 splice site usage. Following transient transfection of the expression vector in MB135 myoblasts, RT-qPCR demonstrated the 4qA166 sequence showed significantly reduced usage of the exon 2-3 splice site relative to the exon 2-6 splice compared to the 4qA161 sequence (figure 4.2.1C). Additional constructs introducing the 4A166 G polymorphism near the splice acceptor site into 4A161 sequence (161-G) or introducing the two 4A161 variants adjacent to the exon 3 splice acceptor into the 4A166 sequence (166-A/G) showed that these were sufficient to alter exon 3 splice site efficiency (Fig 4.2.1C). Together, these transient transfection data demonstrate the potential for the sequence variants in the 4qA166 haplotype to decrease splicing efficiency between exon 2 and 3.

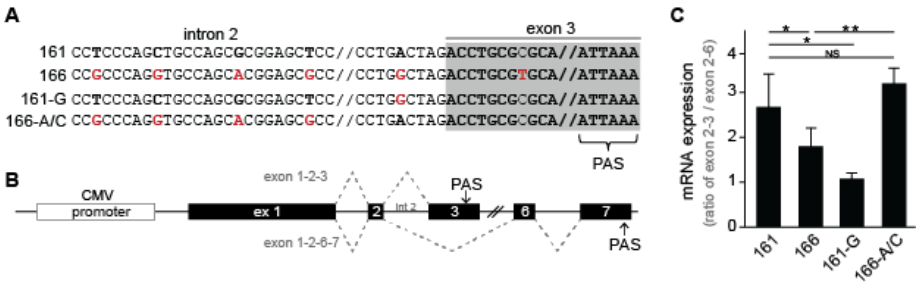


Figure 4.2.1: Single nucleotide polymorphisms affect splicing efficiencies of DUX4 transcripts.
(A) 4qA161 and 4qA166 sequence polymorphisms in DUX4 intron 2 and exon 3. **(B)** A schematic of pCS2-DUX4 expression cassette: the genomic region of DUX4 exon 1 through the first 27bp of intron 3 followed by the genomic region of the last 150bp of intron 6 (2037bp in full) through exon 7 all driven by one CMV promoter. **(C)** RT-qPCR on transfected MB135 for DUX4 exon 2-3 expression level relative to exon 2-6 levels. n = 3 biological samples for each construct. *: p< 0.05; **: p< 0.005; NS: no significant difference.

4.2.2 Cancer cell lines express DUX4 from multiple D4Z4 alleles with decreased DNA methylation

In contrast to DUX4 expression in FSHD muscle cells, which is almost exclusively from the FSHD-permissive 4qA161 allele and polyadenylated in exon 3, 3'RACE and RT-PCR in the

DUX4-FL cancer cell lines indicated *DUX4* mRNA expression from multiple 4qA, 4qB, and 10q D4Z4 locations that used alternative polyadenylation (pA) sites (Table 2). All of the cell lines expressed *DUX4* transcripts with chromosome 10-specific polymorphisms that spliced exon 2 with exon 6 and used a canonical pA signal in exon 6 (Table 2), consistent with *DUX4* isoforms previously identified in testis that arise from D4Z4 repeats of both chromosome 4 and 10[20].

Table 2: DUX4-FL mRNA sequences **Bold** letters = canonical PA signal

type	cell line	clone	assay	Primers	alignment	last 60bp	spliced?	3' feature	3' end bp#
DUX4-FL	G401	G401_2-1	3' RACE		4qA161				
		G401_2-2							
		G401_2-3			4qB163				
	HCC38	HCC38_2-1			4qA161	CTCCGGCCCCCGAAAGGCTGGCCATGCCGACTGTTTGCTCCCGGAGCTCTGCGGGCACCC	ex2-ex3a	exon 3a	+2654
		HCC38_2-5							
		HCC38_2-4			4qB163				
	SKNMC	SKNMC_2-1			4qA161				
		SKNMC_2-2							
		SKNMC_2-3							
	G401	G401_9-1			D4Z4 (no specifying SNPs)	AGCTTTAGGACGCGGGGTTGGGACGGGGTCGGGTGGTTCGGGGCAGGGCGGTGGCCTCTC	no	intron 1	+1483
		G401_9-2							
		G401_9-3							
	SUSA	SUSA_2-1							
		SUSA_2-2							
		SUSA_2-3							
	SKNMC	SKNMC_1-1		exon 1 (+1045 to 3'end)	D4Z4 (no specifying SNPs)	TCGCGGGGAACACCTGGCTGGCTACGGAGGGGCGTGTCTCCGCCCCGCCCCCTCCACCGG	no	intron 1	+1545
		SKNMC_1-3							
	HCC38	HCC38_1-1			D4Z4 (no specifying SNPs)	AGGGGCGTGTCTCCGCCCCGCCCCCTCCACCGGGCTGACCGGCCTGGGATTCCTGCCTTC	no	intron 1	+1572
		HCC38_1-2							
	KLE	KLE_2-1			4qA161	AGAGCCGGCCCAGGTACCAGCAGGTGGGCCGCCTACTGCGCACGCGCGGGTTTGCGGGCA	no	intron 2	+1704
	G401	G401_8-1			4qB163	CTCCACACCGCGGAGAACTGCCATTCTTTCCTGGGCATCCCGGGGATCCCAGAGCCGGCC	ex1-ex2	exon 2	+1654
		G401_8-2							
		G401_8-3							
	SKNMC	SKNMC_3-6			4qB163	CGGTGAGAGACTCCACACCGCGGAGAACTGCCATTCTTTCCTGGGCATCCCGGGGATCCC	ex2-ex3b-ex2	exon 2	
	G401	G401_3-2			10qA166	TTTTGTTTTAATNTCAGTGACAGGGGAACACAGGTAGCTNT**AATATA**GCTGTGANTGAGA	ex2-ex6	intron 6	+11498
		G401_7-1							
	KLE	KLE_6-1							
	HCC38	HCC38_3-1							
	SUSA	SUSA_1-1							
	SKNMC	SKNMC_3-5							
	HCC38	HCC38_4-1			4qB163	CGGTGAGAGACTCCACACCGCGGAGAACTGCCATTCTTTCNTGGGCATCCCGGGGATCCC	ex3b-ex2	exon 2	+1488
	SKNMC	SKNMC_7-1		exon 3B (+1416 to 3'end)	D4Z4 (no specifying SNPs)	GCAACGGCGACGCCCACAGGCATTGCCTCCTTCACGGAGAGAGGGCCTGGCACACTCAAG	no	exon 3b	+2714
	SUSA	SUSA_3-1				AGCCAAAGCGAGGCCCTGCGAGCCTGCTTTGAGCGGAACCCGTACCCGGGCATCGCCACC	no	exon1 (HD1)	+3384
		SUSA_4-1		exon 3B (+2249 to 3'end)	D4Z4 (no specifying SNPs)	GGGCCAGGCACCCGGGACAGGGTGGCAGGGCGCCCGCGCAGGCAGGCGGCCTGTGCAGCG	no	exon1 (HD2)	+3736
	G401	G401_11-1				TTTGAGCGGAACCCGTACCCGGGCATCGCCACCAGAGAACGGCTGGCCCAGGCCATCGGC	no	exon1 (HD1)	+3411

Table2: DUX4 sequences detected in DUX4-FL cancer cell lines with 3'RACE

In addition to this previously described germline *DUX4* isoform, 3'RACE identified several DUX4 RNA isoforms with pA sites shared by multiple cancer cell lines, albeit lacking a canonical pA signal motif. Three of the cell lines (G401, HCC38, and SKNMC) had transcripts originating from 4qA and 4qB haplotypes with an exon 2/3a or 2/3b splice junction and polyadenylation at a site in exon 3a. Many of the cancer cell lines (G401, SUSA, SKNMC, and HCC38) also had *DUX4* transcripts polyadenylated at identical or nearby sites in intron 1, again without a nearby canonical pA motif. The appearance of the same or similarly located pA sites from multiple cells lines suggests that these might represent alternative pA sites, however, it remains very possible that the high GC content in this region creates amplification artifacts and the true pA site might be distal to those mapped in this study, or the RNAs might not be polyadenylated despite priming with oligo-dT.

Decreased DNA methylation in the D4Z4 region is associated with DUX4 expression in FSHD[89] and it is interesting that the five DUX4-FL expressing cancer cell lines showed decreased DNA methylation of the D4Z4 region as determined by the percentage cleavage of a methylation sensitive enzyme (FseI), an assay that measures methylation on all haplotypes of both chromosome 4 and 10[33,78] (Fig. 4.2.2).

Taken together, these studies show *DUX4* mRNA transcripts are using both the exon 2-3a/b splice site and the exon 2-6 site arising from 4qA, 4qB and 10qA hypomethylated D4Z4 regions in the DUX4-FL cancer cell lines. Of note, the transcripts using the distal exon 6 are likely to arise from the last D4Z4 unit, whereas the transcripts using exon 3a or 3b splice acceptor site can arise from internal repeats and indicates that multiple D4Z4 units on different D4Z4 haplotypes contribute to *DUX4* mRNA in the cancer cell lines.

			Permisive	D4Z4 methylation	chr4 #1			chr4 #2			chr10 #1			chr10 #2		
cell line	M/F	clinical status	allele	(%FseI resistant)	units	A/B	SSLP	units	A/B	SSLP	units	A/B	SSLP	units	A/B	SSLP
SUSA	M	testicular germanoma	4qA161/166	46	22	A	166	44	A	161S	9	A	166	44	A	166
KLE	F	uterus adenocarcinoma	none	29	9	B	163/168	33	B	163/168	33	A	166/170	33	A	166/170
HCC-38	F	breast adenocarcinoma	4qA161	7	12	B	163	22	A	161	26	A	166	26	A	161
G-401	M	rhabdoid tumor	4qA166	16	13	B	163	49	A	166	7	A	166	40	AB	166
SK-N-MC	F	neuroepithelioma	none	21	12	B	163/168	23	B	163/168	35	A	166		deleted	
NALM-6	M	Adult B acute lymphoblastic leukemia	4qA	99	22	B	163	33	A	161	4	0	166	19	A	166
AMO-1	F	plasma cell meyloma	4qA	98	12	B	163	42	A	161	23	0	166	62	A	166
KMS-27	M	plasma cell meyloma	4qA	100	21	B	163	32	A	165	4	0	166	21	A	166
KM-H2	M	Hodgkin lymphoma	none	80	22	B	163		deleted		10	A	164	28	A	166
HELA	F	cervical tumor	none	97	9	B	168	9	B	0	21	A	166	39	A	166
MCF7	F	breast adenocarcinoma	?	89												

Figure 4.2.2. Cancer cell lines express DUX4 from multiple alleles, but hypomethylation of the D4Z4 is a unifying theme.

Columns: **(1-3)** Cell line name, sex of cell lines, and cancer type/root tissue adopted from CCLE. (**4)** Summary of SSLP data: type A alleles on chromosome 4 are considered permissive by FSHD standards. **(5)** CpG methylation of D4Z4 macrosatellites on chromosomes 4 and 10 was determined by sensitivity to FseI endonuclease. **(6-7, 9-10, 12-13 & 15-16)** Number of units in D4Z4 macrosatellites on chromosomes 4 and 10 as well as their haplotype as determined with Pulsed Field Gel electrophoresis Southern blotting and hybridization, and SSLP length analysis by PCR. **(11, 14, 17)**.

DUX4 mRNA uses exon 3a and 3b in iPSC, testis, and cancer cell lines versus exon 3 in FSHD myotubes.

Although the DUX4 ORF is encoded in exon 1, DUX4 3'UTR exons vary depending on the alleles it's transcribed from. To evaluate which exons are being utilized in different contexts we designed primers to target the different 3'UTR exon usage (Fig. 4.2.3A). RTqPCR for these different exons demonstrated that the canonical exon 3 associated with FSHD was the most utilized 3' exon in FSHD myotubes (MT200), whereas exon 2-3a or 2-3b exon usage was most common in all other cell types, including human testis, naïve iPSC and cancer cell lines (Fig. 4.2.3B). To see which alleles these exon 2-3a/3b were transcribed from, we sequenced multiple clones from at least 3 separate RT-PCRs on the cancer cell lines and used

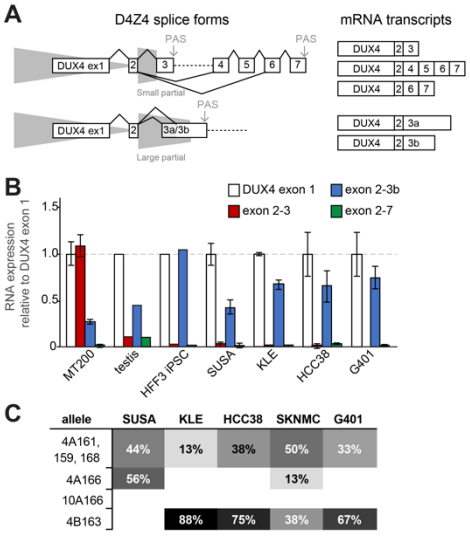


Figure 4.2.3 DUX4 transcripts use different exons and come from different alleles compared to FSHD

(A) Schematic of DUX4 splicing showing transcripts from terminal D4Z4 repeats. (B) RTqPCR for DUX4 exon usage on RNA from denoted cell types. (C) Percentage of transcripts coming from different alleles in each cell line: determined by exon 2 sequence disparities.

SNPs in exon 2 to discriminate between template alleles[20], The results demonstrated that these transcripts were coming from multiple alleles including ones from 4qB, which is not known to contain a PAS (Fig 4.2.3C).

4.3 Discussion

A unifying feature of FSHD is the combination of D4Z4 hypomethylation with a permissive allele. Permissive D4Z4 alleles include 4qA161 and 4qA166 which have polyadenylation signals (PAS) in exon 3, downstream of the terminal D4Z4 unit[33]. 4qA alleles and D4Z4 macrosatellites on chromosome 10 are not known to be permissive. DUX4 transcripts coming from chromosome 10 were identified in the testis[20], but the exon usage in the context of the early embryo and cancer is unknown.

The data presented in this chapter suggests that the majority of DUX4 transcripts in non-FSHD contexts are using exons 2-3b or exons 2-3a. In FSHD, only the terminal D4Z4 unit is known to drive functional DUX4 coding transcripts, by splicing into downstream sequences to access the exon 3 PAS, but these data are particularly interesting because all D4Z4 units contain 3a and 3b splice accepters and these transcripts could very well be coming from transcription of internal units.

Using 3'RACE on cancer cell line RNA, we did discover a novel PAS in intron 6 of 10q alleles, which is also downstream of the D4Z4 and suggests that the terminal D4Z4 unit may be transcribed to access this particular PAS, similar to DUX4 transcripts sequenced in the testis[20]. However, RTqPCR demonstrated that this DUX4 transcript was at much lower abundance than transcripts which used exons 2-3a/3b. Future studies should further interrogate the nature of these DUX4 transcripts. Which transcripts are driving DUX4 protein in non-FSHD contexts has yet to be determined.

Data in this chapter also evaluates the functionality of mutations in and around the splice acceptor of exon 3 in an FSHD context, demonstrating that certain SNPs can reduce splicing efficiency of DUX4 in permissive alleles. Altering the splicing efficiency of exon 3 may reduce disease FSHD severity and seemingly would not have an effect on developmental DUX4 expression as this exon usage does not appear to be necessary for DUX4 in development. Current therapies for FSHD include targeting DUX4 exon 3 with siRNAs and the data here support that avenue, but suggest those therapies may not be effective in treating DUX4-expressing cancers.

4.4 Materials and Methods

Myoblast cultures

Primary human myoblast cell lines were obtained from 3 family members of family Rf204 (numbers from pedigree) and from 2 members of family Rf10 (numbers from pedigree) through the FSHD Research Center at UR Medicine. Primary myoblast cell lines were cultured in DMEM/F10 medium supplemented with 20% heat inactivated fetal bovine serum (FBS), 1% penicillin/streptomycin, 10 ng/ml rhFGF and 1μM dexamethasone. Myoblast cultures were differentiated into myotubes at 80% confluency by culturing them in DMEM/F-12 Glutamax medium supplemented with 1% penicillin/streptomycin and 2% knockout serum for approximately 36 hours. Immortalized WT human myoblasts (MB135) were grown in Ham's F-10 (Gibco) supplemented with 10% FBS (HyClone), 1% penicillin/streptomycin (Thermo Fisher Scientific), 10ng/mL rhFGF (Promega), 1uM dexamethasone (Sigma). All cell lines were cultured at 37°C in a humidified incubator supplied with 5% CO_2.

Transient transductions and RNA extraction

MB135 cultered in 6-well dishes with 2ml F-10 (as above) per well. Each well was transduced with 600ng pCS2 vector, 4uL Lipofectamine 3000 (Invitrogen) and 1.5uL P3000 diluted in 500uL Opti-MEM Reduced Serum Medium (Gibco) overnight. transfection reagents were washed off cells in the morning and allowed to recover for 24hr. RNA was harvested with NucleoSpin RNA kit (Machery-Nagel) according to manufacturer instructions. Isolated RNA was treated with Amp-grade DNase I (Invitrogen), heat inactivated, and reverse transcribed into cDNA using SuperScript IV First-Strand Synthesis System (Invitrogen) following the manufacturer's protocol.

Chapter 5. DUX4 positive autoregulation.

This chapter contains unpublished work.

5.1 Introduction

DUX4 is a potent pioneer early embryonic transcription factor that drives a totipotent program in the cleavage stage embryo and in embryonic stem cells[6,18]. DUX4 is also the causative gene of facioscapulohumeral dystrophy (FSHD)[32]. Currently, the universal mode of DUX4 activation in all of these contexts is believed to be a multifactorial process leading to hypomethylation of the D4Z4, but transcription factors that directly bind to and activate DUX4 expression are relatively unknown. p53 is a transcription factor that has been shown to also enhance DUX4 expression[58] but P53 doesn't appear to drive DUX4 expression when the D4Z4 is hypermethylated, it's normal state in somatic tissues.

Grow et al identified a Dux DNA-binding motif upstream of the Dux ORF and used ChIP-sequencing to demonstrate that Dux binding to this site[58]. These data suggest Dux is contributing to its own transcriptional output. Expressing mouse Dux in human cells and human DUX4 in mouse cells has demonstrated partially conserved transcriptional targets, due to their partially conserved DNA-binding motifs[16]. In this chapter, we report that DUX4 is autoregulated and that mouse Dux and DUX4 can interchangeably drive transcription in murine and human cells.

5.2 Results

5.2.1 Inducible mDux and DUX4 drive expression of endogenous mDux/DUX4

To evaluate if DUX4 is positively autoregulated we employed the use of engineered human myoblasts with an doxycycline-inducible codon-altered DUX4 (MB135-i*DUX4*^{CA})[57] and a codon-altered DUX4 DNA-binding mutant i*DUX4*^{F67A41}. Treating these cells with doxycycline for 12hrs induced *DUX4*^{CA} expression in both lines to similar levels, but wild type *DUX4* (*DUX4*^{WT}) expression was only induced in the line with DNA-binding DUX4 (Fig 5.2.1). These data support DUX4 autoregulation.

To evaluate if mouse Dux (mDux) can drive expression of human DUX4 we engineered human myoblast cell lines containing a doxycycline-inducible codon-altered mouse *Dux* (MB135-i*mDux*^{CA}) and to see if human DUX4 can activate mouse Dux myoblasts, we generated

a C2C12- i*DUX4*^{CA} line. PCR confirmed that inducing either i*mDux*^{CA} or i*DUX4*^{CA} drives expression of endogenous *DUX4*^{WT} in human myoblasts and that either can also drive expression of endogenous *mDux*^{WT} in mouse myoblasts (Fig 5.2.1B). RT-PCR confirmed that both i*mDux*^{CA} and i*DUX4*^{CA} drive expression of endogenous *DUX4*^{WT} in human myoblasts similar levels (Fig 6.2.1C) and both drive expression of wt *mDux* in mouse myoblasts at similar levels (Fig 6.2.1D). Inducing Dux expression for 12hrs in the MB135-i*mDux*^{CA} line and then staining for mDux expression and DUX4 expression demonstrated that i*mDux*^{CA} drives endogenous DUX4 protein expression in human cells either directly or indirectly (Fig 5.2.1E).

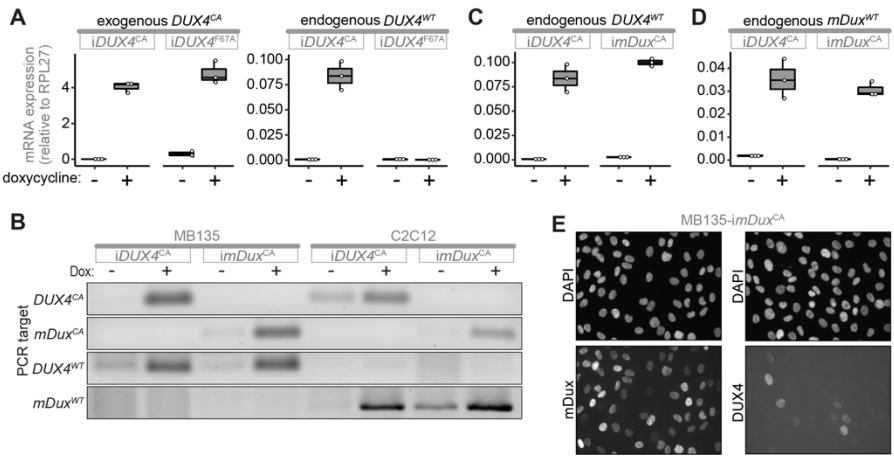


Figure 5.2.1: Codon-altered mDux and DUX4 drive expression of wild-type mDux/DUX4
(A) RTqPCR for i*mDux*^{CA} and *mDux*^{WT} on MB135-i*mDux*^{CA} and MB135-i*mDux*^{F67A} human myoblast lines after 12hrs of doxycycline. **(B)** Agarose gel of RT-PCR products for wildtype and codon altered versions of mouse Dux or DUX4 in denoted doxycycline-inducible myoblast lines after 12hrs of dox. **(C)** RTqPCR for i*mDux*^{CA} and *mDux*^{WT} on MB135-i*mDux*^{CA} and MB135-i*mDux*^{F67A} human myoblast lines aftyer 12hrs of dox. **(D)** RTqPCR for i*mDux*^{CA} and *mDux*^{WT} on MB135-i*mDux*^{CA} and MB135-i*mDux*^{F67A} human myoblast lines after 12hrs of dox. **(E)** Immunofluorescent staining mDux and DUX4 in MB135-i*mDux*^{CA} cells after 12hrs dox.

5.2.2 mDux and DUX4 and have upstream Dux family binding sites

In the mouse genome, there is a perfect mouse Dux binding motif 547bp upstream of the *mDux* translational start codon of the open reading frame ([58] and Fig 5.2.2A). To evaluate if mDux and DUX4 are directly driving transcription of *mDux*, we engineered a series of constructs with different segments of the upstream region of *mDux*, each driving expression of the luciferase gene (Fig 5.2.2B) in pGL3-basic reporters. We co-transfected human myoblasts with these reporters and either pCS2-mDux or pCS2-DUX4. A dual reporter luciferase assay demonstrated that both mDux and DUX4 drove more luciferase expression in the reduced bind site reporter than in the longer promoter reporter, and when we deleted the 9bp binding motif from this construct, luciferase expression was greatly reduced (Fig 5.2.2C). These data suggest direct DNA-binding of mDux and DUX4 at this mDux motif promotes transcription of *mDux*.

To determine where these proteins are binding upstream of DUX4 in the human genome, we took advantage of pre-existing DUX4 ChIP experiments from human myoblasts with a doxycycline-inducible DUX4[21]. When DUX4 expression was induced in these cells, we observed 4 ChIP peaks (1-4) representing potential binding sites (Fig 5.2.2D). Peaks 1 and 2 also existed in the –Dox sample so we focused on the region containing peaks 3 and 4.

To determine if mDux and DUX4 were binding to this region upstream of *DUX4*, we engineered a series of constructs with different segments of the upstream region of *DUX4*, each driving expression of the luciferase gene in the pGL3b (Fig 5.2.2E). After these cells were co-transfected with these constructs and pCS2-*DUX4* or -*mDux*, A dual reporter luciferase assay demonstrated that both mDux and DUX4 drove more luciferase expression in the reduced bind site reporter than in the longer promoter reporter, and when we deleted the 9bp binding motif from this construct, luciferase expression was greatly reduced (Fig 5.2.2F). These data suggest direct DNA-binding of mDux and DUX4 at this mDux motif promotes transcription of *mDux*.

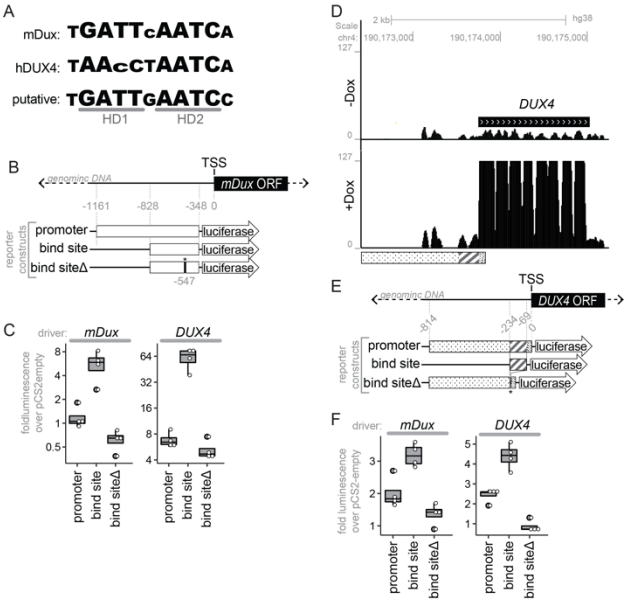


Figure 5.2.2: mDux and DUX4 and have upstream Dux family binding sites
(A) Predicted DUX family binding sites upstream of *mDux*. **(B)** Schematic mDux coding regions and constructs made with upstream sequences. **(C)** Illuminometer readings of firefly normalized to renilla expressed as fold over pCS2-empty drivers with the respective reporter constructs (n=4 separate wells) in human myoblasts. **(D)** Tracks of DUX4 ChIP reads from human myoblasts with doxycycline-inducible DUX4. **(E)** Schematic mDux coding regions and constructs made with upstream sequences. **(F)** Illuminometer readings of firefly normalized to renilla expressed as fold over pCS2-empty drivers with the respective reporter constructs (n=4 separate wells) in mouse myoblasts.

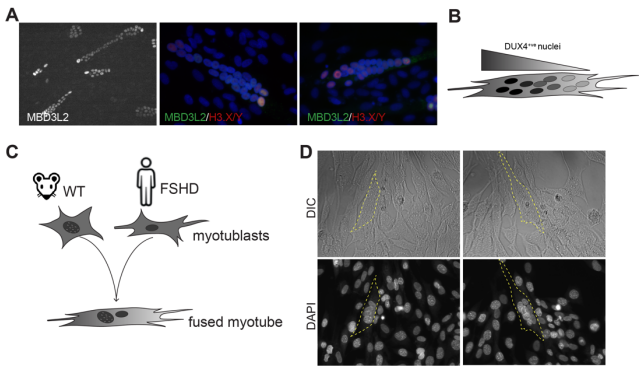


Figure 5.2.3: DUX4 expression gradient in FSHD myotubes and generation of chimeric myotubes.
(A) Immunofluorescence for DUX4 transcriptional targets in FSHD myotube culture. **(B)** Schematic of DUX4 expression spreading across nuclei in FSHD myotubes. **(C)** Schematic of chimeric myotube generation. **(D)** DAPI stained chimeric myotubes (dashed yellow line outlines a single myotube in each frame)

5.2.3 Endogenous DUX4 expression can drive expression of endogenous mDux in chimeric myotubes

FSHD patient-derived myoblast lines can be differentiated into myotubes that express high levels of DUX4[52]. Staining for transcriptional targets of DUX4 in these FSHD myotubes demonstrated a gradient of nuclear expression (Fig 5.2.3A, and illustrated in Fig

5.2.3B). To evaluate if endogenous DUX4 can dive expression of endogenous mDux in adjacent nuclei of the same cell body, we differentiated human FSHD myoblasts mixed with wild-type mouse myoblasts and achieved chimeric myotubes with both mouse and human nuclei (Fig 5.2.3C-D). Immunostaining for the DUX4 target H3Y confirmed that, while FSHD myoblasts do express H3Y (an indicator of DUX4 expression) and mouse myoblasts don't express mDux, fused chimeric myotubes express both H3Y and mDux (Fig 5.2.4A). RT-qPCR confirmed that chimeric myotubes increased *mDux* expression (Fig 5.2.4B). These data support the hypothesis the autoregulation of mDux/DUX4 can occur across nuclei in a shared cell body.

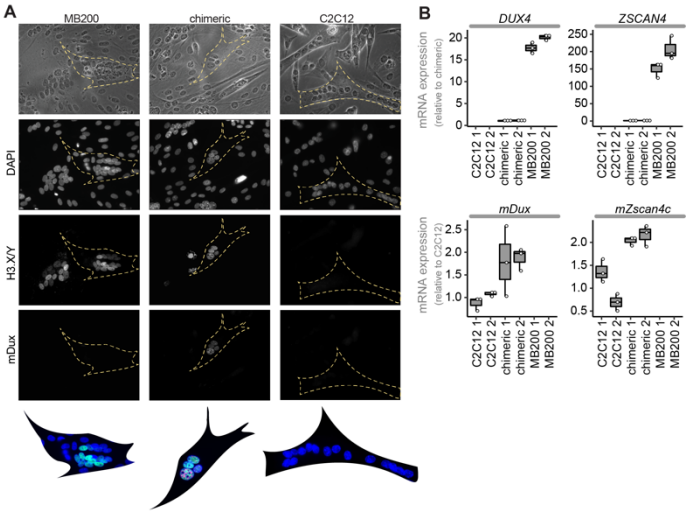


Figure 5.2.4: Endogenous DUX4 expression can drives expression of endogenous mouse Dux in chimeric myotubes. **(A)** Immunofluorescence for mDux and the DUX4 transcriptional target H3Y in in human-mouse chimeric myotubes. **(B)** RTqPCR for human DUX4 or mouse Dux and their respective Zscan4 target genes in human-mouse chimeric myotubes.

5.3 Discussion

The current understanding of DUX4 expression relies on the multifactorial hypomethylation of the D4Z4 macrosatellite which encodes DUX4, but transcription factors that directly bind to and activate DUX4 expression are relatively unknown. p53 is a transcription factor that has been shown to enhance DUX4 expression[58] but P53 doesn't appear to drive DUX4 expression when the D4Z4 is hypermethylated, it's normal state in somatic tissues. The data presented here demonstrate that both pioneer transcription factors, DUX4 and mouse Dux proteins, directly bind to and activate transcription of the loci that encode them, even in control myoblasts which have hypermethylated D4Z4s. Autoregulation has been demonstrated in other master regulator genes such as MyoD[90], but the DUX4 and mouse DUX autoregulation reported here mark the first reported direct activation of these genes.

DUX4 is known as the causative gene in FSHD, in which it drives cell death in myotubes, however, it is also known to be expressed in the early embryo, testis, a subset of cancers and subset naïve pluripotent stem cells in culture, where the cells tolerate its expression and survive. The field has been searching for an explanation for these differences in the functional outcome of DUX4 expression. These data herein layout a model whereby DUX4 expression can be amplified in multinucleated myotubes via DUX4 autoregulation: DUX4

transcripts from one nucleus may be translated into protein which diffuses into adjacent nuclei and activates DUX4 expression in a gradient across the syncytium. Physiological DUX4 expression appears to be transient[91] whereas the aberrant DUX4 expression observed in FSHD myotubes appears to be at abnormally high levels which is recapitulated by DUX4 overexpression systems that result in cell death.

DUX4 autoregulation may also add to the potency of DUX4 expression as it is observed that low levels of transient DUX4 expression have profound and long-lived changes in transcriptomes regardless of cell type. Taken together, this body of work furthers the understanding of the transcriptional activation of DUX4 and may provide insight into the different functional outcomes of DUX4 expression.

5.4 Materials and Methods

Mice

Cell Culture

MB135 and C2C12 myoblast lines were cultured in F-10 10% FBS 1% penicillin/streptomycin. All cultures were maintained at 37C in 5%CO2

Tissue embedding

Immunofluorescence

Cultured cells were fixed in 2% paraformaldehyde 10min, and washed with PBS. Cells were then permeabilized in 0.1% Triton X-100 in PBS for 10 minutes at room temperature. Cells were then incubated overnight at 4°C with DUX4 [E14-3] primary antibody (1:400 dilution) or mDux [E9-generated in house] (1:333 dilution) and washed with PBS three times for 10 minutes each. Cells were incubated in TRITC-conjugated donkey anti-rabbit secondary antibody (1:666 dilution) and TRITC-conjugated donkey anti-rat secondary antibody (1:666 dilution) for 3 hours at room temperature. Cells were washed with PBS, stained with DAPI for 10 minutes at room temperature, and washed with PBS. Cells were imaged with a fluorescence microscope.

Reverse transcription-quantitative PCR (RT-qPCR)

Tissue was pulverized with a mortar and pestle after snap freezing in liquid nitrogen. RNA from pulverized tissue and cultured cells was harvested with the NucleoSpin RNA Kit (Takara) according to the manufacturer's protocol. RNA quality was verified by NanoDrop 2000 (Thermo Scientific). RNA was treated with DNaseI, Amplification Grade (Invitrogen). Reverse transcription was performed in a 20uL reaction: 200-1000ng whole RNA, 1uL dNTP (10mM), 1uL oligo dT primer (10mM), 4uL 5x SSIV Buffer, 1uL DTT (100mM), 1 uL RNaseOUT, and 1uL SSIV RT enzyme. Thermal cycling conditions for reverse transcription were as follows: 50°C for 40 minutes, 55°C for 30 minutes, and 80°C for 10 minutes. Complementary DNA (cDNA) was treated with 1uL of RNaseH and incubated at 37°C for 20 minutes, then diluted 1:5 or 1:4 with RNase-free H2O. Quantitative real-time PCR (qPCR) was performed on the QuantStudio™ 7 Flex Real-Time PCR System in a 10uL reaction: 2uL cDNA, 5uL 2x iTaq Universal SYBR Green Supermix, 0.3uL forward and reverse primer (10uM), and 2.4uL H2O. qPCR primers were synthesized by Integrated DNA Technologies (IDT) and are listed in Star Methods table. Thermal cycling conditions for qPCR were as follows: 50°C for 2 minutes and 95°C for 10 minutes; 40 cycles of 95°C for 15 seconds and 60°C for 60 seconds. For each Biological replicate, qPCR reactions were run in technical triplicates, including -RT controls. Median CT values of the technical triplicates were used for analysis. Gene expression was normalized to the housekeeping gene RPL27 (ribosomal protein L27). P-values were calculated with an unpaired, one-tailed student's t-test assuming unequal variance (type 3).

Generation of cell lines with dox-inducible codon altered Dux transgene.

Polyclonal C2C12-*imDux*^{CA} and MB135-*imDux*^{CA} cell lines were generated with lentiviral transduction, selected, and maintained in 3µg/mL puromycin as previously described for MB135-*iDUX4*^{CA} cells in Jagannathan et al., 2016.

Myotube differentiation.

Cells were grown to ~95% confluence and then switched to DMEM 10ug/mL insulin 10ug/mL transferrin without FBS. Cells were cultured for 48-17hrs until visible myotubes were present before harvesting for RNA or fixing for immunodetection. Mixed cultures were plated with 25% C2C12 and 75% MB200 before differentiating.

Luciferase Assays.

Fragments indicated were cloned into the pGL3b backbone and reverse transfected with Lipofectamine 3000 (ThermoFisher, L3000015) according to manufacturer's protocol into MB135 in 96-well plates (PerkinElmer, 6005070). Briefly: 5uL OPtiMEM was combined with 0.15uL lipofectamine and 5ul OptiMEM was combined with 100ng of plasmid DNA (40ng pCS2-DUX4 or pCS2-mDux, 40ng pGL3b, and 20ng pCS2 luciferase) and 0.2uL P3000 per well. These cocktails were mixed at 1:1 ratio and allowed to incubate 10min at RT before being aliquoted to individual wells, then cells were seeded at $0.03x10^6$ cells/well . 24hrs after transfection, reagents were washed off and cells were allowed to recover for 36hrs before processing with the Dual Luciferase reporter assay system (Promega , E1980) according to manufacturer's protocol. Luciferase levels were quantified with a …Luminometer and were normalized to levels of renilla.

Chapter 6. DUX4 RNA often lacks a polyA tail and is largely localized to the nucleus.

This chapter contains unpublished work.

6.1 Introduction

DUX4 is an early embryonic transcription factor that drives ZGA in the early embryo[18], it's also expressed in the germ cells of the adult testis[20] and is the causative gene of Facioscapulohumeral dystrophy (FSHD)[32]. *DUX4* is encoded in the D4Z4 macrosatellite repeat which doesn't encode a known polyadenylation signal (PAS)[33]. A unifying feature of DUX4 expression in FSHD is the functional PAS encoded downstream of the D4Z4 in the pLam region which the terminal D4Z4 unit uses to stabilize *DUX4* mRNA and drive expression of DUX4 protein[33]. However, not all alleles carry this PAS, but seemingly all embryos and male germ cells express DUX4. Furthermore, not all DUX4-FL cancer cell lines have a 4qA allele that contains a known PAS. Therefore, investigated whether DUX4 could be expressed as a non-polyadenylated (pA-) RNA.

In this chapter, we explore the nature of DUX4 mRNA in a variety of cell types and show that DUX4 mRNA is largely pA- in non-FSHD contexts and is mostly localized to the nucleus in all contexts.

6.2 Results

6.2.1 DUX4 RNA is largely pA-

To determine DUX4 transcripts, in different contexts, contain pA tails, we extracted RNA from FSHD1 myotubes (MB073) and naïve-state iPSC (derived from human foreskin fibroblasts: HFF3)[92] and used oligo dT magnetic beads to enrich for pA+ transcripts. We generated cDNA form the resulting pA+/pA- fractions with oligo dT or random hexamer primers. RT-qPCR confirmed that *RPL27* was enriched in the pA+ fractions for both cell lines and, while *DUX4* exon 1 was also enriched in the pA+ fraction from FSHD1 myotubes, in iPSC *DUX4* exon 1 was enriched in the pA- fraction (Fig. 6.2.1A-B). Using this same method, we evaluated *DUX4* polyadenylation in RNA extracted from human testis and DUX4-FL cancer cell lines and found that, besides FSHD myotubes, *DUX4* exon 1 is largely pA-, similar to pA- *H2Bk* (Fig 6.2.1C-D). Evaluating the relative expression of known *DUX4* splice forms across these conditions, demonstrated that, while transcripts that contained exons 2-3 (FSHD specific) or exons 2-6-7 (testis specific) was largely pA+, it was exon 2-3b usage (the major *DUX4* isoform in non-FSHD

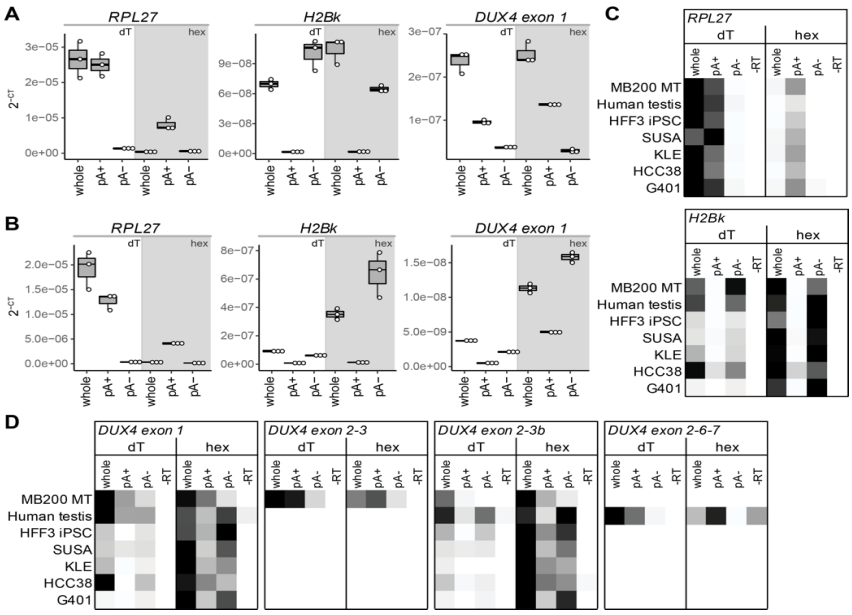


Figure 6.2.1 DUX4 RNA lacks a poly A tail
(A) RT-qPCR on pA-/pA+ fractions of MB073 with oligo dT- or random hexamer-primed RT. **(B)** RT-qPCR on pA-/pA+ fractions of naïve-state HFF3 iPSC with oligo dT- or random hexamer-primed RT. **(C)** Heatmaps displaying relative expression levels of DUX4 and control transcripts measured by RTqPCR after pA-pulldown and priming with either oligo dT or random hexamers.

contexts; Chapter 4) that appeared to be largely pA-. Taken together, these data suggest that that pA+ *DUX4* RNA may be unique to FSHD contexts.

6.2.2 DUX4 RNA is not circular

DUX4 is encoded in a macrosatellite repeat and, as it appears to be largely pA- in non-FSHD contexts, it is a great candidate for forming circular RNA (circRNA)[93,94]. For example, the same splice donors and acceptors known to generate exon 1-2-3a and 1-2-3b-spliced *DUX4* mRNA from the terminal D4Z4 repeat (Fig 6.2.2A) could both be used to splice transcripts from internal repeats (Fig 6.6.2.B) as well as used in backsplicing events to generate *DUX4* circRNA (Fig 6.2.2C). To evaluate if DUX4 RNA is circular we inversed primers at the 5' and 3' end of DUX4 exon 1 for PCR (Fig. 6.2.2C; red arrows) and sequenced promising bands. Although we did get PCR products, none of these sequence contained a reasonable splice junction and treating RNA with RNaseR, which digests linear RNA, demonstrated complete loss of these products (Fig 6.2.2D). These data indicate that DUX4 transcripts are not circRNA in DUX4-FL cancer cell lines.

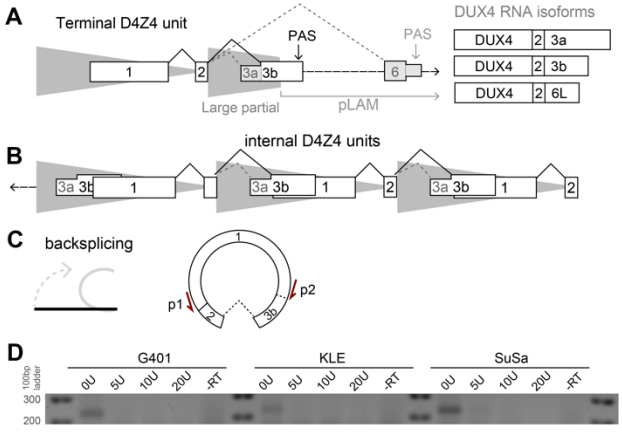


Figure 6.2.2 RNaseR treatment abolished DUX4 inverse RT-PCR bands

(A-C) Schematics demonstrating the splicing of transcripts from the D4Z4. **(A)** Transcripts from the terminal D4Z4 repeat. **(B)** Transcripts from internal D4Z4 repeats. **(C)** Theoretical back-splicing of transcripts from internal D4Z4 repeats with inverse primers (red). **(D)** Agarose gel of inverse RT-PCR products after treatment with denoted concentrations of RNaseR for 30min at 37C.

6.2.3 DUX4 mRNA is enriched in the nucleus

To confirm that DUX4 transcripts were being translocated to the cytoplasm, we fractionated the nucleus from the cytoplasm in FSHD1 myoblasts (MB073) and were surprised to find that the majority of DUX4 transcripts were associated with the nuclear fraction, similar to 5S rRNA and human satellite repeat II (HSATII) (Fig. 6.2.3A). To evaluate whether the endoplasmic reticulum (ER) was in the nuclear or cytoplasmic fractions, we extracted protein used Western Blotting to detect the ER-associated protein GPCR94. While the histone H3 protein was enriched in the nuclear fraction, the ER protein GPCR94 was enriched in the cytoplasmic fraction, similar to the cytoplasmic GAPDH protein (Fig. 6.2.3B). RT-qPCR on nuclear versus cytoplasmic fractions of DUX4-FL cancer cell lines revealed that DUX4 was enriched in the nucleus across all samples evaluated (Fig 6.2.3C). These data suggest that *DUX4* mRNA is enriched in the nucleus regardless of cell type or pA status.

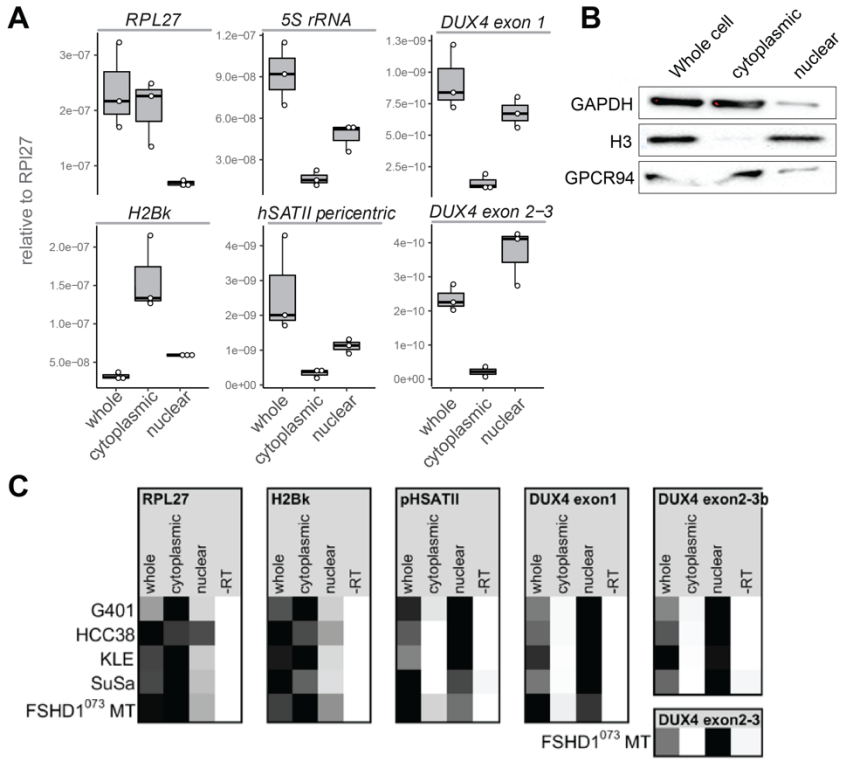


Figure 6.2.3: DUX4 mRNA is enriched in the nucleus
(A) RTqPCR on fractionated MB073 myoblasts for the denoted genes. **(B)** Western Blot of MB073 fractions. **(C)** RTqPCR heatmap summaries demonstrate relative expression of cell line fractionations.

6.2.4 DUX4 subcellular localization is likely sequence dependent

To better evaluate if both pA- and pA+ *DUX4* RNA is localized to the nucleus, we used oligo dT-coated magnetic beads to enrich for pA+ transcripts in nuclear and cytoplasmic fractions from the G401 cell line. RT-qPCR demonstrated that both pA- and pA+ *DUX4* RNA were enriched in the nuclear fractions from the G401 cell line (Fig 6.2.4A). To see if DUX4-like RNAs are also localized to the nucleus we fractionated cells from DUX4-like cancer cell lines. RTqPCR showed that DUX4-like transcripts from KMS27 and AMO1 cell lines, which are likely non-coding RNA, are enriched in the nucleus, while DUX4-like RNA from the NALM6 cell line,

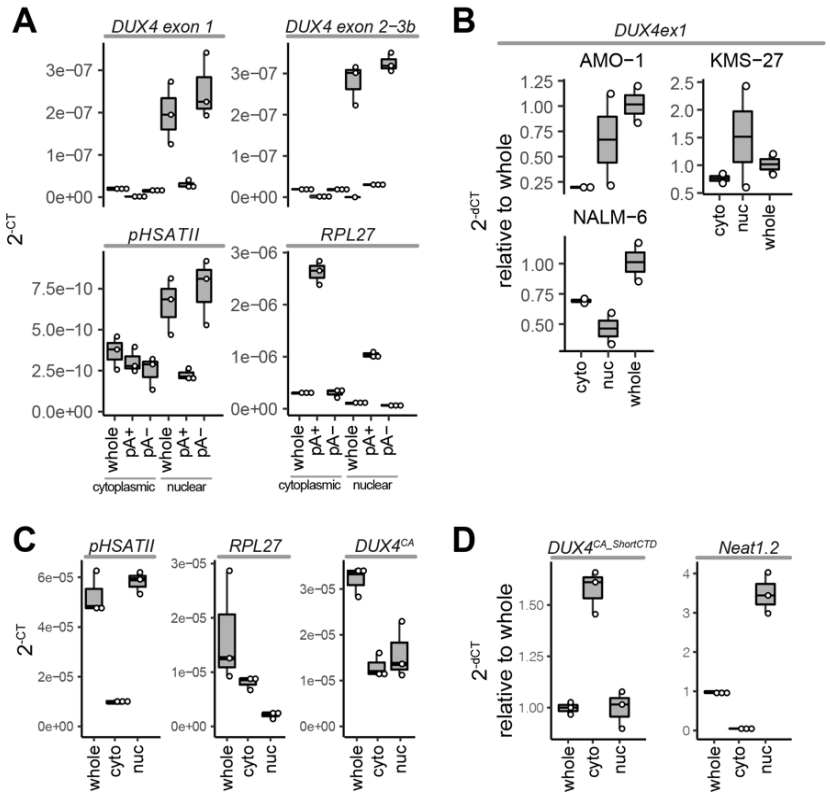


Figure 6.2.4: DUX4 nuclear retention is sequence-dependent

(A) RTqPCR on pA-pulldowns of G401 cellular fractions. **(B)** RTqPCR on fractions of cancer cell lines expressing DUX4-like transcripts. **(C)** RTqPCR on fractions of SUSA-i*DUX4*^{CA} pulsed with doxycycline for 4hrs harvested 4hrs after wash off. **(D)** RTqPCR on fractions of MN135-i*DUX4*^{shortCTD}.

which contains the homeodomain-coding region of DUX4 but the IgH C-terminal coding region, was enriched in the cytoplasmic fraction (Fig 6.2.4B). We therefore postulated that the nuclear localization of DUX4 transcripts was likely sequence dependent. To test this, we expressed a 26% codon-altered *DUX4* transcript (Appendix 1B) in MB135 (MB135-i*DUX4*^{CA})[57] for 8hrs and fractionated these cells. Relative expression of the nuclear versus cytoplasmic fractions showed that codon altered DUX4 was ~50% localized to the nucleus. We then utilized a truncated version of inducible codon-altered *DUX4,* only coding for the last 86aa of the C-terminal domain (*DUX4*^{shortCTD}*;* Appendix 1C),[41] and found it to be enriched in the cytoplasmic fraction in contrast to the non-coding, nuclear RNA *NEAT1.2* (Fig 6.2.4D). These data support the nuclear retention of *DUX4* being sequence dependent.

6.2.5 Loss of nuclear retention sequences does not reverse nuclear enrichment of DUX4 mRNA

We identified the nuclear retention signal AGCCC with sequence restrictions at positions −8 (T or A) and −3 (G or C)[95] repeated twice within the wild-type (WT) *DUX4* ORF (Appendix 1A) and once each in both the 5' and 3' UTRs. The inducible codon-altered DUX4 sequence that we previously demonstrated having ~50% nuclear retention (Fig 6.2.4C) retained one of these sequences in the open reading frame, so we further codon-altered it to remove this signal (*DUX4*$^{CA\text{-}NR}$; Appendix 1C) and generated 3 dox-inducible lines in Rhabdomyosarcoma cells (RD) expressing either i*GFP*, i*DUX4*^{CA}, or i*DUX4CA*^{noNR}. We treated these RD cell lines with

doxycycline for 4hrs followed by cyto/nuc fractionation. RT-qPCR on cellular fractions from the RD-i*GFP* line confirmed that these inducible trans genes behave as expected and their transcripts don't non-discriminately accumulate in the nucleus (Fig 6.2.5A). RT-qPCR on cellular fractions from the i$DUX4^{CA}$ or i$DUX4CA^{noNR}$ RD lines demonstrated that there was no significant change in the nuclear localization of *DUX4* mRNA when this nuclear retention signal was disrupted. (Fig 6.2.5B). To test if these was unique to human cells we genetically engineered mouse the myoblast C2C12 cell line to express either inducible $DUX4^{CA}$ *or* $DUX4^{CA\text{-}NR}$. RTqPCR in these lines demonstrated that the nuclear retention of both transcripts was increased in mouse lines but, again, not significantly different from each other (Fig 6.2.5C). These data demonstrate that we were unable to identify the sequence driving nuclear retention of DUX4 transcripts, and that this nuclear retention is not unique to human cells.

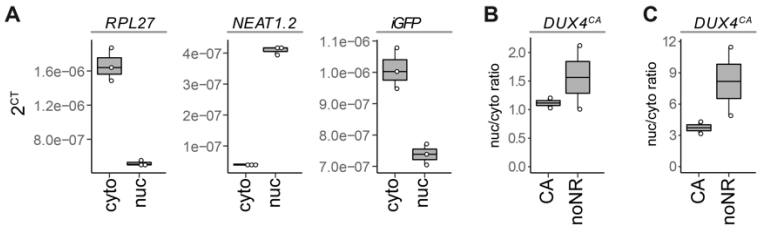


Figure 6.2.5: Loss of nuclear retention sequences does not reverse nuclear enrichment of DUX4 mRNA

(A) RT-qPCR on cellular fractions from the RD-i*GFP* line treated with dox. **(B)** RT-qPCR on cellular fractions from RD-i$DUX4^{CA}$, and RD-i$DUX4^{CA\text{-}NR}$ lines treated with dox; relative RNA graphed as ratio of nuclear to cytoplasmic abundance. **(C)** RT-qPCR on cellular fractions from C2C12-i$DUX4^{CA}$, and C2C12-i$DUX4^{CA\text{-}NR}$ lines treated with dox; relative RNA graphed as ratio of nuclear to cytoplasmic abundance.

6.3 Discussion

DUX4 expression is complicated because there are more than 100 genomic copies of DUX4 per arranged as repeats in the D4Z4 macrosatellite. These are genes expressed at low levels with high GC-content, which makes them hard to detect from the background noise of the genomic DUX4 sequence and, although all these DUX4fl transcripts code for the same protein, there are different splice forms coming from different alleles, few of which are known to actually use a poly A signal.

The fact that DUX4 is encoded in repetitive units, is largely pA- and is potent despite being lowly express, makes it a great candidate for making circular RNA, however our data do not support this model. So then how are these DUX4 transcripts making protein?

To confirm that these transcripts were being exported to the cytoplasm, we started doing nuclear/cytoplasmic fractionations and, to our surprise, we found that DUX4 transcripts are localized to the nucleus, even in the context of FSHD, where there is robust evidence that these transcripts are polyadenylated and translated into protein. This finding is further supported by [96](Figure 2C) which shows DUX4 transcripts enriched in the nucleus of FSHD myotubes.

Our data suggest that the nuclear retention of DUX4 could be sequence dependent. And although we did identify a putative nuclear retention signal, altering this signal had no effect on the nuclear localization of DUX4 transcripts. In addition to not knowing what sequence is driving the nuclear retention of DUX4 transcripts, we also don't know which transcripts are being translated into protein. Future directions in these regards should focus on defining which *DUX4* isoforms are driving DUX4 protein in addition to what sequence drives the nuclear retention of DUX4 transcripts.

6.4 Materials and Methods

Cell Culture

MB135, MB073 and MB200 myoblast lines were cultured in F-10 10% FBS 1% penicillin/streptomycin. SUSA AMO1, KMS27 and NALM6 were cultured in RPMI 10%FBS. G401 were cultered in McCoys 10%FBS, All cell lines were maintained at 37C with 5% CO2.

Western Blot

Cells were harvested in RIPA buffer containing protease and phosphatase inhibitors and sonicated on low for 5 minutes. The whole cell lysate was centrifuged at 12,0000 RCF for 10 minutes and the supernatant transferred to a new tube. Protein quantification was performed using the Pierce BCA Protein Assay Kit. Either 5ug or 10ug of whole cell lysate was heated in LDS Sample Buffer with 2.5% β-mercaptoethanol at 70°C for 10 minutes with shaking. For SDS-PAGE, proteins were loaded onto a 4-12% NuPAGE Bis-Tris gel and run at 100V for about 2 hours in NuPAGE MOPS SDS Running Buffer with 300uL NuPAGE antioxidant. Proteins were transferred to a 0.2µm PVDF membrane at 30V for 1 hour in NuPAGE Transfer Buffer with 10% methanol. Membranes were blocked in 5% milk in TBST for 1 hour, then incubated at 4°C overnight with rocking. Blots were washed three times (15 minutes each) with TBST and incubated with HRP-conjugated secondary antibody for 1 hour. Blots were washed three times (15 minutes each) with TBST, and bands were detected via chemiluminescence with either SuperSignal West Pico PLUS Substrate, SuperSignal West Femto Substrate, or SuperSignal West Atto Substrate.

Reverse transcription-quantitative PCR (RT-qPCR)

Tissue was pulverized with a mortar and pestle after snap freezing in liquid nitrogen. RNA from pulverized tissue and cultured cells was harvested with the NucleoSpin RNA Kit (Takara) according to the manufacturer's protocol. RNA quality was verified by NanoDrop 2000 (Thermo Scientific). RNA was treated with DNaseI, Amplification Grade (Invitrogen). Reverse transcription was performed in a 20uL reaction: 200-1000ng whole RNA, 1uL dNTP (10mM), 1uL oligo dT primer (10mM), 4uL 5x SSIV Buffer, 1uL DTT (100mM), 1 uL RNaseOUT, and 1uL SSIV RT enzyme. Thermal cycling conditions for reverse transcription were as follows: 50°C for 40 minutes, 55°C for 30 minutes, and 80°C for 10 minutes. Complementary DNA (cDNA) was treated with 1uL of RNaseH and incubated at 37°C for 20 minutes, then diluted 1:5 or 1:4 with RNase-free H2O. Quantitative real-time PCR (qPCR) was performed on the QuantStudio™ 7 Flex Real-Time PCR System in a 10uL reaction: 2uL cDNA, 5uL 2x iTaq Universal SYBR Green Supermix, 0.3uL forward and reverse primer (10uM), and 2.4uL H2O. qPCR primers were synthesized by Integrated DNA Technologies (IDT) and are listed in Star Methods table. Thermal cycling conditions for qPCR were as follows: 50°C for 2 minutes and 95°C for 10 minutes; 40 cycles of 95°C for 15 seconds and 60°C for 60 seconds. For each Biological replicate, qPCR reactions were run in technical triplicates, including -RT controls. Median CT values of the technical triplicates were used for analysis. Gene expression was normalized to the housekeeping gene RPL27 (ribosomal protein L27). P-values were calculated with an unpaired, one-tailed student's t-test assuming unequal variance (type 3).

Generation of cell lines with dox-inducible codon altered Dux transgene.

Polyclonal C2C12-*imDux*^{CA} and MB135-*imDux*^{CA} cell lines were generated with lentiviral transduction, selected, and maintained in 3µg/mL puromycin as previously described for MB135-*iDUX4*^{CA} cells in Jagannathan et al., 2016.

PolyA enrichment

Tissue was digested for 30min in Collagenase A, dispase I, and DNase in RPMI 30min at 37C strained through a 70uM filter and pelleted at 300RCF. cells were resuspended in PBS 1%FBS and stained with 1:50 APC/FIRE or BV605 mouse anti-mouse cKit (Biolegend) 20-30min at 4°C in dark. Cells were washed with PBS and filtered through 35uM strainer caps before analyzing and sorting with the BD Symphony 6. Cells were sorted into 4°C media and pelleted at 300RCF before being processed for RT-qPCR as above. Flow analyses were repeated for publication using FlowJo10.

Nuclear/Cytoplasmic Fractionation

Tissue was digested for 30min in Collagenase A, dispase I, and DNase in RPMI 30min at 37C strained through a 70uM filter and pelleted at 300RCF. cells were resuspended in PBS 1%FBS and stained with 1:50 APC/FIRE or BV605 mouse anti-mouse cKit (Biolegend) 20-30min at 4°C in dark. Cells were washed with PBS and filtered through 35uM strainer caps before analyzing and sorting with the BD Symphony 6. Cells were sorted into 4°C media and pelleted at 300RCF before being processed for RT-qPCR as above. Flow analyses were repeated for publication using FlowJo10.

Chapter 7: Discussion

DUX4 is a potent driver of the totipotency program, but its mechanisms of expression are not well understood, partly because DUX4 genetics are complicated. There are 100s of DUX4 open reading frames (ORFs) in the human genome, but the data presented here suggest that there is a clear difference in a developmental *DUX4* transcript from that of a FSHD *DUX4* transcript. The *DUX4* mRNA detected in testis iPSC and in cancer cell lines mainly consists of exons 1-2-3a or 1-2-3b and largely lacks a poly A tail, whereas the majority of DUX4 mRNA detected in FSHD lines mainly utilizes the exons 1-2-3 and is pA+. This is important to know when detecting DUX4 expression and the RNA level, because targeting spliced *DUX4* transcript can increase accuracy and using random hexamers to prime cDNA for PCR rather than oligo dT can increase the sensitivity of the assay. Furthermore, these data demonstrate that developmental *DUX4* transcripts are likely being transcribed from internal D4Z4 repeats on chromosomes 4 and 10, while in the context of FSHD, *DUX4* is transcribed from the terminal D4Z4 repeat on chromosome 4. Thus, the transcriptional activation of DUX4 may differ in development and cancer versus FSHD.

The transcriptional activation of DUX4 is a complicated multifactorial process. Reduction in D4Z4 copy number and/or mutations in chromatin modifiers of the D4Z4 can reduce its methylation status, poising cells to more easily express DUX4; cell lines with hypomethylated D4Z4s contain rare cells transiently expressing DUX4 and are more sensitive to p53 activation of DUX4. How can such a lowly and transiently expressed gene have such a strong impact on cell state? DUX4 amplifies itself in two ways: 1) DUX4 protein binds to upstream of the DUX4 ORF and drives transcription of itself (data herein), and 2) DUX4 drives transcription of noncanonical histones H3Y, which lack tails and incorporate into DUX4 transcriptional target genes keeping them open for prolonged expression[91]. Furthermore, DUX4 is master regulator gene, driving expression of hundreds of target genes[16,21].

It is interesting that, in FSHD myotubes, DUX4 and target genes can be detected spreading across nuclei in a gradient demonstrating its self-perpetuation within a syncytium. This may be the reason that FSHD myotubes express such higher levels of DUX4 than cells with a single nucleus (FSHD myoblasts, iPSC, and cancer cell lines). Furthermore, this excessive expression of DUX4 may be the reason FSHD myotubes die, whereas cells that are transiently expressing DUX4 appear to live and divide.

DUX4 expression has been reported to inhibit cells from responding to certain immune stimuli and has been well demonstrated to downregulate MHC-I expression, both of which

suggest the endowment of an immune evasion phenotype. Although more studies need to be done to determine the functional immune evasion phenotype of DUX4+ cells, it is interesting that we detected Dux+ trophoblast giant cells in murine placentas at the fetal-maternal interface. Cells of the placenta are known to have reduced MHC-I expression and to modulate the maternal immune system. It is attractive to think that DUX4 drives fetal immune evasion properties in normal development during implantation. Using Dux -/- mice, future studies will elucidate the necessity of Dux expressing cells in the placenta and will help us better understand the function of DUX4 in human development.

Although the zygote could be considered totipotent, the zygote is borrowing a gene expression program from the oocyte and the DUX4-driven ZGA program marks the beginning of autonomous totipotency. DUX4 has been proposed to drive totipotency in stem cell culture, but the true totipotency of these cells have yet be challenged with directed differentiation and teratoma assays. The implications of this DUX4-driven totipotent program in cancers is not yet fully understood, but we show that DUX4+ cancer cells can take on an immune evasive phenotype and that certain chemotherapies and cell stress can promote this state. These data together with the correlated increased plasticity/potency of undifferentiated cells does not bode well for DUX4+ cancer prognosis. Although these DUX4+ cells only make up a small fraction of cancer cells in a dish or solid tumor, these types of characteristics could make DUX4+ cancer cells a posterchild for surviving drug bottlenecks and driving cancer progression. I believe that the transient DUX4+ state marks a new type of cancer stem cell.

The bodies of work presented here are multifaceted and beg to be further studied in multiple regards. Regarding mouse mDux in development (Chapter 2), our working hypothesis is that mDux expressed in placental cells modulates the mother's immune system and promotes embryo engraftment, therefore future studies should focus on defining the mechanistic effects of mDux expression in the placenta. Regarding DUX4 expression in cancers (Chapter 3), future directions need to implement lineage tracing models, organoids, and use patient derived xenograft models to trace the cells in solid tumor environments with and without chemotherapy treatments to see if DUX4 positive cells are truly driving cancer progression. Regarding DUX4 positive autoregulation (Chapter 5), future research should concretely define the DUX4 binding site within the D4Z4. Finally, regarding the pA- and nuclear *DUX4* transcripts (Chapter 6), it is still unknown whether these are translated into DUX4 protein, therefore future studies should define which *DUX4* transcripts are driving DUX4 protein expression in non-FSHD contexts. There is still much to understand about DUX4 expression in development and cancer. These

proposed research directions will further our basic understanding of DUX4 expression and may lead to better treatments for DUX4+ cancers and patients with FSHD.

References

1. Jukam, D., Shariati, S.A.M., and Skotheim, J.M. (2017). Zygotic Genome Activation in Vertebrates. Dev Cell *42*, 316-332. 10.1016/j.devcel.2017.07.026.
2. Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. Nat Rev Mol Cell Biol *10*, 91-103. 10.1038/nrm2618.
3. Marikawa, Y., and Alarcon, V.B. (2009). Establishment of trophectoderm and inner cell mass lineages in the mouse embryo. Mol Reprod Dev *76*, 1019-1032. 10.1002/mrd.21057.
4. Gurtner, G.C., Werner, S., Barrandon, Y., and Longaker, M.T. (2008). Wound repair and regeneration. Nature *453*, 314-321. 10.1038/nature07039.
5. Bacakova, L., Zarubova, J., Travnickova, M., Musilkova, J., Pajorova, J., Slepicka, P., Kasalkova, N.S., Svorcik, V., Kolska, Z., Motarjemi, H., and Molitor, M. (2018). Stem cells: their source, potency and use in regenerative therapies with focus on adipose-derived stem cells - a review. Biotechnol Adv *36*, 1111-1126. 10.1016/j.biotechadv.2018.03.011.
6. Taubenschmid-Stowers, J., Rostovskaya, M., Santos, F., Ljung, S., Argelaguet, R., Krueger, F., Nichols, J., and Reik, W. (2022). 8C-like cells capture the human zygotic genome activation program in vitro. Cell Stem Cell *29*, 449-459 e446. 10.1016/j.stem.2022.01.014.
7. D'Alterio, C., Scala, S., Sozzi, G., Roz, L., and Bertolini, G. (2020). Paradoxical effects of chemotherapy on tumor relapse and metastasis promotion. Semin Cancer Biol *60*, 351-361. 10.1016/j.semcancer.2019.08.019.
8. Korentzelos, D., Clark, A.M., and Wells, A. (2020). A Perspective on Therapeutic Pan-Resistance in Metastatic Cancer. Int J Mol Sci *21*. 10.3390/ijms21197304.
9. Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., Bell, G.W., Regev, A., and Weinberg, R.A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet *40*, 499-507. 10.1038/ng.127.
10. Gupta, P.B., Pastushenko, I., Skibinski, A., Blanpain, C., and Kuperwasser, C. (2019). Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. Cell Stem Cell *24*, 65-78. 10.1016/j.stem.2018.11.011.
11. Chew, G.L., Campbell, A.E., De Neef, E., Sutliff, N.A., Shadle, S.C., Tapscott, S.J., and Bradley, R.K. (2019). DUX4 Suppresses MHC Class I to Promote Cancer Immune Evasion and Resistance to Checkpoint Blockade. Dev Cell *50*, 658-671 e657. 10.1016/j.devcel.2019.06.011.
12. Dhatchinamoorthy, K., Colbert, J.D., and Rock, K.L. (2021). Cancer Immune Evasion Through Loss of MHC Class I Antigen Presentation. Front Immunol *12*, 636568. 10.3389/fimmu.2021.636568.
13. De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat Genet *49*, 941-945. 10.1038/ng.3858.
14. Leidenroth, A., and Hewitt, J.E. (2010). A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. BMC Evol Biol *10*, 364. 10.1186/1471-2148-10-364.

15. Leidenroth, A., Clapp, J., Mitchell, L.M., Coneyworth, D., Dearden, F.L., Iannuzzi, L., and Hewitt, J.E. (2012). Evolution of DUX gene macrosatellites in placental mammals. Chromosoma *121*, 489-497. 10.1007/s00412-012-0380-y.
16. Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. Nat Genet *49*, 935-940. 10.1038/ng.3846.
17. Ren, W., Gao, L., Mou, Y., Deng, W., Hua, J., and Yang, F. (2022). DUX: One Transcription Factor Controls 2-Cell-like Fate. Int J Mol Sci *23*. 10.3390/ijms23042067.
18. Hendrickson, P.G., Dorais, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat Genet *49*, 925-934. 10.1038/ng.3844.
19. Eckersley-Maslin, M.A., Svensson, V., Krueger, C., Stubbs, T.M., Giehr, P., Krueger, F., Miragaia, R.J., Kyriakopoulos, C., Berrens, R.V., Milagre, I., et al. (2016). MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. Cell Rep *17*, 179-192. 10.1016/j.celrep.2016.08.087.
20. Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J., and Miller, D.G. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet *6*, e1001181. 10.1371/journal.pgen.1001181.
21. Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., and Tapscott, S.J. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Dev Cell *22*, 38-51. 10.1016/j.devcel.2011.11.013.
22. Tawil, R. (2008). Facioscapulohumeral muscular dystrophy. Neurotherapeutics *5*, 601-606. 10.1016/j.nurt.2008.07.005.
23. Wijmenga, C., Frants, R.R., Brouwer, O.F., Moerer, P., Weber, J.L., and Padberg, G.W. (1990). Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. Lancet *336*, 651-653. 10.1016/0140-6736(90)92148-b.
24. Wijmenga, C., Padberg, G.W., Moerer, P., Wiegant, J., Liem, L., Brouwer, O.F., Milner, E.C., Weber, J.L., van Ommen, G.B., Sandkuyl, L.A., and et al. (1991). Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. Genomics *9*, 570-575. 10.1016/0888-7543(91)90348-i.
25. Upadhyaya, M., Lunt, P., Sarfarazi, M., Broadhead, W., Farnham, J., and Harper, P.S. (1992). The mapping of chromosome 4q markers in relation to facioscapulohumeral muscular dystrophy (FSHD). Am J Hum Genet *51*, 404-410.
26. Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., and et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet *2*, 26-30. 10.1038/ng0992-26.
27. van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., and Frants, R.R. (1993). FSHD associated

DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum Mol Genet *2*, 2037-2042. 10.1093/hmg/2.12.2037.

28. Bakker, E., Wijmenga, C., Vossen, R.H., Padberg, G.W., Hewitt, J., van der Wielen, M., Rasmussen, K., and Frants, R.R. (1995). The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle Nerve Suppl *2*, S39-44.
29. Gabriels, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D., and Belayew, A. (1999). Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene *236*, 25-32. 10.1016/s0378-1119(99)00267-x.
30. Lemmers, R.J., de Kievit, P., Sandkuijl, L., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2002). Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. Nat Genet *32*, 235-236. 10.1038/ng999.
31. Lemmers, R.J., Wohlgemuth, M., Frants, R.R., Padberg, G.W., Morava, E., and van der Maarel, S.M. (2004). Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. Am J Hum Genet *75*, 1124-1130. 10.1086/426035.
32. Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O., et al. (2007). The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord *17*, 611-623. 10.1016/j.nmd.2007.04.002.
33. Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. Science *329*, 1650-1653. 10.1126/science.1189044.
34. Tian, L., Shao, Y., Nance, S., Dang, J., Xu, B., Ma, X., Li, Y., Ju, B., Dong, L., Newman, S., et al. (2019). Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia. Nat Commun *10*, 2789. 10.1038/s41467-019-10637-8.
35. Yoshimoto, T., Tanaka, M., Homme, M., Yamazaki, Y., Takazawa, Y., Antonescu, C.R., and Nakamura, T. (2017). CIC-DUX4 Induces Small Round Cell Sarcomas Distinct from Ewing Sarcoma. Cancer Res *77*, 2927-2937. 10.1158/0008-5472.CAN-16-3351.
36. Chen, Z., and Zhang, Y. (2019). Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. Nat Genet *51*, 947-951. 10.1038/s41588-019-0418-7.
37. De Iaco, A., Verp, S., Offner, S., Grun, D., and Trono, D. (2020). DUX is a non-essential synchronizer of zygotic genome activation. Development *147*. 10.1242/dev.177725.
38. Bosnakovski, D., Gearhart, M.D., Ho Choi, S., and Kyba, M. (2021). Dux facilitates post-implantation development, but is not essential for zygotic genome activationdagger. Biol Reprod *104*, 83-93. 10.1093/biolre/ioaa179.
39. Yaffe, D., and Saxel, O. (1977). Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature *270*, 725-727. 10.1038/270725a0.

40. Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascencao, K., Rummel, C., Ovchinnikova, S., et al. (2019). Gene expression across mammalian organ development. Nature *571*, 505-509. 10.1038/s41586-019-1338-5.
41. Spens, A.E., Sutliff, N.A., Bennett, S.R., Campbell, A.E., and Tapscott, S.J. (2023). Human DUX4 and mouse Dux interact with STAT1 and broadly inhibit interferon-stimulated gene induction. Elife *12*. 10.7554/eLife.82057.
42. Erlebacher, A. (2013). Immunology of the maternal-fetal interface. Annu Rev Immunol *31*, 387-411. 10.1146/annurev-immunol-032712-100003.
43. Meacham, C.E., and Morrison, S.J. (2013). Tumour heterogeneity and cancer cell plasticity. Nature *501*, 328-337. 10.1038/nature12624.
44. Dagogo-Jack, I., and Shaw, A.T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol *15*, 81-94. 10.1038/nrclinonc.2017.166.
45. Batlle, E., and Clevers, H. (2017). Cancer stem cells revisited. Nat Med *23*, 1124-1134. 10.1038/nm.4409.
46. Prasetyanti, P.R., and Medema, J.P. (2017). Intra-tumor heterogeneity from a cancer stem cell perspective. Mol Cancer *16*, 41. 10.1186/s12943-017-0600-4.
47. Akiyama, T., Xin, L., Oda, M., Sharov, A.A., Amano, M., Piao, Y., Cadet, J.S., Dudekula, D.B., Qian, Y., Wang, W., et al. (2015). Transient bursts of Zscan4 expression are accompanied by the rapid derepression of heterochromatin in mouse embryonic stem cells. DNA Res *22*, 307-318. 10.1093/dnares/dsv013.
48. Fu, X., Djekidel, M.N., and Zhang, Y. (2020). A transcriptional roadmap for 2C-like-to-pluripotent state transition. Sci Adv *6*, eaay5181. 10.1126/sciadv.aay5181.
49. Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2014). Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. Skelet Muscle *4*, 12. 10.1186/2044-5040-4-12.
50. Daxinger, L., Tapscott, S.J., and van der Maarel, S.M. (2015). Genetic and epigenetic contributors to FSHD. Curr Opin Genet Dev *33*, 56-61. 10.1016/j.gde.2015.08.007.
51. Statland, J.M., and Tawil, R. (2016). Facioscapulohumeral Muscular Dystrophy. Continuum (Minneap Minn) *22*, 1916-1931. 10.1212/CON.0000000000000399.
52. Rickard, A.M., Petek, L.M., and Miller, D.G. (2015). Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. Hum Mol Genet *24*, 5901-5914. 10.1093/hmg/ddv315.
53. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehar, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603-607. 10.1038/nature11003.
54. Choi, S.H., Gearhart, M.D., Cui, Z., Bosnakovski, D., Kim, M., Schennum, N., and Kyba, M. (2016). DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. Nucleic Acids Res *44*, 5161-5173. 10.1093/nar/gkw141.
55. Yasuda, T., Tsuzuki, S., Kawazu, M., Hayakawa, F., Kojima, S., Ueno, T., Imoto, N., Kohsaka, S., Kunita, A., Doi, K., et al. (2016). Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. Nat Genet *48*, 569-574. 10.1038/ng.3535.

56. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol *19*, 15. 10.1186/s13059-017-1382-0.
57. Jagannathan, S., Shadle, S.C., Resnick, R., Snider, L., Tawil, R.N., van der Maarel, S.M., Bradley, R.K., and Tapscott, S.J. (2016). Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. Hum Mol Genet *25*, 4419-4431. 10.1093/hmg/ddw271.
58. Grow, E.J., Weaver, B.D., Smith, C.M., Guo, J., Stein, P., Shadle, S.C., Hendrickson, P.G., Johnson, N.E., Butterfield, R.J., Menafra, R., et al. (2021). p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. Nat Genet *53*, 1207-1220. 10.1038/s41588-021-00893-0.
59. Ghandi, M., Huang, F.W., Jane-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., 3rd, Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature *569*, 503-508. 10.1038/s41586-019-1186-3.
60. Freed-Pastor, W.A., and Prives, C. (2012). Mutant p53: one name, many proteins. Genes Dev *26*, 1268-1286. 10.1101/gad.190678.112.
61. Willis, A., Jung, E.J., Wakefield, T., and Chen, X. (2004). Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. Oncogene *23*, 2330-2338. 10.1038/sj.onc.1207396.
62. Boettcher, S., Miller, P.G., Sharma, R., McConkey, M., Leventhal, M., Krivtsov, A.V., Giacomelli, A.O., Wong, W., Kim, J., Chao, S., et al. (2019). A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. Science *365*, 599-604. 10.1126/science.aax3649.
63. Young, J.M., Whiddon, J.L., Yao, Z., Kasinathan, B., Snider, L., Geng, L.N., Balog, J., Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2013). DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. PLoS Genet *9*, e1003947. 10.1371/journal.pgen.1003947.
64. Yoshihara, M., Kirjanov, I., Nykanen, S., Sokka, J., Weltner, J., Lundin, K., Gawriyski, L., Jouhilahti, E.M., Varjosalo, M., Tervaniemi, M.H., et al. (2022). Transient DUX4 expression in human embryonic stem cells induces blastomere-like expression program that is marked by SLC34A2. Stem Cell Reports *17*, 1743-1756. 10.1016/j.stemcr.2022.06.002.
65. Eckersley-Maslin, M., Alda-Catalinas, C., Blotenburg, M., Kreibich, E., Krueger, C., and Reik, W. (2019). Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. Genes Dev *33*, 194-208. 10.1101/gad.321174.118.
66. Dion, C., Roche, S., Laberthonniere, C., Broucqsault, N., Mariot, V., Xue, S., Gurzau, A.D., Nowak, A., Gordon, C.T., Gaillard, M.C., et al. (2019). SMCHD1 is involved in de novo methylation of the DUX4-encoding D4Z4 macrosatellite. Nucleic Acids Res *47*, 2822-2839. 10.1093/nar/gkz005.
67. Ruebel, M.L., Vincent, K.A., Schall, P.Z., Wang, K., and Latham, K.E. (2019). SMCHD1 terminates the first embryonic genome activation event in mouse two-cell embryos and contributes to a transcriptionally repressive state. Am J Physiol Cell Physiol *317*, C655-C664. 10.1152/ajpcell.00116.2019.

68. Huang, Z., Yu, J., Cui, W., Johnson, B.K., Kim, K., and Pfeifer, G.P. (2021). The chromosomal protein SMCHD1 regulates DNA methylation and the 2c-like state of embryonic stem cells by antagonizing TET proteins. Sci Adv *7*. 10.1126/sciadv.abb9149.
69. Campbell, A.E., Belleville, A.E., Resnick, R., Shadle, S.C., and Tapscott, S.J. (2018). Facioscapulohumeral dystrophy: activating an early embryonic transcriptional program in human skeletal muscle. Hum Mol Genet *27*, R153-R162. 10.1093/hmg/ddy162.
70. Tihaya, M.S., Mul, K., Balog, J., de Greef, J.C., Tapscott, S.J., Tawil, R., Statland, J.M., and van der Maarel, S.M. (2023). Facioscapulohumeral muscular dystrophy: the road to targeted therapies. Nat Rev Neurol *19*, 91-108. 10.1038/s41582-022-00762-2.
71. Oliva, J., Galasinski, S., Richey, A., Campbell, A.E., Meyers, M.J., Modi, N., Zhong, J.W., Tawil, R., Tapscott, S.J., and Sverdrup, F.M. (2019). Clinically Advanced p38 Inhibitors Suppress DUX4 Expression in Cellular and Animal Models of Facioscapulohumeral Muscular Dystrophy. J Pharmacol Exp Ther *370*, 219-230. 10.1124/jpet.119.259663.
72. Cruz, J.M., Hupper, N., Wilson, L.S., Concannon, J.B., Wang, Y., Oberhauser, B., Patora-Komisarska, K., Zhang, Y., Glass, D.J., Trendelenburg, A.U., and Clarke, B.A. (2018). Protein kinase A activation inhibits DUX4 gene expression in myotubes from patients with facioscapulohumeral muscular dystrophy. J Biol Chem *293*, 11837-11849. 10.1074/jbc.RA118.002633.
73. Hogan, B., Fellous, M., Avner, P., and Jacob, F. (1977). Isolation of a human teratoma cell line which expresses F9 antigen. Nature *270*, 515-518. 10.1038/270515a0.
74. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564-576 e516. 10.1016/j.cell.2017.06.010.
75. Himeda, C.L., Jones, T.I., Virbasius, C.M., Zhu, L.J., Green, M.R., and Jones, P.L. (2018). Identification of Epigenetic Regulators of DUX4-fl for Targeted Therapy of Facioscapulohumeral Muscular Dystrophy. Mol Ther *26*, 1797-1807. 10.1016/j.ymthe.2018.04.019.
76. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674. 10.1016/j.cell.2011.02.013.
77. Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70. 10.1016/s0092-8674(00)81683-9.
78. van Overveld, P.G., Lemmers, R.J., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat Genet *35*, 315-317. 10.1038/ng1262.
79. Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun *8*, 14049. 10.1038/ncomms14049.
80. Levine, J.H., Simonds, E.F., Bendall, S.C., Davis, K.L., Amir el, A.D., Tadmor, M.D., Litvin, O., Fienberg, H.G., Jager, A., Zunder, E.R., et al. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell *162*, 184-197. 10.1016/j.cell.2015.05.047.

81. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587 e3529. 10.1016/j.cell.2021.04.048.
82. P, H. (2022). SeuratDisk: Interfaces for HDF5-Based Single Cell File Formats. https://mojaveazure.github.io/seurat-disk/ , https://github.com/mojaveazure/seurat-disk.
83. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128. 10.1186/1471-2105-14-128.
84. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res *44*, W90-97. 10.1093/nar/gkw377.
85. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90. 10.1002/cpz1.90.
86. Hu, C., Li, T., Xu, Y., Zhang, X., Li, F., Bai, J., Chen, J., Jiang, W., Yang, K., Ou, Q., et al. (2022). CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. Nucleic Acids Res. 10.1093/nar/gkac947.
87. Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis. https://ggplot2.tidyverse.org.
88. Hesterberg, T. (2022). Resample: Resampling Functions. https://CRAN.R-project.org/package=resample.
89. Lemmers, R.J., Goeman, J.J., van der Vliet, P.J., van Nieuwenhuizen, M.P., Balog, J., Vos-Versteeg, M., Camano, P., Ramos Arroyo, M.A., Jerico, I., Rogers, M.T., et al. (2015). Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. Hum Mol Genet *24*, 659-669. 10.1093/hmg/ddu486.
90. Tapscott, S.J., Lassar, A.B., Davis, R.L., and Weintraub, H. (1989). 5-bromo-2'-deoxyuridine blocks myogenesis by extinguishing expression of MyoD1. Science *245*, 532-536. 10.1126/science.2547249.
91. Resnick, R., Wong, C.J., Hamm, D.C., Bennett, S.R., Skene, P.J., Hake, S.B., Henikoff, S., van der Maarel, S.M., and Tapscott, S.J. (2019). DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. Cell Rep *29*, 1812-1820 e1815. 10.1016/j.celrep.2019.10.025.
92. Campbell, A.E., Shadle, S.C., Jagannathan, S., Lim, J.W., Resnick, R., Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2018). NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. Elife *7*. 10.7554/eLife.31023.
93. Zhou, W.Y., Cai, Z.R., Liu, J., Wang, D.S., Ju, H.Q., and Xu, R.H. (2020). Circular RNA: metabolism, functions and interactions with proteins. Mol Cancer *19*, 172. 10.1186/s12943-020-01286-3.
94. Wilusz, J.E. (2015). Repetitive elements regulate circular RNA biogenesis. Mob Genet Elements *5*, 1-7. 10.1080/2159256X.2015.1045682.

95. Zhang, B., Gunawardane, L., Niazi, F., Jahanbani, F., Chen, X., and Valadkhan, S. (2014). A novel RNA motif mediates the strict nuclear localization of a long noncoding RNA. Mol Cell Biol *34*, 2318-2329. 10.1128/MCB.01673-13.
96. Amini Chermahini, G., Rashnonejad, A., and Harper, S.Q. (2019). RNAscope in situ hybridization-based method for detecting DUX4 RNA expression in vitro. RNA *25*, 1211-1217. 10.1261/rna.070177.118.

Appendix 1: *DUX4* sequences

A. *DUX4*^{WT}:

ATGGCCCTCCCGACACCCTCGGACAGCACCCTCCCCGCGGAAGCCCGGGGACGAGGACGGCGACGGAGAC
TCGTTTGGACCCCGAGCCAAAGCGAGGCCCTGCGAGCCTGCTTTGAGCGGAACCCGTACCCGGGCATCGC
CACCAGAGAACGGCTGGCCCAGGCCATCGGCATTCCGGAGCCCAGGGTCCAGATTTGGTTTCAGAATGAG
AGGTCACGCCAGCTGAGGCAGCACCGGCGGGAATCTCGGCCCTGGCCCGGGAGACGCGGCCCGCCAGAAG
GCCGGCGAAAGCGGACCGCCGTCACCGGATCCCAGACCGCCCTGCTCCTCCGAGCCTTTGAGAAGGATCG
CTTTCCAGGCATCGCCGCCCGGGAGGAGCTGGCCAGAGAGACGGGCCTCCCGGAGTCCAGGATTCAGATC
TGGTTTCAGAATCGAAGGGCCAGGCACCCGGGACAGGGTGGCAGGGCGCCCGCGCAGGCAGGCGGCCTGT
GCAGCGCGGCCCCCGGCGGGGGTCACCCTGCTCCCTCGTGGGTCGCCTTCGCCCACACCGGCGCGTGGGG
AACGGGGCTTCCCGCACCCCACGTGCCCTGCGCGCCTGGGGCTCTCCCACAGGGGGCTTTCGTGAGCCAG
GCAGCGAGGGCCGCCCCCGCGCTGCAGCCCAGCCAGGCCGCGCCGGCAGAGGGGATCTCCCAACCTGCCC
CGGCGCGCGGGGATTTCGCCTACGCCGCCCCGGCTCCTCCGGACGGGGCGCTCTCCCACCCTCAGGCTCC
TCGCTGGCCTCCGCACCCGGGCAAAAGCCGGGAGGACCGGGACCCGCAGCGCGACGGCCTGCCGGGCCCC
TGCGCGGTGGCACAGCCTGGGCCCGCTCAAGCGGGGCCGCAGGGCCAAGGGGTGCTTGCGCCACCCACGT
CCCAGGGGAGTCCGTGGTGGGGCTGGGGCCGGGGTCCCCAGGTCGCCGGGGCGGCGTGGGAACCCCAAGC
CGGGGCAGCTCCACCTCCCCAGCCCGCGCCCCCGGACGCCTCCGCCTCCGCGCGGCAGGGGCAGATGCAA
GGCATCCCGGCGCCCTCCCAGGCGCTCCAGGAGCCGGCGCCCTGGTCTGCACTCCCCTGCGGCCTGCTGC
TGGATGAGCTCCTGGCGAGCCCGGAGTTTCTGCAGCAGGCGCAACCTCTCCTAGAAACGGAGGCCCCGGG
GGAGCTGGAGGCCTCGGAAGAGGCCGCCTCGCTGGAAGCACCCCTCAGCGAGGAAGAATACCGGGCTCTG
CTGGAGGAGCTTTAG

B. *DUX4*^{CA}

ATGGCATTGCCTACACCTTCAGACTCTACGCTGCCTGCAGAGGCTAGGGGAAGAGGTAGACGGCGGCGAT
TGGTGTGGACTCCATCACAATCCGAAGCTCTTCGCGCATGCTTCGAGCGCAATCCCTATCCGGGGATTGC
CACAAGGGAGAGGCTTGCACAGGCTATCGGAATCCCGGAACCGAGAGTGCAGATCTGGTTCCAAAATGAA
CGCTCTCGGCAGCTCAGACAGCATCGCAGGGAGTCCCGCCCGTGGCCAGGAAGAAGGGGACCACCTGAAG
GAAGAAGAAAACGCACAGCGGTGACTGGCAGCCAAACGGCTCTGCTGCTCCGCGCTTTCGAGAAAGATCG
GTTCCCCGGAATTGCCGCACGCGAAGAACTCGCCAGAGAAACTGGGCTCCCAGAATCACGAATACAGATT
TGGTTCCAGAACCGCAGAGCAAGACACCCAGGCCAGGGGGGACGGGCACCTGCTCAGGCCGGTGGACTCT
GCTCTGCTGCCCCTGGGGGCGGCCATCCAGCACCTTCCTGGGTGGCTTTCGCTCATACTGGCGCTTGGGG
TACCGGGCTGCCTGCTCCGCATGTTCCCTGTGCTCCAGGGGCCCTCCCGCAGGGAGCGTTTGTTTCCCAG
GCAGCTAGGGCTGCACCTGCCCTGCAACCATCACAGGCAGCGCCAGCTGAAGGCATCAGCCAACCCGCCC
CAGCCCGCGGAGATTTTGCTTATGCAGCGCCAGCACCTCCAGACGGTGCCCTGAGCCACCCCCAAGCCCC
CAGATGGCCCCCTCACCCTGGTAAGTCCCGGGAAGACCGCGATCCCCAACGAGATGGACTGCCCGGTCCT
TGCGCTGTGGCCCAGCCAGGACCTGCTCAAGCCGGCCCTCAGGGGCAAGGAGTGCTGGCCCCACCTACAA
GCCAGGGATCTCCCTGGTGGGGTTGGGGACGCGGACCTCAGGTTGCTGGAGCCGCTTGGGAGCCTCAGGC
CGGAGCTGCACCGCCGCCACAACCGGCCCCTCCCGACGCGTCAGCGTCCGCCCGACAAGGCCAGATGCAG
GGAATCCCAGCACCTAGCCAAGCTCTTCAAGAGCCTGCCCCTTGGAGCGCACTGCCGTGTGGGCTGCTCC
TGGATGAACTCCTGGCTAGCCCAGAATTTCTCCAGCAGGCACAGCCACTCCTGGAAACAGAAGCTCCGGG
AGAGCTCGAAGCCTCCGAAGAAGCAGCAAGCCTGGAGGCACCTCTTTCCGAGGAGGAGTATAGAGCCCTT
CTGGAAGAACTTTGA

C. *DUX4*$^{CAshort\ CTD}$

GCGTCAGCGTCCGCCCGACAAGGCCAGATGCAGGGAATCCCAGCACCTAGCCAAGCTCTTCAAGAGCCTG
CCCCTTGGAGCGCACTGCCGTGTGGGCTGCTCCTGGATGAACTCCTGGCTAGCCCAGAATTTCTCCAGCA
GGCACAGCCACTCCTGGAAACAGAAGCTCCGGGAGAGCTCGAAGCCTCCGAAGAAGCAGCAAGCCTGGAG
GCACCTCTTTCCGAGGAGGAGTATAGAGCCCTTCTGGAAGAACTTTGA

D. *DUX4*^{CAnoNR}

ATGGCATTGCCTACACCTTCAGACTCTACGCTGCCTGCAGAGGCTAGGGGAAGAGGTAGACGGCGGCGAT
TGGTGTGGACTCCATCACAATCCGAAGCTCTTCGCGCATGCTTCGAGCGCAATCCCTATCCGGGGATTGC
CACAAGGGAGAGGCTTGCACAGGCTATCGGAATCCCGGAACCGAGAGTGCAGATCTGGTTCCAAAATGAA
CGCTCTCGGCAGCTCAGACAGCATCGCAGGGAGTCCCGCCCGTGGCCAGGAAGAAGGGGACCACCTGAAG
GAAGAAGAAAACGCACAGCGGTGACTGGCAGCCAAACGGCTCTGCTGCTCCGCGCTTTCGAGAAAGATCG
GTTCCCCGGAATTGCCGCACGCGAAGAACTCGCCAGAGAAACTGGGCTCCCAGAATCACGAATACAGATT
TGGTTCCAGAACCGCAGAGCAAGACACCCAGGCCAGGGGGGACGGGCACCTGCTCAGGCCGGTGGACTCT
GCTCTGCTGCCCCTGGGGGCGGCCATCCAGCACCTTCCTGGGTGGCTTTCGCTCATACTGGCGCTTGGGG
TACCGGGCTGCCTGCTCCGCATGTTCCCTGTGCTCCAGGGGCCCTCCCGCAGGGAGCGTTTGTTTCCCAG
GCAGCTAGGGCTGCACCTGCCCTGCAACCATCACAGGCAGCGCCAGCTGAAGGCATCAGCCAACCCGCTC
CTGCTCGCGGAGATTTTGCTTATGCAGCGCCAGCACCTCCAGACGGTGCCCTGAGCCACCCTCAGGCTCC
CAGATGGCCCCCTCACCCTGGTAAGTCCCGGGAAGACCGCGATCCCCAACGAGATGGACTGCCCGGTCCT
TGCGCTGTGGCCCAGCCAGGACCTGCTCAAGCCGGCCCTCAGGGGCAAGGAGTGCTGGCCCCACCTACAA
GCCAGGGATCTCCCTGGTGGGGTTGGGGACGCGGACCTCAGGTTGCTGGAGCCGCTTGGGAGCCTCAGGC
CGGAGCTGCACCGCCGCCACAACCGGCCCCTCCCGACGCGTCAGCGTCCGCCCGACAAGGCCAGATGCAG
GGAATCCCAGCACCTAGCCAAGCTCTTCAAGAGCCTGCCCCTTGGAGCGCACTGCCGTGTGGGCTGCTCC
TGGATGAACTCCTGGCTAGTCCAGAATTTCTCCAGCAGGCACAGCCACTCCTGGAAACAGAAGCTCCGGG
AGAGCTCGAAGCCTCCGAAGAAGCAGCAAGCCTGGAGGCACCTCTTTCCGAGGAGGAGTATAGGGCTCTT
CTGGAAGAACTTTGA