

©Copyright 2025

Ellen Paatela

Before and after DUX4: deconstructing its web of silencers and defining its long-term
impact on cellular processes

Ellen Paatela

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Stephen J Tapscott, Chair

Toshio Tsukiyama

Sita Kugel

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Before and After DUX4: Deconstructing its Web of Silencers and Defining its Long-Term
Impact on Cellular Processes

Ellen Paatela

Chair of the Supervisory Committee:

Stephen J Tapscott

Department of Neurology

Facioscapulohumeral dystrophy (FSHD) is driven by a loss of epigenetic repression at the D4Z4 repeat region, leading to aberrant expression of early embryonic transcription factor Double Homeobox 4 (DUX4) and an embryonic transcriptional program in skeletal muscle. Mechanisms that initiate and maintain epigenetic repression at the D4Z4 are complex and incompletely understood. Here, I designed a functional silencing reporter system to identify specific D4Z4 sequences that drive silencing activity and discovered that one discrete sequence, termed D4Z4-S5, was sufficient for sequence-dependent silencer recruitment. This sequence drives silencing activity through both *de novo* DNA methylation and repressive histone modifications, driven specifically by several previously identified D4Z4 silencers: SETDB1, ATF7IP,

SIN3A/3B, and LRIF1. The reporter system also showed promise as a robust platform for further discovery of FSHD therapeutics and disease modifiers, with candidate therapeutic p38 pathway inhibitors showing increased repression in D4Z4-S5. Our findings identify a key D4Z4 regulatory sequence that drives epigenetic repression and proposes a novel system for further screening of FSHD disease modifiers and therapeutics.

The second focus of this work is based on the long-term effects of DUX4 in cancer. Expression of DUX4 has been identified in a wide variety of solid tissue cancers. DUX4 was recently found to be a driver of immune evasion and is associated with poor prognosis in response to immune-mediated cancer therapies due to a loss of antigen presentation machinery. Though the transcriptional activity of DUX4 has been well characterized, recent studies have implicated DUX4 in post-transcriptional modulation of cellular processes. Here, I describe the mechanism of DUX4-mediated translational suppression, leading to translational reprogramming. I also describe optimization attempts to develop a DUX4-lineage tracing system for further study of long-term effects of DUX4 activity in cancer and FSHD.

Table of Contents

<i>Table of Contents</i>	5
<i>Acknowledgements</i>	11
<i>Chapter 1: Introduction</i>	12
<i>DUX4</i> expression in the early embryo drives the first wave of zygotic genome activation	12
<i>DUX4</i> misexpression in skeletal muscle drives pathogenesis of facioscapulohumeral dystrophy	13
Epigenetic regulation of <i>DUX4</i> expression in FSHD pathogenesis	14
Current therapeutic strategies for targeting <i>DUX4</i> in FSHD	15
Post-transcriptional effects of <i>DUX4</i> expression	17
<i>DUX4</i> expression in cancer leads to immune evasion and induction of a stem-cell like program	17
<i>Chapter 2: A discrete region of the D4Z4 is sufficient to initiate epigenetic silencing</i>	19
Introduction	19
Results	20
Specific regions in the LRIF1 promoter and the D4Z4 suppress expression of an integrated reporter	20
Epigenetic repression of D4Z4-S5 contributes to silencing activity	22
Histone methyltransferase and de-acetylase activity contributes to silencing at D4Z4-S5 and LRIF1-FLr	23
SMCHD1 and LRIF1 are enriched at D4Z4-S5 and LRIF1-FLr yet display different silencing roles	24
D4Z4-S5 and LRIF1-FLr cell lines are sensitive to p38 pathway inhibition and can be used to screen for novel FSHD therapeutics	25
Discussion	26

Materials and Methods	28
Figures	33
Figure 2.1: A safe harbor system identifies silencing activity driven by specific sequences in the LRIF1 promoter.	33
Figure 2.2: Segmentation of the D4Z4 reveals one sequence with silencing activity.	34
Figure 2.3: GFP silencing activity of D4Z4-S5 is partially due to de novo DNA methylation and repressive histone modifications.	36
Figure 2.4: Loss of histone methyltransferases and deacetylases specifically relieves D4Z4-S5 and LRIF1-FLr GFP silencing.	38
Figure 2.5: SMCHD1 and LRIF1 are enriched at D4Z4-S5 and LRIF1-FLr yet display different silencing roles.	40
Figure 2.6: Candidate p38 inhibitor FSHD therapeutics inhibit GFP expression specifically in D4Z4-S5 and LRIF1-FLr.	41
Supplementary Figures	43
Supplementary Figure S2.1: PCR genotyping reveals proper integration of LRIF1 constructs, and silencing activity of LRIF1-S3, but not LRIF1-FL, is independent of sequence orientation.	43
Supplementary Figure S2.2: D4Z4-S5 silencing activity is orientation-dependent and increases over time in culture.	44
Supplementary Figure S2.3: LRIF1-S3 contains a conserved sequence that conveys silencing activity.	46
Supplementary Figure S2.4: Further segmentation of D4Z4-S5.4 does not narrow region that confers silencing activity.	47
Supplemental Figure S2.5: Treatment of D4Z4-S5 or LRIF1-FLr cells with 5aza-dC or Entinostat initially increases GFP expression, followed by GFP re-silencing over time.	48
Supplementary Figure S2.6: Knockdown of MBD1, CHD4, or TRIM28 does not rescue GFP expression in D4Z4-S5 and LRIF1-FLr cells.	50
Supplementary Figure S2.7: RT-qPCR analysis confirms GFP expression changes and sufficient knockdown efficiency of identified epigenetic regulators.	52
Supplementary Figure S2.8: RT-qPCR and western blot analyses confirm changes in GFP expression and sufficient knockdown of SMCHD1 and LRIF1.	54
Supplementary Figure S2.9: D4Z4-S5 and LRIF1-FLr show differential sensitivity to candidate FSHD therapeutics.	56
Tables	58
Table 2.1: Characteristics of AAVS1 construct insert sequences.	58

Chapter 3: Using D4Z4-S5 and LRIF1-FLr reporter constructs for unbiased discovery of novel FSHD regulators _____ 59

Introduction _____ 59

Results _____ 59

enChIP-MS as an unbiased discovery tool to identify direct regulators of D4Z4-S5 and LRIF1-FLr _____ 59

A live cell imaging kinase inhibitor screen confirms p38 inhibitors as regulators of DUX4 expression. _____ 61

Discussion _____ 62

Materials and Methods _____ 63

Figures _____ 68

Figure 3.1: Design and validation of AAVS1-targeting enChIP system. _____ 68

Figure 3.2: enChIP-MS analysis confirms binding of many previously identified D4Z4-regulatory proteins to D4Z4-S5. _____ 70

Figure 3.3: enChIP-MS analysis also confirms binding of many previously identified D4Z4-regulatory proteins to LRIF1-FLr. _____ 72

Figure 3.4: Live-cell imaging kinase inhibitor screen of GFP fluorescence in CTRL, D4Z4-S5, and LRIF1-FL cells. _____ 73

Figure 3.5: Live-cell imaging GFP confluency analysis of candidate DUX4 inhibitors in AAVS1 reporter lines reflects previous flow cytometry results. _____ 74

Figure 3.6: Kinase Inhibitor Regularization (KiR) analysis of AAVS1 reporter kinase inhibitor screen. _____ 75

Tables _____ 77

Table 3.1: Kinase inhibitors used in live cell imaging screen. _____ 77

Chapter 4: The transcription factor DUX4 orchestrates translational reprogramming by broadly suppressing translation efficiency and promoting expression of DUX4-induced mRNAs _____ 78

Introduction _____ 78

Results _____ 80

DUX4 activity induces prolonged suppression of antigen presentation factors ____ 80

DUX4 modulates multiple pathways involved in translational regulation _____ 81

Transient DUX4 activity broadly suppresses nascent protein synthesis _____	83
Ribosome footprinting reveals a DUX4-induced loss of 5' ribosome occupancy _	84
Genome-wide polysome profiling identifies defects in translation initiation and elongation _____	85
Translation of DUX4-induced mRNAs _____	86
Expression of endogenous DUX4 correlates with translational suppression and reduced MHC-I surface antigens in cancer cells _____	87
Discussion _____	87
Materials and Methods _____	89
Figures _____	96
Figure 4.1: Brief expression of DUX4 results in long-term suppression of MHC-I and iProteasome subunits. _____	96
Figure 4.2: A pulse of DUX4 negatively regulates the status of multiple translational regulators and broadly suppresses nascent protein synthesis. _____	97
Figure 4.3: Ribosome footprinting and polysome profiling reveal DUX4 suppressive effects on translation efficiency. _____	99
Figure 4.4: DUX4 orchestrates translational reprogramming through broad inhibition of translation concurrent with de novo translation of DUX4 target mRNAs. _____	101
Supplementary Figures _____	103
Supplementary Figure S4.1: Subcellular mRNA localization via cell fractionation. _____	103
Supplementary Figure S4.2: Treatment with MG132 or Bafilomycin does not rescue DUX4-induced protein suppression of MHC-I and the iProteasome. _____	105
Supplementary Figure S4.3: MHC-I protein stability does not require iProteasome catalytic subunits PSMB8 or PSMB9. _____	107
Supplementary Figure S4.4: Validation of IFN γ -induced mRNA expression levels with Everolimus, Torin2, and 4EGI-1 treatments. _____	109
Supplementary Figure S4.5: Oxidative stress, DNA damage, and hypoxia-induced cell stress pathways moderately suppress MHC-I and PSMB9. _____	111
Supplementary Figure S4.6: Quality control analysis of Ribo-seq data _____	112
Supplementary Figure S4.7: Genome-wide quantification of ribosome-protected mRNA fragments using RNA sequencing (RNA-seq) and ribosome profiling (Ribo-seq). _____	113
Supplementary Figure S4.8: Polysome profiling and 5' UTR analysis. _____	114

Chapter 5: Designing a lineage tracing system to track DUX4-expressing cells and their progeny	116
Introduction	116
Results	117
Design of a classical tamoxifen-inducible Cre recombinase-based system to track DUX4-expressing cells and their progeny.	117
Treatment with doxorubicin or etoposide increases DUX4 expression and the number of GFP+ cells in the polyclonal line, but not in monoclonal lines	119
A doxycycline-inducible Cre recombinase lineage tracing system to tightly control Cre expression and stoplight excision.	120
Monoclonal selection was ineffective in producing constitutive control lines but revealed marginal success in DUX4 lineage tracing lines.	121
Antisense oligonucleotide (ASO)-mediated knockdown of DUX4 partially decreases Cre expression in G401 H3Y1-rtTA lineage tracing monoclonal lines.	122
Design and pilot trial of endogenous DUX4 target gene knock-in of lineage tracing machinery.	123
Discussion	123
Materials and Methods	125
Figures	129
Figure 5.1: Pilot testing of the 4'OH-Tamoxifen-inducible Cre system shows some promise in temporal control of lineage tracing.	129
Figure 5.2: Increasing expression of DUX4 with doxorubicin or etoposide treatment does not improve temporal efficiency of ERT2CreERT2-based lineage tracing system.	131
Figure 5.3: Doxycycline-inducible Cre recombinase system increases efficacy of Stoplight construct Cre excision.	133
Figure 5.4: G401 PGK-rtTA monoclonal lines are not responsive to DOX induction.	135
Figure 5.5: Enhancing DUX4 expression with doxorubicin shows robust response in G401 lineage tracing monoclonal lines treated with doxycycline, but monoclonal expression of GFP is still leaky.	136
Figure 5.6: Enhancing DUX4 expression with "Catch and Release" shows robust response in G401 lineage tracing monoclonal lines treated with doxycycline, but monoclonal expression of GFP is still leaky.	138
Figure 5.7: Stoplight excision and GFP expression in G401 monoclonal lines is partially DUX4-dependent, though leaky GFP expression is still prevalent.	140

Figure 5.8: Myoblast polyclonal lineage tracing lines with rtTA knocked into DUX4 target loci are partially responsive to etoposide treatment. _____ 142

Chapter 6: Discussion _____ **144**

Epigenetic regulation of the D4Z4 and implications for FSHD therapeutic screens _____ 144

Long term effects of DUX4 expression in development and disease _____ 147

Chapter 7: References _____ **148**

Acknowledgements

I would like to take this opportunity to thank several people who have made this accomplishment possible.

First, to my advisor Dr. Stephen Tapscott, thank you for your guidance and mentorship throughout my PhD. Your extensive expertise in biology, research processes, and mentorship has made this work possible. I would also like to thank my committee, Drs. Adam Geballe, Andrew Hsieh, Sita Kugel, Toshio Tsukiyama, and Michael Gale, for their support over the years. Their input and advice into my project over the years has helped shape my work into what it has become. Finally, I would like to thank members of the Tapscott lab, both past and present, who have provided an immeasurable amount of insight and support into my projects over the years: Dr. Danielle Hamm, Dr. Tessa Arends, Dr. Andrew Smith, Dr. Amy Spens, Dr. Chao-Jen Wong, Sean Bennett, Aidan O'Donnell, Esteban Abeyta, Faith St. Amant, Yee Nip, Nicholas Sutliff, and Colton Robbins. This supportive community of brilliant scientists has shaped how I think and conduct research and has led to this body of work.

In addition to my scientific support, I could not have completed my PhD without a wonderful community of people outside the lab. First, I would like to thank my parents Leanna and Jeremy, and my siblings Lauren and Graham for always encouraging me to follow my passions throughout life. They, along with my extended family, have been instrumental in supporting my pursuit of this degree. To my partner Nick, you have been the rock I needed to keep me going through this challenging process. To all my friends, both old and new, your support means the world to me. I look forward to the next chapter of life with you all.

Chapter 1: Introduction

DUX4 expression in the early embryo drives the first wave of zygotic genome activation

In the earliest stages of placental mammalian development post-fertilization, cellular processes are driven by maternally deposited RNAs and proteins present in the oocyte. After these deposited RNAs and proteins are degraded, the cleavage stage (4-8 cell stage in humans and the 2-cell stage in mice) occurs and the embryo undergoes the maternal to zygotic transition (MZT), where the embryonic genome is activated and begins expressing its own genes in a process called zygotic genome activation (ZGA) [1]. Some of the main drivers of ZGA in placental mammals are *DUXC*-family genes, such as *Double Homeobox 4* (*DUX4*) in humans and murine *Dux* (*Dux*) in mouse, which induce transcription of over a hundred genes involved in early embryonic developmental processes [2,3]. This ZGA transcriptional program is characterized by expression of totipotency and retroviral genes, elongation of telomeres, chromocenter formation, and chromatin relaxation, leading to further activation of downstream embryonic transcriptional programs [2,4,5]. Expression of *Dux* in mouse embryonic stem cells (mESCs) or *DUX4* in human ESCs promotes ZGA gene expression and cycling into a totipotent 2C- or 4-8C-like state in culture [2,3,6]. Knockout (KO) of *Dux* in mouse zygotes prevents exit from the 2C-like state and morula and blastocyst formation [3], though the essentiality of *Dux* for ZGA activation is disputed [7,8]. This potent induction of totipotency genes by *DUXC*-family members is tightly regulated and important for embryonic development.

The *DUX4* gene, like other *DUXC* family members, is thought to have arisen via retrotransposition of an ancestral *DUXC* sequence because of the lack of introns and similarity between *DUXC* family homeodomain sequences. Located in the highly repetitive D4Z4 region of the human genome, *DUX4* expression is limited to a short 'pulse' in cleavage stage embryos and low levels in the testes; in other somatic tissues, this region is highly epigenetically silenced [2,3,9]. The *DUX4* gene is located within each 3.3 kilobase (kb) tandem repeat unit of the D4Z4 macrosatellite arrays located in the subtelomeric regions of chromosomes 4q and 10q [10–14]. Each tandem repeat unit in the D4Z4 contains the coding region of *DUX4*, and humans typically have between 1 and 150 copies of the D4Z4 [15]. Regulation of *DUX4* expression from the D4Z4 in embryonic development and somatic cells is an area of active research.

DUX4 misexpression in skeletal muscle drives pathogenesis of facioscapulohumeral dystrophy

Facioscapulohumeral dystrophy (FSHD) is the third most common muscular dystrophy, estimated to effect one in every 8-10,000 individuals [16], and is an autosomal dominant disease characterized by progressive wasting of muscles in the face and upper extremities [17,18]. In the 1990s, the causal gene mutation in FSHD was narrowed to chromosome 4q and eventually to deletions in, but not full loss of [19], the D4Z4 macrosatellite on chromosome 4q35 [14,20–24]. Further study of the genetics of FSHD revealed that a putative gene (now identified as *DUX4*) was located within each 3.3 kb D4Z4 element [25], and that only those with the 4qA variant of chromosome 4q, present in about ~50% of the population, were associated with FSHD diagnosis [26,27]. Later, *DUX4* was detected in FSHD patient samples and was shown to be cytotoxic in myoblasts, implicating *DUX4* expression in the muscle wasting phenotype presented in FSHD [28]. Finally, the molecular mechanism of *DUX4* induction of FSHD was identified as a misexpression of *DUX4* due to hypomethylation of the D4Z4 combined with the presence of a 4qA allele, which contains a poly(A) signal downstream of the D4Z4 repeat region, allowing for expression of *DUX4* from the terminal D4Z4 repeat [29]. The discovery of *DUX4* as the causal gene of FSHD spurred a new wave of research into FSHD biology and led to further research into FSHD pathogenesis.

FSHD is primarily an autosomal dominantly-inherited disease, requiring a permissive 4qA allele containing a poly(A) site for *DUX4* expression at the distal 3' end of the D4Z4 array [26]. Up to 30% of cases are caused by sporadic *de novo* mutations with no family history, however, and many FSHD patient families include asymptomatic carriers, indicating the inheritance, penetrance and severity of the disease are highly variable [30–32]. FSHD is characterized by progressive weakening of the muscles in the face (facio-), shoulders (scapulo-) and upper arms (humeral), and some cases, weakness of muscles in the respiratory system [17,18]. Age at symptom onset is extremely variable between patients, with average onset in early adulthood but many instances of onset in infancy or much later into adulthood [31,33]. FSHD muscle is generally characterized by presence of cellular necrosis, immune cell infiltration, fat infiltration, and intramuscular edema that worsens over time [33].

Current diagnoses define two forms of FSHD with distinct mechanisms of D4Z4 de-repression. FSHD1, which represents ~95% of FSHD cases, is caused by chromatin relaxation due to a loss of D4Z4 repeat copy number from 8-100 to 1-10 copies [23,24]. The other ~5% of cases are classified as FSHD2, characterized by intermediate repeat copy number (8-20 copies) and hypomethylation of the D4Z4 due to mutation of epigenetic regulators, such as Structural Maintenance of Chromosomes flexible Hinge Domain Containing 1 (SMCHD1) (>95% of FSHD2 cases), its binding partner Ligand

Dependent Nuclear Receptor Interacting Factor 1 (LRIF1), or DNA methyltransferase 3 beta (DNMT3B) [34–36]. Because FSHD1 and FSHD2 form a disease continuum based on both repeat number and activity of epigenetic regulators, methods to counter FSHD-causing mutations are focused on silencing *DUX4* expression either by increasing the number of D4Z4 repeats or increasing the activity of silencers.

Epigenetic regulation of *DUX4* expression in FSHD pathogenesis

Extensive studies from our lab and others have revealed several mechanisms that regulate epigenetic silencing of the D4Z4, one of which is DNA methylation. A hallmark of FSHD is DNA hypomethylation at the D4Z4 [37,38], sometimes driven by mutations in DNA methyltransferases, such as DNMT3B [36,39,40]. In addition to Southern blot analysis of D4Z4 repeat array size, D4Z4 DNA methylation status has increasingly been integrated into clinical practice for FSHD diagnosis [41]. Several “hotspots” have been identified as areas of significant DNA hypomethylation in individuals with FSHD compared to controls, indicating methylation gradient changes in FSHD vary across the D4Z4 repeat unit [42–45]. The presence of these specific FSHD-associated differentially methylated regions indicates there may be specific sequences necessary for driving silencing of the D4Z4 region and subsequent *DUX4* expression.

D4Z4 silencing is also driven by several repressive histone modifications. A previous study in our lab identified many of these drivers by precipitation of D4Z4 repeats and unbiased proteomics analysis [46]. Components of the Nucleosome Remodeling and Deacetylase (NuRD) and Chromatin Assembly Factor-1 (CAF-1) complexes were identified as suppressors of *DUX4* expression, likely through conferring histone deacetylase (HDAC) activity and modulation of chromatin accessibility, respectively [46]. Other HDACs, SIN3 Transcription Regulator Family Members A and B (SIN3A and SIN3B), also suppressed *DUX4* expression in myoblasts [46].

Repressive histone H3 lysine 9 trimethylation (H3K9me3) loss at the D4Z4 is also associated with FSHD [47]. Loss of histone lysine methyltransferase (HKMT) SET Domain Bifurcated Histone Lysine Methyltransferase 1 (SETDB1) and Tripartite Motif Containing 28 (TRIM28), which complexes with HKMT Suppressor of Variegation 3-9 Homolog 1 (SUV39H1), led to increased *DUX4* expression in myoblasts [46,48]. Several other chromatin-related D4Z4 repressors have also been identified, including FSHD2 disease modifiers SMCHD1 and LRIF1, SMCHD1 interactor RuvB-like AAA ATPase 1 (RUVBL1), histone H3 lysine 4 demethylase Lysine Demethylase 1A (KDM1A), transcriptional repressor Yin Yang 1 (YY1), endogenous RNA interference (RNAi) pathway members Dicer 1 (DICER1) and Argonaut RISC catalytic component 2 (AGO2), and chromatin binding proteins Heterochromatin Protein 1 gamma (HP1 γ), Cohesin, and CCCTC-binding factor (CTCF) [9,34,35,46,47,49–53].

Amongst all these D4Z4 repressors, FSHD2 driver SMCHD1 has specifically been suggested as an important modifier of FSHD disease severity, as FSHD1 patients with *SMCHD1* mutations present more aggressive disease progression [54]. In addition, restoration of normal levels of wild-type SMCHD1 in cells with *SMCHD1* haploinsufficiency can suppress DUX4 expression [55,56], indicating that increasing levels of SMCHD1 may provide a therapeutic approach for both forms of FSHD. Although its exact mechanistic basis is still largely unclear, SMCHD1 has been extensively studied for its roles in both establishment and maintenance of H3K9me3 and higher order chromatin structure of repetitive regions, such as the X chromosome, throughout development [57–60]. SMCHD1 chromatin immunoprecipitation sequencing (ChIP-seq) analysis has shown large contiguous areas of binding in these repetitive regions, including the D4Z4, suggesting that SMCHD1 is recruited by broad epigenetic silencing signatures, such as repressive histone modifications and CpG methylation, rather than by distinct factor binding sequences. Additionally, this analysis also revealed ~30 isolated single-copy autosomal regions with distinct peaks of both SMCHD1 and LRIF1 binding, including the *LRIF1* promoter. Knockdown of SMCHD1 also increased LRIF1 expression, indicating that SMCHD1 is acting as a repressor at this locus [56]. This distinct binding to single-copy loci could indicate SMCHD1/LRIF1 recruitment to specific sequence motifs rather than by distributed chromatin silencing mechanisms. Determining, mechanistically, how DUX4 is silenced by SMCHD1 and other regulators will provide valuable insight into FSHD pathogenesis, as well as identify targets for therapeutic intervention.

Current therapeutic strategies for targeting DUX4 in FSHD

There is currently no cure for FSHD, so current approaches used in the clinic focus on alleviation of symptoms by non-specific treatments to augment muscle function. Physical therapy, β 2-adrenergic agonists, myostatin inhibitors, and creatine have been administered to combat loss of muscle mass, glucocorticoids have been used to combat inflammation and antioxidants have been administered to combat oxidative stress found in FSHD muscle [61]. As the molecular basis of FSHD pathogenesis has been increasingly elucidated, focus has shifted toward specific approaches to limit DUX4 activity.

The search for successful DUX4-inhibitory drugs has been challenging due to the sporadic and variable nature of disease pathogenesis in patients and the difficulties of detecting low levels of *DUX4* RNA and protein, as *DUX4* is expressed sporadically in ~0.1%-1% of cells in patient myoblast samples [9,62,63]. Many screens for identifying suppressors of DUX4 activity use artificial reporter systems driven by DUX4 target gene promoters in FSHD myoblasts. Several compounds have been identified as inhibitors of *DUX4* target gene expression using these systems, including p38 pathway inhibitors

[64–66], beta-2-adrenergic receptor agonists [67,68], bromodomain and extraterminal (BET) protein inhibitors [67], phosphodiesterase (PDE) inhibitors [68], G-quadruplex stabilizers [69], and Wnt pathway agonists [70]. The p38 inhibitor losmapimod was recently in a Phase III clinical trial (ClinicalTrials.gov Identifier: NCT04264442) for treatment of FSHD, though the trial did not meet its endpoint for suppression of *DUX4* target gene activation. Another FSHD clinical trial with beta-2-adrenergic receptor agonist Clenbuterol is currently underway (ClinicalTrials.gov Identifier: NCT06721299). Despite progress made in the identification of potential FSHD therapeutics, the search for more specific therapeutic avenues is still underway.

An exciting new area of FSHD therapeutic research is in the application of oligonucleotide therapies [63]. Several antisense oligonucleotide (ASO)-mediated approaches have shown promise in reducing *DUX4* transcript levels in FSHD cells and mouse models [63]. In studies with siRNAs and ASOs targeting the *DUX4* poly(A) site, splice sites, or the *DUX4* promoter, *DUX4* transcript levels were reduced in FSHD muscle cells [51,71,72]. Phosphorodiamidate morpholino oligomers (PMOs) targeting the *DUX4* poly(A) signal, locked nucleic acid (LNA) gapmers targeting *DUX4* exon 3, U7-small nuclear RNA (U7-snRNA) antisense transcripts targeting *DUX4* splice sites, microRNAs (miRNAs) targeting *DUX4*, and constrained ethyl gapmer (cEt) and 2'-O methoxyethyl gapmers targeting the *DUX4* coding region inhibited *DUX4* expression in FSHD muscle cells and mouse models [73–82]. Finally, DNA aptamers, which bind transcription factors and reduce their ability to bind and induce target gene expression, have been successfully used to reduce *DUX4* transcriptional activity [83,84]. Current barriers to the adaptability of oligonucleotide therapies into the clinic include off-target effects, inefficient delivery and uptake into cells, cost, and interruption of endogenous RNA pathways [63]. As the field of oligonucleotide therapies advances, we hope some of these highlighted systems will be effective in treating FSHD.

Clustered regularly interspaced short palindromic repeats (CRISPR)-based approaches have also been undertaken in FSHD therapeutic intervention. CRISPR/Cas9 editing of the *DUX4* poly(A) signal to reduce *DUX4* transcript expression has had mixed success [85–87], while correction of and SMCHD1 mutation in FSHD2 cells decreased *DUX4* levels [55]. CRISPR-based interference (CRISPRi) approaches, such as single guide RNA (sgRNA)-mediated D4Z4 targeting of catalytically dead *Streptococcus pyogenes* Cas9 fused to epigenetic repressor proteins (dCas9-repressor), have also been assessed as potential therapeutic approaches to FSHD treatment [47,48,86,88–90] and have shown promise in silencing the D4Z4 to reduce *DUX4* misexpression. With all these promising approaches, the problem of how to deliver these therapeutics to patients has become an area of focus. Systemic delivery to reach every muscle in the body is difficult, with the only current method being adeno-associated virus (AAV)-mediated delivery [63]. This method has been implemented in several clinical trials for other muscular dystrophies, but is limited by high cost, sub-

optimal efficiency of delivery, and the immunogenicity of delivering foreign DNA and proteins to cells [91]. As gene therapy technology progresses, the hope is that these therapies will provide successful therapeutic avenues for the treatment of FSHD.

Post-transcriptional effects of *DUX4* expression

In addition to induction of an early embryonic transcriptional program, *DUX4* expression in cells leads to a complex array of post-transcriptional modifications that can impact cellular fate decisions. A well-studied example of post-transcriptional gene regulation is the effect of *DUX4* on mRNA stabilization due to inhibition of nonsense-mediated decay (NMD) mechanisms. In cells without *DUX4* expression, the presence of NMD machinery will prevent mRNAs with premature termination codons (PTCs) from translation [92]. In the presence of *DUX4*, NMD machinery is inhibited, leading to the translation of many PTC-containing mRNAs and induction of cellular stress, apoptosis, insufficient proteasomal degradation, and protein aggregation [93,94]. In one study, *DUX4* induced post-transcriptional suppression of over 200 proteins despite no change or even upregulation at the RNA level [95]. *DUX4* inhibition of nonsense-mediated decay machinery has also been shown to induce production of nonfunctional truncated proteins, leading to protein loss [96–98]. Our own work, described in **Chapter 4**, highlights an additional method of *DUX4*-induced post-transcriptional regulation; a brief pulse of *DUX4* leads to a long-term reduction of multiple translation initiation and elongation proteins, leading to changes in the translome, or the proteins that are preferentially translated [99]. The culmination of these studies highlights the importance of *DUX4* not just as a transcription factor, but as a driver of long-term changes in cellular phenotypes.

***DUX4* expression in cancer leads to immune evasion and induction of a stem-cell like program**

An emerging focus in *DUX4* biology is the discovery of *DUX4* expression in cancer. A landmark study from our lab and other collaborators identified *DUX4* as one of the top hits in a screen for genes with pan-cancer-specific expression [100]. Though several studies have identified *DUX4* fusion proteins as drivers of cancer pathogenesis [101,102], these cancer cells did not express full length *DUX4*. Our lab found that expression of full length, transcriptionally active *DUX4* is correlated with resistance to immune checkpoint blockade in many solid tissue cancers; patients with high *DUX4* expression had a lower survival rate and poorer response to checkpoint blockade immunotherapies [100,103]. This was found to be partially due to an inhibition in MHC Class I (MHC-I) antigen presentation upon *DUX4* expression [100]. It was later

determined that DUX4 inhibits MHC-I through multiple different mechanisms. The DUX4 protein itself inhibits immune signaling, as DUX4 interacts directly with STAT1 to prevent induction of interferon-stimulated genes [104]. Described in detail in **Chapter 4**, my work with others in the lab has determined that DUX4 has also been implicated in long term suppression of MHC-I proteins through inhibition of nascent protein translation [99]. Ultimately, the immune evasion effect and its prevalence in cancers makes DUX4 an interesting target for cancer therapeutic intervention.

In addition to immune evasion, *DUX4* expression has been implicated in induction of a totipotent-like program in a subset of cancer cells. In classic cell culture conditions, the DUX4 protein and its target H3Y1 are expressed in about 0.1% of many cancer cell lines [105]. This induction of DUX4 leads to the expression of 8C- and ZGA-like early embryonic transcriptional programs in a small subset of cancer cells, in addition to suppression of MHC-I mRNA expression and presentation [105]. Additionally, treatment of cancer cells with common chemotherapeutic p53 inhibitors doxorubicin and etoposide increases the number of DUX4-positive cells in cancer and FSHD lines [105,106], indicating common therapeutic interventions may exacerbate *DUX4* expression and further induce an early embryonic program with immune evasion.

This induction of DUX4 and an early embryonic program in many different cancer lines proposes the interesting hypothesis that DUX4 expression in a rare subset of cancer cells leads to a cancer stem cell-like phenotype. Because DUX4 expression is correlated with decreased translation [99] and translational inhibition is associated with increased cancer cell plasticity [107,108], DUX4 inhibition of translation in a subset of cancer cells could lead to increased viability in response to harsh environments, nutrient deprivation, and therapeutic intervention.

Chapter 2: A discrete region of the D4Z4 is sufficient to initiate epigenetic silencing

A version of this chapter is currently in review at Human Molecular Genetics as:

A discrete region of the D4Z4 is sufficient to initiate epigenetic silencing

Ellen M. Paatela, Faith G. St. Amant, Danielle C. Hamm, Sean R. Bennett, Taranjit S. Gujral, Silvère M. van der Maarel, and Stephen J. Tapscott

A version of this chapter is currently a preprint on Biorxiv as:

A discrete region of the D4Z4 is sufficient to initiate epigenetic silencing

Ellen M. Paatela, Faith G. St. Amant, Danielle C. Hamm, Sean R. Bennett, Taranjit S. Gujral, Silvère M. van der Maarel, Stephen J. Tapscott

bioRxiv 2025.02.19.639175; doi: <https://doi.org/10.1101/2025.02.19.639175>

Introduction

Facioscapulohumeral dystrophy (FSHD) is the third most common muscular dystrophy, estimated to affect one in every 8-10,000 individuals [16], and is characterized by progressive wasting of muscles in the face and upper extremities [17,18]. Current treatments for FSHD focus on alleviation of symptoms, as there is no cure [109]. FSHD is caused by misexpression of the transcription factor double homeobox 4 (*DUX4*) in skeletal muscle [9]. *DUX4* is normally expressed as a short 'burst' in the cleavage stage embryo and at low levels in the testes and is epigenetically silenced in most other somatic tissues [2,3,9]. Loss of this epigenetic repression can lead to *DUX4* misexpression and activation of an embryonic program in skeletal muscle, eventually leading to FSHD pathogenesis [62,110,111].

The *DUX4* gene is located within each 3.3 kilobase (kb) tandem repeat unit of the D4Z4 macrosatellite arrays located in the subtelomeric regions of chromosomes 4q and 10q [10–14]. There are two genetic causes of FSHD [26]: FSHD1, which represents ~95% of FSHD cases, is caused by a loss of D4Z4 repeat copy number from 8-100 to 1-10 copies [23,24], and FSHD2 (~5% of cases) which is due to mutation of epigenetic regulators, such as Structural Maintenance of Chromosomes flexible Hinge Domain Containing 1 (SMCHD1) (>95% of FSHD2 cases), its binding partner Ligand Dependent Nuclear Receptor Interacting Factor 1 (LRIF1), or DNA methyltransferase 3 beta (DNMT3B) [34–36]. Both FSHD1 and FSHD2 result in D4Z4 chromatin relaxation and DNA hypomethylation with the subsequent mis-expression of *DUX4* in skeletal muscle.

While SMCHD1 binding to the *DUX4* locus and the mouse *Dux* locus is implicated in the initial epigenetic silencing of *DUX4* expression during development and establishment of DNA methylation [112,113], many factors have been identified that

maintain epigenetic repression of *DUX4* expression in somatic cells and skeletal muscle [110,114,115]. A previous study in our lab identified many of these drivers by precipitation of D4Z4 repeats and unbiased proteomics analysis [46]. Components of the Nucleosome Remodeling and Deacetylase (NuRD) complex together with SIN3 Transcriptional Regulator Family Members A and B (SIN3A and SIN3B) and Chromatin Assembly Factor-1 (CAF-1) complexes were identified as suppressors of *DUX4* expression [46]. Histone H3 lysine 9 trimethylation (H3K9me3) loss at the D4Z4 is associated with FSHD [47] and loss of histone lysine methyltransferase (HKMT) SET Domain Bifurcated Histone Lysine Methyltransferase 1 (SETDB1) and Tripartite Motif Containing 28 (TRIM28), which complexes with HKMT Suppressor of Variegation 3-9 Homolog 1 (SUV39H1), led to increased *DUX4* expression in myoblasts [46,48]. Several other chromatin-related D4Z4 suppressors have also been identified, including FSHD2 disease modifiers SMCHD1 and LRIF1, SMCHD1 interactor RuvB-like AAA ATPase 1 (RUVBL1), histone H3 lysine 4 demethylase Lysine Demethylase 1A (KDM1A), transcriptional repressor Yin Yang 1 (YY1), endogenous RNA interference (RNAi) pathway members Dicer 1 (DICER1) and Argonaut RISC catalytic component 2 (AGO2), and chromatin binding proteins Heterochromatin Protein 1 gamma (HP1 γ), Cohesin, and CCCTC-binding factor (CTCF) [9,34,35,46,47,49–53]. The large number of regulatory factors at the D4Z4 has made identification of the most critical silencing pathways difficult.

In this study, we designed a silencing reporter system to identify specific sequences in the D4Z4 that are sufficient for epigenetic silencing. Our silencing reporter system identified one discrete sequence in the D4Z4, termed D4Z4-S5, that confers epigenetic silencing activity and shows enrichment for SMCHD1 and LRIF1 binding. A small interfering RNA (siRNA) panel of known D4Z4-regulatory proteins revealed several factors necessary to silence the D4Z4-S5 reporter, including SETDB1, its binding partner Activating Transcription Factor 7 Interacting Protein (ATF7IP), SIN3A/3B, and LRIF1. The D4Z4-S5 reporter was also further suppressed by p38 inhibitors, which are known to suppress *DUX4* expression in FSHD skeletal muscle cells. Our study shows that initiation of epigenetic silencing, as measured by our reporter assay, is not distributed across the full D4Z4 unit but instead limited to one region and is regulated by specific factors. This reporter system also provides a simple and robust tool for identification of key D4Z4 regulators and potential therapeutics.

Results

Specific regions in the LRIF1 promoter and the D4Z4 suppress expression of an integrated reporter

To functionally assess the effect of specific D4Z4 sequences on silencing activity, we generated reporter constructs with the test sequence of interest placed upstream of

a CMV promoter driving mEmerald green fluorescence protein (*GFP*) expression (**Fig. 2.1A**) and used homologous recombination to introduce these reporter constructs into the *AAVS1* (*PPP1R12C*) locus in HeLa cells. The *AAVS1* locus is considered a “safe harbor” due to its low intrinsic silencing and no discernible phenotype upon perturbation [116]. HeLa cells were used to avoid the dynamic regulation of D4Z4 expression during myogenic differentiation [117,118].

Amongst D4Z4 suppressors, SMCHD1 has an established role in initiation and maintenance of D4Z4 epigenetic repression. Previous ChIP analysis established that SMCHD1 and LRIF1 bind a discrete region adjacent to the *LRIF1* promoter and suppress *LRIF1* expression [56,60]. Therefore, as a positive control for silencing activity, we inserted the full-length 617 nucleotide (nt) region containing the SMCHD1 ChIP-seq peak (termed LRIF1-FL) upstream of the CMV promoter into the reporter construct (**Fig. 2.1B**) and introduced this into the HeLa *AAVS1* locus. Because this region is directly upstream of the *LRIF1* transcriptional start site (TSS), we reversed the orientation of the LRIF1-FL sequence (LRIF1-FLr) to avoid changes in GFP expression due to transcription from the LRIF1 promoter. The directionality of integration within the *AAVS1* construct is shown in **Fig. 2.1B**. For comparison, we also introduced a 200 nt region (CTRL, chr11:114,494,095-114,494,295), previously identified as a sequence with no regulatory activity [119], as a control and likely neutral sequence. Cells were expanded under selection as a polyclone and accurate integration was confirmed by polymerase chain reaction (PCR) genotyping (**Suppl. Fig. S2.1A, B**). To ensure low GFP levels were due to silencing and not a stable GFP^{Low} population, GFP^{High} cells were sorted via fluorescence activated cell sorting (FACS) and cultured for two weeks to allow for construct silencing. Flow cytometry showed that nearly 80% of the CTRL cells expressed GFP at relatively high levels, whereas cells with LRIF1-FLr had approximately 50% of cells showing low GFP fluorescence, approaching a bimodal distribution (**Fig. 2.1C**). Reversing the orientation of LRIF1-FL was necessary for silencing, as flow cytometry of the LRIF1-FL forward sequence (LRIF1-FLf) did not induce silencing, indicating that the presence of the *LRIF1* promoter sequence directly upstream of the TSS was likely driving transcriptional activation (**Suppl. Fig. S2.1C-E**). Fragmentation of the LRIF1-FLr into five overlapping segments showed silencing activity above control levels and was restricted to the 112 nt LRIF1-S3r subfragment (**Fig. 2.1C**) centered under the SMCHD1/LRIF1 ChIP peaks [56] (**Fig. 2.1B**), providing a positive control to show that the integrated reporter can be silenced by a region shown to bind SMCHD1/LRIF1. To ensure that directionality was not a specific driver of silencing activity, the reverse complement of LRIF1-S3, LRIF1-S3f, which contains the sequence in the correct orientation relative to the endogenous *LRIF1* TSS, was integrated into the *AAVS1* system (**Suppl. Fig. S2.1C, D**). Regardless of orientation, LRIF1-S3 showed GFP suppression compared to the control sequence, indicating this sequence is truly a silencer regulatory element (**Suppl. Fig. S2.1E**).

To distinguish whether D4Z4 silencing is localized to a specific region of the D4Z4 or is broadly distributed, the 3.3kb D4Z4 repeat unit was segmented into 14 overlapping fragments, cloned into the reporter construct, and integrated at the AAVS1 locus (**Fig. 2.2A**, **Suppl. Fig. S2.2A**). Flow cytometry revealed that over 40% of cells showed substantial GFP suppression with a single 401 nt region, D4Z4 Segment 5 (D4Z4-S5), with minimal silencing activity in other D4Z4 regions (**Fig. 2.2B**). D4Z4-S5 was the only region with a greater percentage of GFP^{Low} cells than the CTRL sequence. Interestingly, this GFP silencing activity was orientation dependent, as reversing the sequence (D4Z4-S5r) did not show increased GFP suppression compared to the CTRL reporter (**Suppl. Fig. S2.2B-D**). To ensure that this bimodal distribution of GFP fluorescence in D4Z4-S5 was due to *de novo* AAVS1 silencing and not due to a stable population of GFP^{Low} cells in the polyclonal reporter line, cells containing the CTRL reporter, or cell lines with D4Z4 reporter constructs were sorted via FACS for GFP^{High} cells and maintained in culture. At 1- and 5-weeks post sort, flow cytometry was performed and showed substantial GFP suppression at week 5 relative to control levels only in D4Z4-S5, confirming that the GFP suppression is not due to expression artifacts in a polyclonal population of cells, but by sequence-dependent silencing over time (**Suppl. Fig. S2.2E, F**). Further fragmentation of D4Z4-S5 (**Fig. 2.2C**) localized the repressive region to a 146 nt fragment (D4Z4-S5.4, **Fig. 2.2C, D**, **Suppl. Fig. S2.2G**), which showed a larger population of GFP^{Low} cells than the control.

GFP silencing regions D4Z4-S5, D4Z4-S5.4, LRIF1-FL, and LRIF1-S3 have similar GC content and percentage of CpG dinucleotides to other regions of the D4Z4 (**Table 2.1**), suggesting that other sequence-specific aspects may account for silencing activity. LRIF1-S3 contains two regions with high conservation in placental mammals predicted by PhyloP [120] (**Suppl. Fig. S2.3A, B**); a minimal subregion (LRIF1-S3f Minimal Sufficiency), containing these motifs is sufficient for GFP silencing, whereas deletion of one of the regions (LRIF1-S3f Δ site#1) appears to reduce GFP silencing (**Suppl. Fig. S2.3C**). These results indicate site#1 is important for GFP silencing activity, though motif analysis did not reveal any obvious regulator candidates in either conserved site. In contrast to LRIF1-S3, PhyloP predicted no conserved motifs among 100 representative vertebrates in D4Z4-S5.4, and fractionation to smaller fragments did not strongly map silencing activity to a specific smaller region (**Suppl. Fig. S2.4A-C**).

Epigenetic repression of D4Z4-S5 contributes to silencing activity

The D4Z4-S5 region partially overlaps with the previously described DR1 and 5P regions of the D4Z4 [42,121], and D4Z4-S5.4 is centered in the 5P region. Previous studies have shown that both regions have relatively high percentage CpG methylation and are specifically hypomethylated in FSHD [42–44,121]. The D4Z4 also shows high levels of repressive histone modifications, where a loss of these modifications is associated with DUX4 expression and FSHD pathogenesis [40,47,56,118]. To

determine the contributions of DNA methylation and repressive histone modifications to GFP silencing in our AAVS1 reporter system, D4Z4-S5 and LRIF1-FLr cells were treated with the DNA demethylating agent 5-aza-2'-deoxycytidine (5-aza-dC) or the histone deacetylase inhibitor entinostat for three days, and GFP fluorescence levels were quantified via flow cytometry at several timepoints after treatment (**Fig. 2.3A**). Treatment with either drug increased the proportion of GFP^{High} cells in the population, indicating a loss of GFP suppressive mechanisms (**Fig. 2.3B, C, Suppl. Fig. S2.5A, B**). Increased time in culture also revealed a re-emergence of the GFP^{Low} population and a restoration of pre-treatment median GFP fluorescence intensity (MFI) levels at days 7 and 10 (**Fig. 2.3B, C, right**). These results suggest the silencing in D4Z4-S5 and LRIF1-FLr is due, in part, to epigenetic mechanisms.

To determine whether GFP suppression is correlated with higher DNA methylation, D4Z4-S5 and LRIF1-FLr cells were sorted into GFP^{High} and GFP^{Low} populations (**Fig. 2.3D**), genomic DNA was bisulfite converted, AAVS1 reporter construct regions were PCR-amplified, and amplicons were sequenced using long-read nanopore technology to determine methylation status. Compiling two PCR amplicons for each, higher methylation in GFP^{Low} cells was distributed throughout most CpG sites in D4Z4-S5 (**Fig. 2.3E, left**) and larger differences in CpG methylation between GFP^{High} and GFP^{Low} populations were concentrated in one region, between 300 and 500nt, of LRIF1-FLr (**Fig. 2.3E, right**). AAVS1 PCR amplicons from both D4Z4-S5 and LRIF1-FLr showed higher levels of overall CpG methylation per read in the GFP^{Low} cells compared to the GFP^{High} cells (**Fig. 2.3F**). In the GFP^{Low} cells, less than half of the reads are completely unmethylated in both D4Z4-S5 amplicons and LRIF1-FLr PCR#2 (in LRIF1-FLr PCR#1, less than 60% of reads are unmethylated), while in the GFP^{High} populations, 60-80% of reads are completely unmethylated in all AAVS1 amplicons, correlating unmethylated AAVS1 DNA with higher GFP fluorescence (**Fig. 2.3F**). The presence of unmethylated reads in the GFP^{Low} cell populations, however, indicates that DNA methylation is only partially responsible for GFP silencing, and suggests that other epigenetic mechanisms also contribute to sequence-dependent GFP suppression in D4Z4-S5 and LRIF1-FLr.

Histone methyltransferase and de-acetylase activity contributes to silencing at D4Z4-S5 and LRIF1-FLr

Previously, we showed that specific histone methyltransferases, de-acetylases, and components of the NuRD complex were necessary to maintain D4Z4 epigenetic repression in human myoblasts [46]. To determine if these known D4Z4 regulators are also responsible for sequence-dependent silencing in D4Z4-S5 and LRIF1-FLr reporter constructs, a focused siRNA-mediated knockdown (KD) panel of these factors [46] was performed in reporter cells. Loss of CHD4 or MBD1 (**Suppl. Fig. S2.6A-D**), components of the NuRD complex, did not strongly de-repress GFP expression in the D4Z4-S5 or

LRIF1-FLr constructs, though loss of MBD1 led to a small but significant increase in the average GFP MFI relative to siCTRL. However, loss of the H3K9 histone methyltransferase SETDB1 robustly de-repressed GFP expression in both D4Z4-S5 and LRIF1-FLr when compared to CTRL cells (**Fig. 2.4A, Suppl. Fig. S2.7A**). The relative GFP MFI increased roughly 4- and 7-fold in D4Z4-S5 and LRIF1-FLr cells, respectively, while decreasing in CTRL cells (**Fig. 2.4A, right**). Loss of ATF7IP, a binding partner of SETDB1 that increases SETDB1 stability and methyltransferase activity [122], significantly increased GFP fluorescence in all reporter cell lines; however, the fold change in relative MFI is an order of magnitude greater for D4Z4-S5 and LRIF1-FLr cells relative to CTRL (**Fig. 2.4B, Suppl. Fig. S2.7B**). ChIP-quantitative PCR (qPCR) analysis revealed that loss of SETDB1 or ATF7IP lead to a significant loss of H3K9me3 specifically at AAVS1 D4Z4-S5 and LRIF1-FLr inserts, whereas the endogenous loci and control regions (a gene desert region (GDR) that has high baseline H3K9me3 and a promoter region of housekeeping gene ribosomal protein L13a (RPL13A) that has low baseline H3K9me3) did not show a change in H3K9me3 levels (**Fig. 2.4C, Suppl. Fig. S2.7C**). Of interest, knockdown of sequence-specific SETDB1/ATF7IP recruitment factor TRIM28 did not de-repress these regions (**Suppl. Fig. S2.6E, F**) suggesting alternative mechanisms are required for recruitment and/or silencing activity of SETDB1 and ATF7IP at D4Z4-S5 and LRIF1-FLr.

In addition to H3K9 methyltransferases, loss of other repressive histone modifying enzymes, such as histone deacetylases SIN3A and SIN3B and H3K4 demethylase KDM1A, were previously found to relieve suppression of DUX4 in myoblasts [46]. Knockdown of SIN3A and SIN3B in our AAVS1 reporter cells showed a 2.5-fold increase in GFP fluorescence in D4Z4-S5 and LRIF1-FLr cells and not in CTRL cells, indicating histone deacetylation may also contribute to silencing (**Fig. 2.4D, Suppl. Fig. S2.7D**). Interestingly, knockdown of KDM1A did not show a substantial increase in GFP fluorescence, with changes in relative GFP MFI less than 1.5-fold (**Fig. 2.4E, Suppl. Fig. S2.7E**), indicating demethylation of histone H3 on lysine 4 (H3K4) is likely not a driver of sequence-specific GFP suppression in D4Z4-S5 and LRIF1-FLr.

SMCHD1 and LRIF1 are enriched at D4Z4-S5 and LRIF1-FLr yet display different silencing roles

Prior studies from our lab and others document SMCHD1/LRIF1 binding to the endogenous region corresponding to LRIF1-FLr [56,59,60]. ChIP-qPCR analysis of our silencing reporter cells confirmed SMCHD1 and LRIF1 binding to the endogenous *LRIF1* promoter region and identified enriched binding at both the LRIF1-FLr and D4Z4-S5 transgene AAVS1 inserts (**Fig. 2.5A**). Of additional interest, ChIP-qPCR also identified enrichment of these factors at the endogenous D4Z4-S5.4 region (**Fig. 2.5A**). Compared to CTRL cells, knockdown of LRIF1 (short and long isoforms) increased reporter expression in both LRIF1-FLr and D4Z4-S5 cells with a significant increase in

GFP MFI (**Fig. 2.5B**, **Suppl. Fig. S2.8A-C**), supporting our ChIP data showing specific LRIF1 binding and silencer recruitment activity. In contrast, knockdown of SMCHD1 shows a significant decrease in reporter expression for all cell lines, indicating a potential positive regulatory role for reporter expression (**Fig. 2.5C**, **Suppl. Fig. S2.8A-C**). Although SMCHD1 can have a role as a transcriptional co-activator in some contexts [123], SMCHD1 is known to negatively regulate *LRIF1* expression [56]. GFP suppression upon loss of SMCHD1 can potentially be explained by increased levels of the silencer LRIF1; however, differences in LRIF1 RNA and protein levels were not significant (**Suppl. Fig. S2.8A-C**). In summation, LRIF1 and SMCHD1 are recruited in a sequence-specific manner to D4Z4-S5 and LRIF1-FLr, though only LRIF1 seems to contribute to silencing activity in our assays.

D4Z4-S5 and LRIF1-FLr cell lines are sensitive to p38 pathway inhibition and can be used to screen for novel FSHD therapeutics

Several studies have identified p38 inhibitors, beta-2-adrenergic receptor agonists, bromodomain and extraterminal (BET) inhibitors, and G-quadruplex stabilizers as repressors of *DUX4* expression, with several of these drugs advancing to clinical trials for treatment of FSHD [61,64,65,67,69,124,125]. To assess the efficacy of our AAVS1 silencing reporter system as a tool for screening FSHD therapeutics, CTRL, LRIF1-FLr, and D4Z4-S5 cells were treated with several candidate drugs at concentrations sufficient to suppress *DUX4* and *DUX4* target genes in FSHD myoblasts. Compared to vehicle dimethyl sulfoxide (DMSO) treatment, p38 inhibitors losmapimod (ClinicalTrials.gov Identifier: NCT04264442) and SB203580 substantially suppressed GFP expression in a dose-dependent manner in D4Z4-S5, and to a lesser degree LRIF1-FLr, cells when compared to CTRL cells (**Fig. 2.6A, B**). Higher doses of losmapimod and SB203580 significantly decreased the median GFP fluorescence intensity and increased the proportion of GFP^{Low} cells in D4Z4-S5, while CTRL and LRIF1-FLr cells showed similar trends but to a lesser degree (**Fig. 2.6C, D**). These results indicate that p38 pathway inhibition may reduce *DUX4* expression through interactions with the D4Z4-S5 sequence.

In contrast, beta-2-adrenergic receptor agonists clenbuterol (ClinicalTrials.gov Identifier: NCT06721299) and formoterol did not change levels of GFP expression in any of the cell lines (**Suppl. Fig. S2.9A-D**). Treatment trends showed increases in GFP MFI levels and decreased proportions of GFP^{Low} cells at higher doses, indicating that these drugs likely do not impact *DUX4* expression through regulation of the D4Z4-S5 sequence. Treatment with BET inhibitor (+)-JQ1 also did not substantially alter levels of GFP fluorescence, though the highest dose had a significant suppressive effect on D4Z4-S5 and LRIF1-FLr (**Suppl. Fig. S2.9E, F**). Surprisingly, treatment with G-quadruplex stabilizer berberine strongly suppressed GFP in a dose dependent manner specifically in D4Z4-S5 cells (**Suppl. Fig. S2.9G, H**), despite the lack of known D4Z4 G-

quadruplex ligands in the D4Z4-S5 sequence [69]. Further study is necessary to determine the mechanism of berberine-mediated transcriptional silencing through D4Z4-S5.

Finally, to assess the efficacy of D4Z4-S5 as a platform for large scale therapeutic screens, GFP fluorescence changes in our AAVS1 reporter cells were assessed in a live-cell imaging system. Cells were treated with 5-aza-dC to ensure strong GFP expression before GFP^{High} cells were sorted via FACS into multi-well plates. These cells were then cultured in the presence of p38 inhibitors losmapimod or doramapimod for multiple days, during which live cell imaging quantified GFP⁺ cell confluency per well every 2 hours (**Fig. 2.6E**). p38-inhibitor treatment continuously suppressed LRIF1-FLr GFP expression over a six-day period, whereas D4Z4-S5 showed transient suppression maximal at 3.5 days with substantial recovery during longer treatment (**Fig. 2.6F**). Together these data indicate that LRIF1-FLr and D4Z4-S5 repressive activity is modulated by p38 signaling and that these AAVS1 reporter cells provide a sensitive platform for live-cell imaging-based assays to identify additional targets and compounds that modulate expression.

Discussion

Although many prior studies have described epigenetic repressors of the D4Z4 repeats in somatic cells, it was unknown whether DNA sequences sufficient to initiate silencing were broadly distributed across the 3.3kb D4Z4 unit, or whether multiple copies of the D4Z4 array were necessary for silencing. The major finding of our current study is that, from the entire 3.3 kb D4Z4, only one discrete region of ~146 nt was sufficient to initiate epigenetic silencing. This region overlaps with the 5P region of the D4Z4 [42–44], a region that was relatively hypomethylated in FSHD cells compared to controls. We showed that both DNA methylation and repressive histone modifications contributed to the silencing activity of this D4Z4 region, as treatment with DNA methyltransferase inhibitor 5-aza-dC or HDAC inhibitor entinostat partly relieved its suppressive effect. In an siRNA panel of several known D4Z4 regulators, suppression of reporter expression was relieved by knockdown of SETDB1, its binding partner ATF7IP, HDACs SIN3A/SIN3B, and LRIF1. In addition, enhanced suppression of the reporter by treatment with p38 inhibitors indicated that this pathway regulates the activity of a factor, or factors, involved in epigenetic silencing, and supports the efficacy of this reporter system as a platform for FSHD drug screening.

In our reporter system, only a subset of the factors previously shown to regulate D4Z4 epigenetic silencing were necessary for the repressive activity of D4Z4-S5. SETDB1 and its enhancer of nuclear stability cofactor ATF7IP [122] were necessary for silencing; however, TRIM28, part of a canonical SETDB1 repressive complex, was not necessary to suppress D4Z4-S5, despite its role in maintaining silencing of the D4Z4

region as shown in several other studies [46,89,86,88,126]. Future studies might test other factors in complexes with SETDB1, such as components of the human silencing hub (HUSH) complex [127–129]. The decreased H3K9me3 levels at the AAVS1 D4Z4-S5 reporter region upon knockdown of SETDB1 or ATF7IP suggest that this histone modification might be necessary to initiate silencing followed subsequently by DNA methylation. This agrees with our observations of D4Z4-S5 reporter DNA methylation, where low levels of GFP were correlated with higher levels of DNA methylation, though notably, many of the silenced cells had little or no DNA methylation. The presence of GFP-repressed cells with no DNA methylation suggests that silencing occurs initially through factor-mediated repression, such as SETDB1-mediated H3K9me3 deposition. The maintenance of high levels of H3K9me3 at the endogenous D4Z4 sequence corresponding to D4Z4-S5, despite loss of SETDB1/ATF7IP, suggests that factors associated with adjacent regions might be sufficient to maintain this repressive chromatin modification. In future studies, genome-wide clustered regularly interspaced short palindromic repeats (CRISPR) knockout screens have the potential to discover additional factors and pathways necessary to both initiate and maintain repression of D4Z4-S5.

Based on our ChIP analysis, the enrichment of SMCHD1 and LRIF1 at the D4Z4-S5 reporter was less substantial than at the endogenous LRIF1 promoter and, as noted above, the failure of SMCHD1 knockdown to de-repress D4Z4-S5 did not support a direct role in silencing through this region. This agrees with previous work in control myoblasts containing hypermethylated D4Z4 regions, where loss of SMCHD1 was not sufficient to alter H3K9me3 levels in the D4Z4 and led to only minor *DUX4* de-repression compared to levels present in FSHD2 myoblasts [56] and is also consistent with a recent report showing that SMCHD1 can play a role as a transcriptional co-activator in some contexts [123]. Future studies will need to determine whether extending the D4Z4-S5 fragment size or creating a multi-copy array will enhance SMCHD1 repressive activity.

The search for successful FSHD therapeutics has been challenging due to the sporadic and variable nature of disease pathogenesis in individuals and the difficulties of detecting low levels of *DUX4* RNA and protein, as *DUX4* is expressed sporadically in ~0.1%-1% of cells in patient myoblast samples [9,62,63]. Using artificial reporter screens in muscle cells, several compounds have been identified that inhibit expression of *DUX4* and its target genes [64–70]. For our reporter assay, we used a non-muscle cell line, HeLa cells, to avoid the dynamic regulation of the D4Z4 during skeletal muscle differentiation [118] and because of the higher efficiency of integration into the safe harbor locus. Extending these studies to human skeletal muscle cells will be important to confirm that D4Z4-S5, and no other D4Z4 segments, have similar silencing activity, and to assess the response during differentiation of myoblasts to differentiated muscle cells. To compare the efficacy of our system in screening FSHD therapeutics to

previously used muscle cell models, we assayed the effect of several *DUX4* inhibitors in D4Z4-S5 HeLa cells and identified the p38 inhibitor losmapimod as a repressor of GFP expression. Currently, the necessary targets of the p38 pathway that contribute to *DUX4* expression have not been elucidated, and our reporter system indicates that p38 pathway members may modulate expression through D4Z4-S5. Longer-term studies in our system, however, revealed that p38-inhibitor-induced silencing in D4Z4-S5 did not persist for the full time-course of treatment. This is consistent with a recent study showing p38 inhibition of *DUX4* expression in early myogenesis but not late myogenesis [66], suggesting that compensatory pathways might limit the duration of efficacy. In this context, the D4Z4-S5 reporter cell line might be particularly useful for drug screens to identify suppressors of D4Z4 without the secondary complications of altering myogenesis and with the added value of long-term treatment monitoring.

Ultimately, the discovery of D4Z4-S5-mediated silencing activity in this isolated reporter system provides a robust platform for identification of new D4Z4-regulating factors, pathways, and therapeutics. We used HeLa cells to establish initial proof-of-concept use of this system and believe it will be easily adaptable for other cell types. Flow cytometry of GFP expression is a simple and sensitive system for future CRISPR- or RNAi-mediated screens of epigenetic regulators that may interact with this important D4Z4-S5/5P regulatory sequence. This study also shows that cells with D4Z4-S5 can be used for temporal studies in multi-well live-cell imaging screens to detect robust changes in GFP+ cell confluency over time in response to treatment.

Materials and Methods

Cell culture

HeLa cells were grown in DMEM High Glucose with L-glutamine (Gibco, #11965092), supplemented with 10% fetal bovine serum (FBS) (Cytiva Hyclone, #SH30396.03) and 1% penicillin/streptomycin (Gibco, #15140122) and incubated at 37°C supplemented with 5% CO₂. Cells were passaged by treatment with 0.25% trypsin- Ethylenediaminetetraacetic acid (EDTA) (Gibco, #25200056) for 5 minutes at 37°C followed by growth media addition and subculturing. For the *de novo* silencing experiment, HeLa cells were treated with growth media supplemented with 30μM 5-aza-2'-deoxycytidine (Millipore Sigma #189825), 5μM entinostat (N/A) or equivalent volume of DMSO (Sigma-Aldrich, #D2650) for 12 hours before washout and addition of growth media for 12 hours, with 12 hour on/off treatments repeated for three days. Cells were then harvested 1, 3, 7, and 10 days after the final washout for harvest of genomic DNA (gDNA) and flow cytometry analysis of GFP. Cells were treated with *DUX4* inhibitors losmapimod (Sigma Aldrich, #SML3596), SB203580 (Selleckchem, #S1076), clenbuterol (Sigma Aldrich, #C5423), formoterol (Sigma-Aldrich, #F9552), (+)- JQ1 (Cayman Chemical, #11187-1), or berberine (Sigma Aldrich, #B3251) at escalating doses for three days before harvest for flow cytometry analysis.

AAVS1 silencing reporter knock-in cell line generation

Various D4Z4, LRIF1 promoter region, and control sequences were cloned into the enhancer region of the AAVS1 HA(L)-CMV-mEmerald-hPGK-PuroR-AAVS1 HA(R), parental plasmid pMK232 (CMV-OsTIR1-PURO) Addgene#72834, which was a gift from the lab of Robert Bradley (FHCC). This plasmid was linearized by inverse PCR (Phusion Plus PCR Master Mix, Thermo Scientific #F631S), to exclude the CMV enhancer region. The backbone was purified using NucleoSpin Gel and PCR cleanup mini kit (Machery-Nagel, #740609). Inserts were amplified via PCR with primers (IDT) containing backbone-overlapping sequences with flanking HindIII and Sall cut sites for easier subsequent cloning (primers [IDT] listed in Supplementary Material, Table S3) and gel purified. D4Z4 segments were amplified from the lambda42 fragment, which contains 2.5 copies of the D4Z4 [24]. LRIF1 promoter region and control fragments were PCR amplified from HeLa genomic DNA. These insert fragments were cloned into the PCR-amplified or HindIII-HF- (NEB, #R3104L) and Sall-HF- (NEB, #R3138L) digested linearized AAVS1 backbone via HiFi DNA Assembly Master mix (NEB, #E2621L) according to manufacturers' protocols. Plasmids were transformed into Stbl3 competent cells, extracted via PureLink Quick Plasmid Miniprep Kit (Invitrogen, #K210011), verified via sanger sequencing (FHCC Genomics core), and further grown and expanded for extraction via Nucleobond Xtra Maxi EF kit (Macherey-Nagel, #740424.50). AAVS1 donor plasmids and pX459-sgAAVS1 (Addgene #184403), which contains a single guide RNA (sgRNA) for the AAVS1 site and Cas9, were transfected into HeLa cells with Lipofectamine 3000 (Invitrogen, #L3000015) according to manufacturer's protocol. HeLa lines were selected for proper AAVS1 knock-in using puromycin selection (Sigma-Aldrich, #P8833) at 1.5 μ g/mL. Cells were cultured in puromycin for 1 week and then sorted for GFP^{High} cells via FACS. These GFP⁺ lines were then expanded without puromycin for two weeks before experimentation and analysis for GFP expression levels via flow cytometry. All cloning and sequencing primers (IDT) are listed in Supplementary Material, Table S3. Cell lines were genotyped for correct construct integration by extraction of genomic DNA according to manufacturer's instructions with GeneJet Genomic DNA Purification kit (Thermo Scientific, #K0722), PCR amplification (Supplementary Material, Fig. S1A, primers listed in Supplementary Material, Table S3) using Phusion Plus PCR Master Mix (Thermo Scientific, Cat#F631S), and gel electrophoresis of PCR products on 1% agarose in 1X TAE gels stained with ethidium bromide. CTRL, LRIF1-FLr, and D4Z4-S5 cell line genotyping PCR products were purified using the NucleoSpin Gel and PCR cleanup mini kit (Machery-Nagel, #740609) and sequenced using Premium PCR Sequencing performed by Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation.

Flow cytometry and FACS

Cells were trypsinized and resuspended in FACS Buffer (1X phosphate buffered saline (PBS) with 2%FBS and 5mM EDTA) and kept on ice until analysis. For flow cytometry analysis, live cell samples were run on one of several analyzers (BD Biosciences: FACSCanto II, LSR Fortessa x50, FACSymphony A3, or FACSymphony A5). For FACS, cells were sorted into tubes or plates with growth media using a BD Biosciences FACSymphony S6 or FACS Aria II instrument. GFP (mEmerald) fluorescence was quantified using Blue 488nm (200mW) laser readouts. For all samples

shown on the same graph, the same voltage parameters and machine was used for consistency. This analysis was supported by the Flow Cytometry Shared Resource (RRID:SCR_022613) of the Fred Hutch/University of Washington/Seattle Children's Cancer Consortium (P30 CA015704). FlowJo v10.8 software (BD Biosciences) was used for all flow cytometry and FACS analysis. For all GFP fluorescence intensity graphs, cells were first gated for live cells to eliminate debris using forward and side scatter (FSC and SSC), then gated on single cells to eliminate doublets using FSC-Height and FSC-Width. GFP channel values reflect log scale distribution of fluorescence intensity.

Quantitative reverse transcription PCR (RT-qPCR)

Total cellular RNA was isolated using the Nucleospin RNA kit (Machery-Nagel, #740955.250) according to the manufacturer protocol. 0.5-1 μ g RNA was treated with Amplification grade-DNaseI (Thermo Fisher, #18068015), heat inactivated with EDTA (Thermo Fisher, #AM9260G), and reverse transcribed into cDNA using random hexamers or oligo-dT with the Superscript IV First-Strand Synthesis System (Invitrogen, #18091050) following manufacturer instructions. Quantitative PCR was run on a QuantStudio 7 Flex (Applied Biosystem) using iTaq SYBR Green Supermix (Bio-Rad, #1725124). Primers (IDT) are listed in Supplementary Material, Table S3.

Immunoblotting

Cells were lysed in radioimmunoprecipitation assay (RIPA) buffer (150mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 1% sodium dodecyl sulfate (SDS), 25mM Tris-HCl pH7.4) with protease and phosphatase inhibitors (Pierce, #A32961) Protein lysate was sonicated in a Bioruptor Sonication System (Diagenode) and cleared by centrifugation at 16,000xg. Lysate was quantified via Pierce bicinchoninic acid (BCA) assay (Thermo Scientific, #23225). Protein lysates were run on NuPAGE 4-12% Bis Tris polyacrylamide gels (Invitrogen, #NP0322BOX) using NuPAGE MOPS buffer (Invitrogen, #NP0001) and transferred onto polyvinylidene fluoride (PVDF) membranes (Invitrogen, #LC2002). Membranes were blocked in PBS-Tween (PBST) buffer (0.1% Tween-20 in 1XPBS) with 5% non-fat dry milk before incubation overnight at 4°C with primary antibody diluted in PBST buffer (primary antibodies listed in Supplementary Material, Table S3). Membranes were incubated in horseradish peroxidase-conjugated secondary antibody (Supplementary Material, Table S3) for one hour at room temperature. SuperSignal chemiluminescent substrate (Thermo Scientific, #34580) was added to membranes, and a Mini-medical 90 processor (AFP Manufacturing) was used for detection on film.

siRNA transfections

HeLa AAVS1 reporter cells were seeded into 6-well plates 24 hours before transfection. 25-50pmol of siRNA was diluted in OptiMEM (Gibco, #31985070) and transfected into cells using Lipofectamine RNAiMAX (Invitrogen #13778150) following manufacturer's instructions. 48 hours after transfection, wells were changed to fresh growth media and the siRNA transfection was repeated. 48 hours after the second transfection, cells were harvested for flow cytometry of GFP fluorescence, RNA isolation for RT-qPCR analysis, and immunoblot analysis.

Bisulfite sequencing and % CpG Methylation Analysis

HeLa AAVS1 reporter lines were harvested, pelleted, and flash frozen in a dry ice/ethanol slurry before storage at -80°C. Genomic DNA was isolated using the GeneJet Genomic DNA Purification kit (Thermo Scientific, #K0722) following manufacturer instructions. 200-500ng of genomic DNA was subjected to bisulfite conversion using EZ DNA Methylation-Gold Kit (Zymo Research, #D5005) following manufacturer instructions. Bisulfite-converted DNA was then used for PCR amplification of the AAVS1 insert regions using ZymoTaq DNA Polymerase (Zymo Research, #E2001) and degenerate primers (IDT, Supplementary Material, Table S3). PCR products were gel purified using the NucleoSpin Gel and PCR cleanup mini kit (Machery-Nagel, #740609) and sequenced using Premium PCR Sequencing performed by Plasmidsaurus using Oxford Nanopore Technology with custom analysis and annotation. Reads were aligned to insert contigs using BWAMeth.py [130] then filtered to only includes reads covering 90% of the contig length using samtools [131]. After subsetting, reads were imported into R [132] using GenomicAlignments [133], converted to a dataframe via BioStrings [134] and percent methylation was calculated using the tidyverse package of tools [135]. Raw sequencing reads are presented in Supplementary Material, Table S2.

ChIP-qPCR

ChIP-qPCR was adapted from previously published protocols [136,137]. HeLa AAVS1 cells were trypsinized, washed with 1xPBS, and fixed in 1% Formaldehyde Ultrapure EM grade (EMS, #15710) for 10 minutes at room temperature with gentle rocking. Formaldehyde was quenched with 125mM glycine, cells were washed with PBS and then lysed in Lysis Buffer (Protease and phosphatase inhibitors (Pierce, #A32961), 1% Na-deoxycholate, 1% SDS, 5mM EDTA pH 8, and 50mM Tris pH 8 in 1xPBS). Lysate was sonicated in Biorupter Plus TPX Polymethyl pentene tubes (Diagenode, C30010010) using a Biorupter Plus Sonication System (Diagenode, B01020001) until chromatin was sheared to a range of 200-600bp; a small aliquot of sheared chromatin was incubated with RNase A at 37°C for one hour, treated with Proteinase K overnight at 65°C, and purified using a NucleoSpin Gel and PCR cleanup mini kit (Machery-Nagel, #740609) before concentration was determined using a Qubit and chromatin was run on a 2% agarose gel in Tris/Borate/EDTA (TBE) buffer to assess fragment size. Chromatin was centrifuged at 16,000xg for 10 min at 4°C to clear insoluble fraction, and 10% of the chromatin was reserved for input DNA.

For immunoprecipitation, chromatin was diluted in Dilution Buffer (1% Triton X-100, 150mM NaCl, 2mM EDTA pH 8, and 20mM Tris pH 8 in PBS) to reduce SDS concentration to <0.1%. For LRIF1 and SMCHD1 ChIP, 5µg of antibody (rbαSMCHD1 Abcam #Ab31865, rbαAnti-Ligand-dependent nuclear receptor-interacting factor 1, Millipore #ABE1008, or Rabbit IgG control Millipore #PP64) was added to 30µg of chromatin and incubated with rotation overnight at 4°C. Agarose Fast Flow Prot A (Millipore Sigma #16-156) beads were pre-blocked overnight at 4°C in Dilution Buffer + 2% bovine serum albumin (BSA). For histone ChIP, 2-4µg of antibody (rbαH3K9me3 Abcam #Ab8898, or Rabbit IgG control Millipore #PP64) was added to ~2-5µg of chromatin and incubated with rotation overnight at 4°C. Dynabeads Protein A (Invitrogen #10002D) beads were pre-blocked overnight at 4°C in Dilution Buffer + 2%

BSA. Samples were incubated with beads at 4°C for 2 hours, and washed with Wash Buffer I (protease and phosphatase inhibitors (Pierce, #A32961), 0.1% SDS, 1% Triton X-100, 2mM EDTA pH 8, 20mM Tris pH 8, 150mM NaCl), washed with Wash Buffer II (protease and phosphatase inhibitors, 0.1% SDS, 1% Triton X-100, 2mM EDTA pH 8, 20mM Tris pH 8, 500mM NaCl), washed with Wash Buffer III (protease and phosphatase inhibitors, 0.25M LiCl, 1% NP-40, 1% Na-deoxycholate, 1mM EDTA pH 8, 10mM Tris pH 8, 500mM NaCl), and washed twice with ice cold PBS. Beads and attached chromatin were then de-crosslinked by resuspension in Ultrapure water with 1% SDS and 0.1M NaHCO₃, incubated with RNase A (Thermo Scientific, # R1253) for 1 hour at 37°C, and incubated overnight at 65°C with Proteinase K (Thermo Scientific, #EO0491). IP DNA was then purified using NTB Buffer (Machery-Nagel, #740595.150) and the NucleoSpin Gel and PCR cleanup mini kit (Machery-Nagel, #740609). Quantitative PCR of IP DNA was run on a QuantStudio 7 Flex (Applied Biosystem) using iTaq SYBR Green Supermix (Bio-Rad, #1725124).

Incucyte tracking of GFP+ cell confluence

HeLa AAVS1 lines were seeded at 80% confluency and 24 hours later were treated with 5-aza-2'-deoxycytidine (Millipore Sigma #189825) for 12 hours. Cells were then incubated with standard media for 12 hours; this was repeated twice more, for a total of 36 hours of 5-aza-2'-deoxycytidine treatment in 72 hours. 24 hours after washing out the third round of 5-aza-2'-deoxycytidine and replacement with standard media, GFP+ cells were sorted by FACS at a density of 10,000 cells per well in a 96-well plate with 100uL media per well. 24 hours later, 100uL of media with 2x concentration of losmapimod (Sigma Aldrich, SML3596) or doramapimod (MedChem Express, #HY-10320) (1μM, final concentration 0.5μM) and 96-well plates were placed in the Incucyte S3 Live Cell Analysis system, kept at 37°C with 5% CO₂. Wells were imaged for GFP+ and brightfield cell confluency, cell count, and GFP Mean Fluorescence intensity every 2 hours for 6 days. On day 3, plates were removed from the Incucyte between images and media was changed and supplemented with fresh p38 inhibitors. GFP+ cell confluency per well was normalized to t0 for that well and to the average of DMSO treated wells to analyze differences in p38 inhibitor-treated wells.

Figures

Figure 2.1: A safe harbor system identifies silencing activity driven by specific sequences in the *LRIF1* promoter.

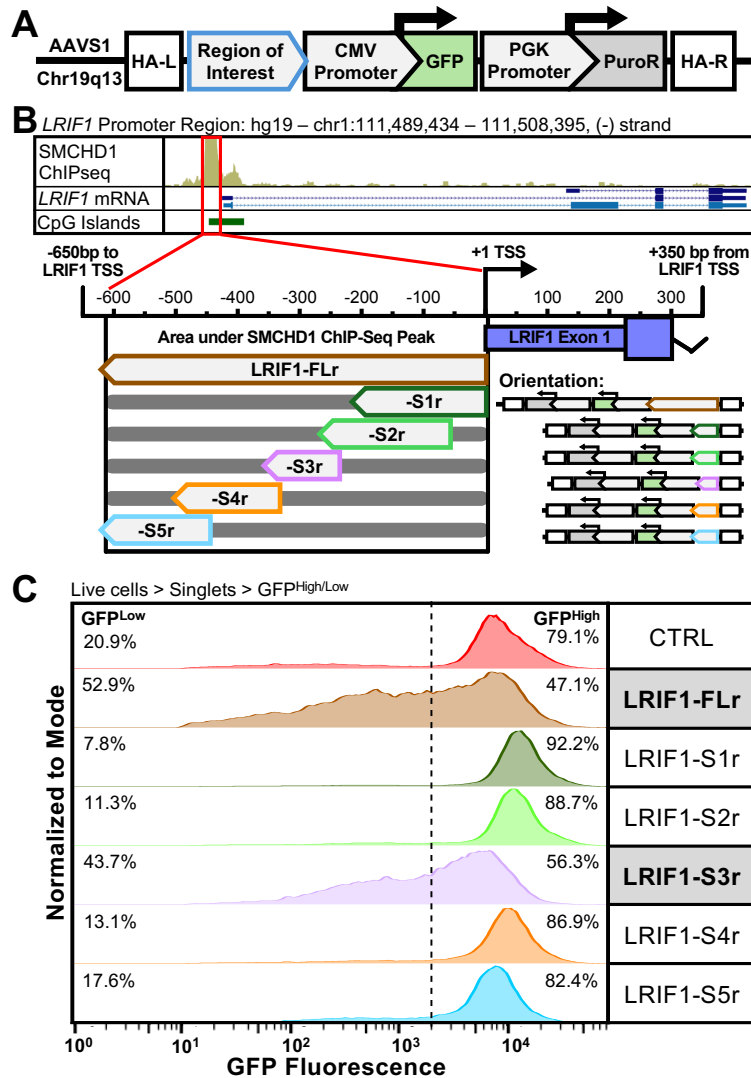


Figure 2.1: (A) Schematic of the safe harbor construct integrated into the *AAVS1* locus in HeLa cells. A sequence of interest is inserted upstream of a constitutive CMV promoter driving GFP. HA-L and HA-R= left and right homology arm. **(B)** Top: UCSC genome browser track of SMCHD1 ChIP-seq peak [60], *LRIF1* mRNA, and CpG islands at the *LRIF1* promoter region. Bottom: Sequence maps of *LRIF1* promoter region segments. Arrowheads indicate orientation of sequence relative to genomic sequence and construct schematics indicate orientation of sequences relative to integrated reporter constructs (TSS = transcription start site, bp = base pair). **(C)** Flow cytometry histogram of GFP fluorescence in cells containing various integrated *LRIF1* promoter silencing constructs (n=1 biological replicate of polyclonal line). Histograms are normalized to the mode. Percentages of cell populations classified GFP^{High} or GFP^{Low} are indicated.

Figure 2.2: Segmentation of the D4Z4 reveals one sequence with silencing activity.

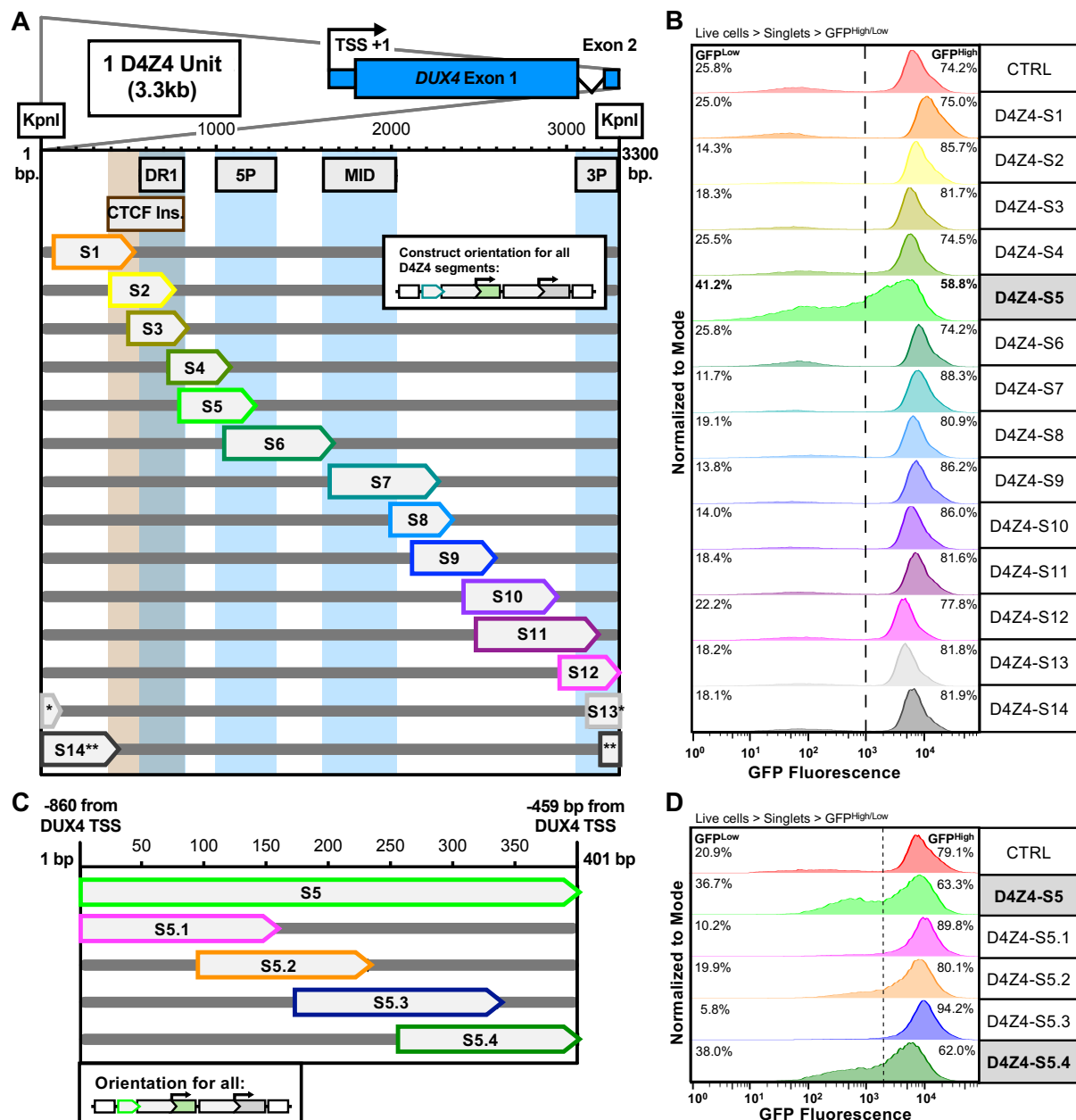


Figure 2.2: (A) Schematic of one D4Z4 repeat unit and sequence segments inserted into the silencing reporter construct (TSS = transcription start site, bp = base pair). Previously identified differentially methylated regions are highlighted (30, 41-44). Asterisks (*) indicate segment spans two D4Z4 units, across the KpnI site. **(B)** Flow cytometry histogram of GFP fluorescence in cells containing various integrated D4Z4 segment constructs (n=1 biological replicate of polyclonal line). Histograms are normalized to the mode. Percentages of cell populations classified GFP^{High} or GFP^{Low} are indicated. **(C)** Schematic of D4Z4-S5 and sequence sub-fragments inserted into the silencing reporter construct (TSS = transcription start site, bp = base pair). **(D)** Flow

cytometry histogram of GFP fluorescence in cells (n=1 biological replicate of polyclonal line) containing various integrated D4Z4-S5 subfragment silencing constructs. Histograms are normalized to the mode. Percentages of cell populations classified GFP^{High} or GFP^{Low} are indicated.

Figure 2.3: GFP silencing activity of D4Z4-S5 is partially due to de novo DNA methylation and repressive histone modifications.

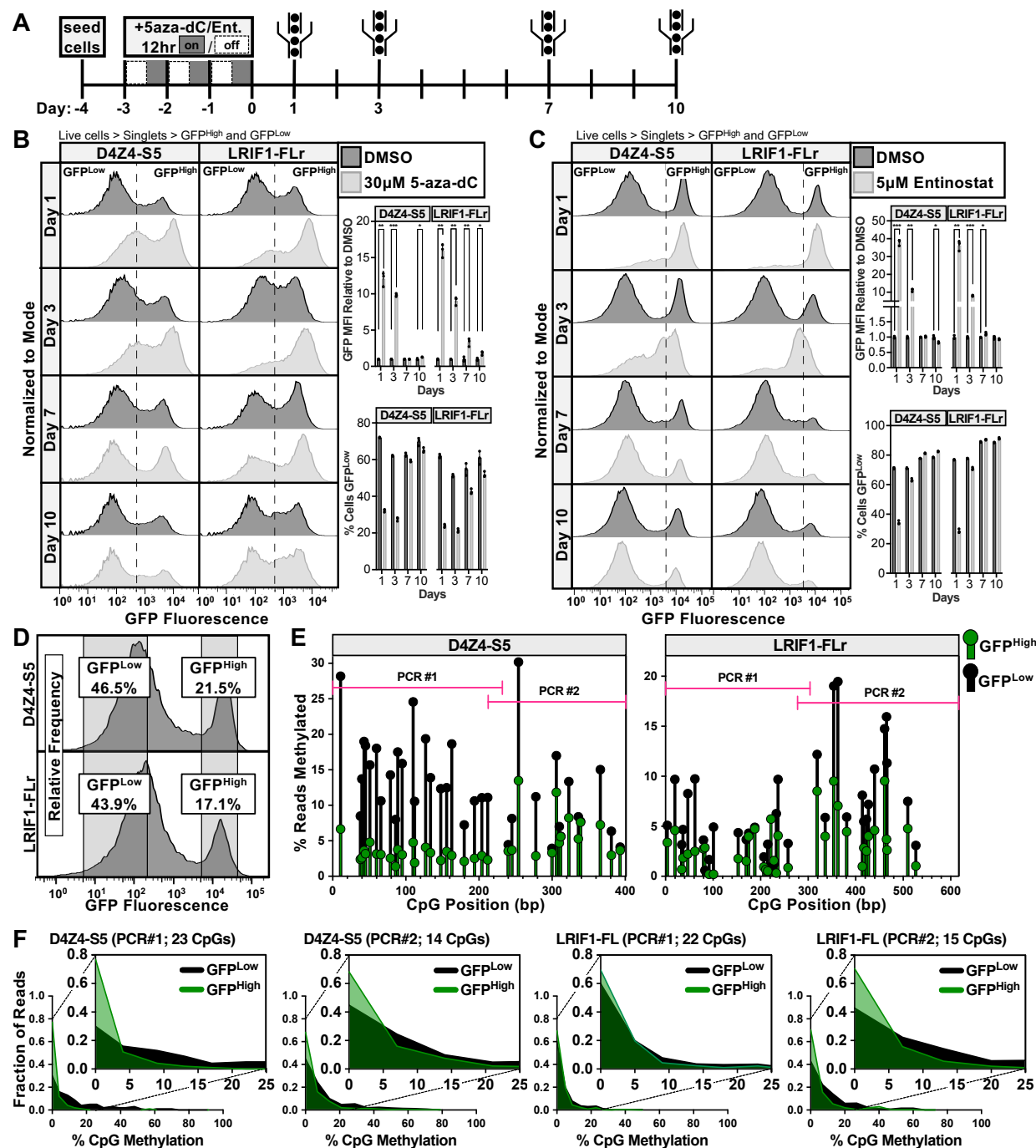


Figure 2.3: (A) Schematic of experimental time course. Cells were treated with 5-aza-dC or entinostat (Ent.) for 3 days with a 12 hour on/off schedule and harvested for flow cytometry analysis at 1, 3, 7, and 10 days after washout. **(B, C)** Left: Representative singleton GFP fluorescence histograms of D4Z4-S5 and LRIF1-FLr cells at various timepoints after 5-aza-dC **(B)** or entinostat **(C)** treatment. Histograms are normalized to the mode of the population. Right top: Fold change in GFP MFI at various timepoints

after treatment compared to the DMSO control. Data represent mean $\pm$ SD of biological replicates, $n=3$. Statistical significance was determined by Welch's t-test: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Right bottom: Percentage of GFP^{Low} cells in the population at various time points after treatment. Data represent mean $\pm$ SD of biological replicates, $n=3$. **(D)** GFP fluorescence histograms of D4Z4-S5 and LRIF1-FLr, plotted based on relative frequency. Gray boxes indicate FACS sorting gates for GFP^{High} and GFP^{Low} populations used in subsequent DNA methylation analysis. **(E)** Lollipop plots of methylation levels at individual CpGs for D4Z4-S5 (left) and LRIF1-FLr (right). The percentage of nanopore sequencing reads of bisulfite PCR products with methylation (no C>T conversion) at each indicated CpG is plotted for GFP^{Low} and GFP^{High} populations. **(F)** Fraction of bisulfite PCR nanopore sequencing whole reads with various percentages of CpG methylation for GFP^{Low} and GFP^{High} cell populations. Inset graph shows read fractions with less than 25% total CpG methylation.

Figure 2.4: Loss of histone methyltransferases and deacetylases specifically relieves D4Z4-S5 and LRIF1-FLr GFP silencing.

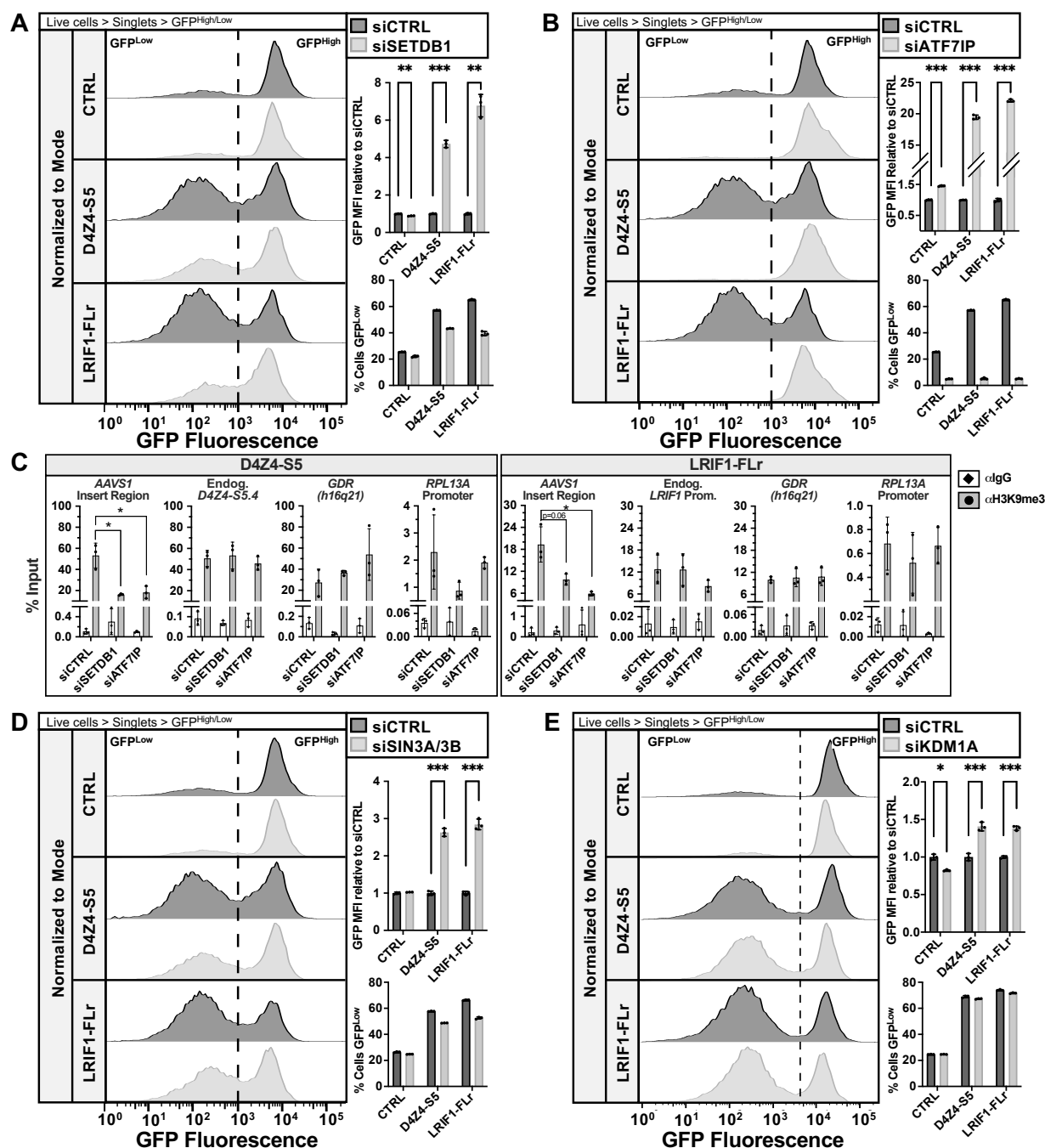


Figure 2.4: (A, B, D, E) Left: Representative singleton GFP fluorescence histograms of CTRL, D4Z4-S5 and LRIF1-FLr cells treated with CTRL or SETDB1 (A), ATF7IP (B), SIN3A/B (D), or KDM1A (E) siRNAs. Histograms are normalized to the mode of the population. Right Top: Fold change in GFP MFI upon siRNA-mediated knockdown of SETDB1 (A), ATF7IP (B), SIN3A/B (D), or KDM1A (E) compared to siCTRL. Data represent mean $\pm$ SD of biological replicates, $n=3$. Statistical significance was

determined by Welch's t-test: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Right bottom: Percentage of GFP^{Low} cells in the population at various time points after siRNA-mediated knockdown of SETDB1 (**A**), ATF7IP (**B**), SIN3A/B (**D**), or KDM1A (**E**) compared to siCTRL. Data represent mean $\pm$ SD of biological replicates, $n=3$. (**C**) ChIP-qPCR of H3K9me3 occupancy at various loci in D4Z4-S5 and LRIF1-FLr cells upon treatment with siCTRL, siSETDB1, or siATF7IP (GDR = gene desert region with high H3K9me3, the *RPL13A* promoter is an active gene control for low H3K9me3). Data is graphed as the percent of total chromatin input that was immunoprecipitated by IgG or H3K9me3 antibodies. Data represent mean $\pm$ SD of biological replicates, $n=3$. Statistical significance was determined by Welch's t-test: * $p < 0.05$.

Figure 2.5: SMCHD1 and LRIF1 are enriched at D4Z4-S5 and LRIF1-FLr yet display different silencing roles.

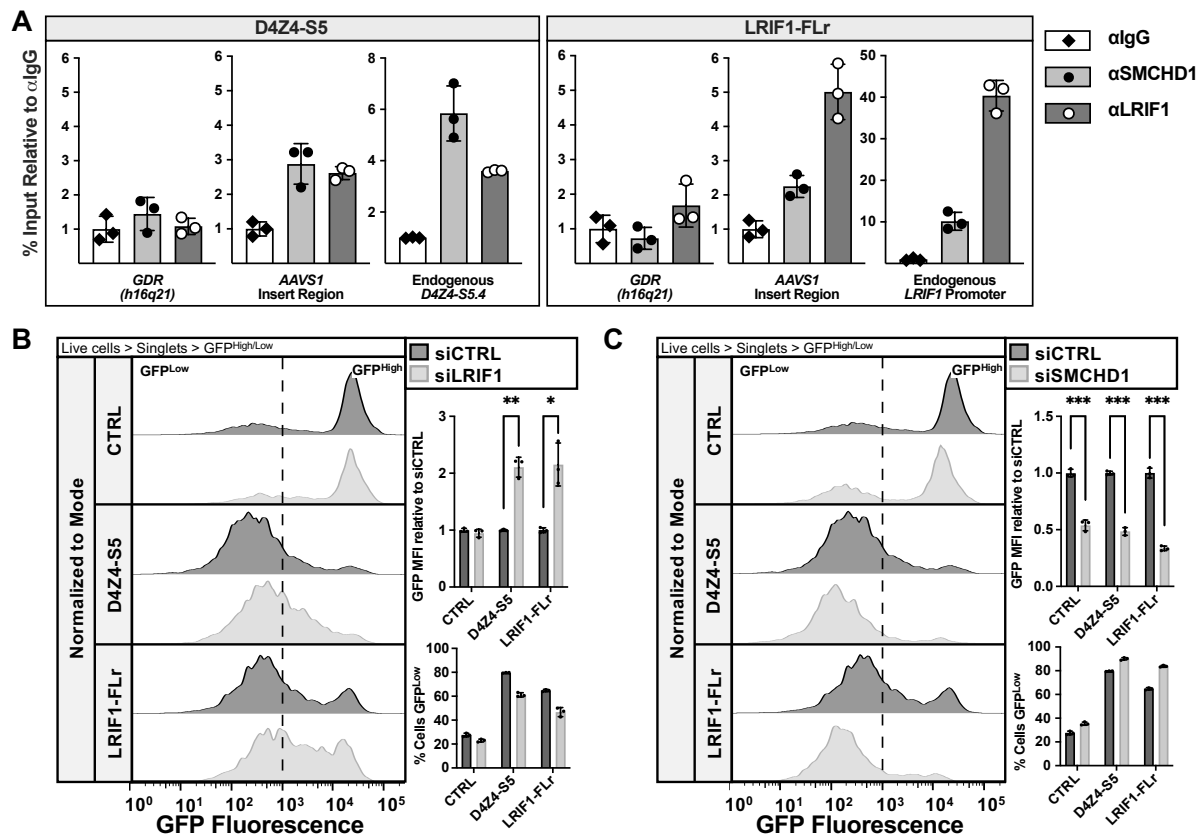


Figure 2.5: (A) ChIP-qPCR analysis of SMCHD1 and LRIF1 occupancy at various endogenous loci and the AAVS1 silencing construct transgene in D4Z4-S5 and LRIF1-FLr cells (GDR = gene desert region). Data represents the % of input chromatin immunoprecipitated by SMCHD1 and LRIF1 antibodies relative to IgG of three technical replicates (n=1 biological replicate). **(B, C)** Left: Representative singleton GFP fluorescence histograms of CTRL, D4Z4-S5 and LRIF1-FLr cells treated with CTRL or LRIF1 **(B)** or SMCHD1 **(C)** siRNAs. Histograms are normalized to the mode of the population. Right Top: Fold change in GFP MFI upon siRNA-mediated knockdown of LRIF1 **(B)** or SMCHD1 **(C)** compared to siCTRL. Data represent mean $\pm$ SD of biological replicates, n=3. Statistical significance was determined by Welch's t-test: *p<0.05, **p<0.01, ***p<0.001. Right bottom: Percentage of GFP^{Low} cells in the population at various time points after siRNA-mediated knockdown of LRIF1 **(B)** or SMCHD1 **(C)** compared to siCTRL. Data represent mean $\pm$ SD of biological replicates, n=3.

Figure 2.6: Candidate p38 inhibitor FSHD therapeutics inhibit GFP expression specifically in D4Z4-S5 and LRIF1-FLr.

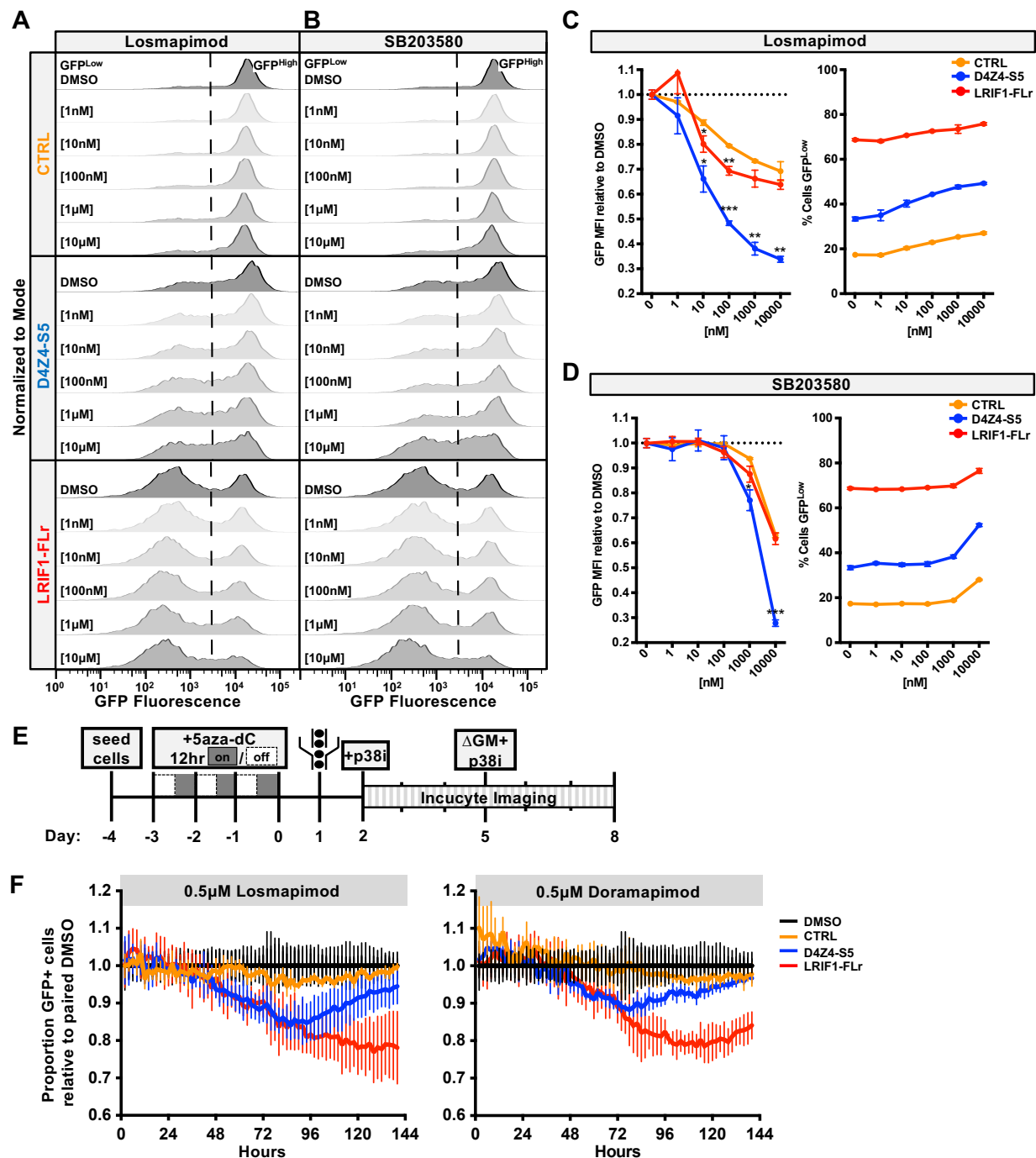
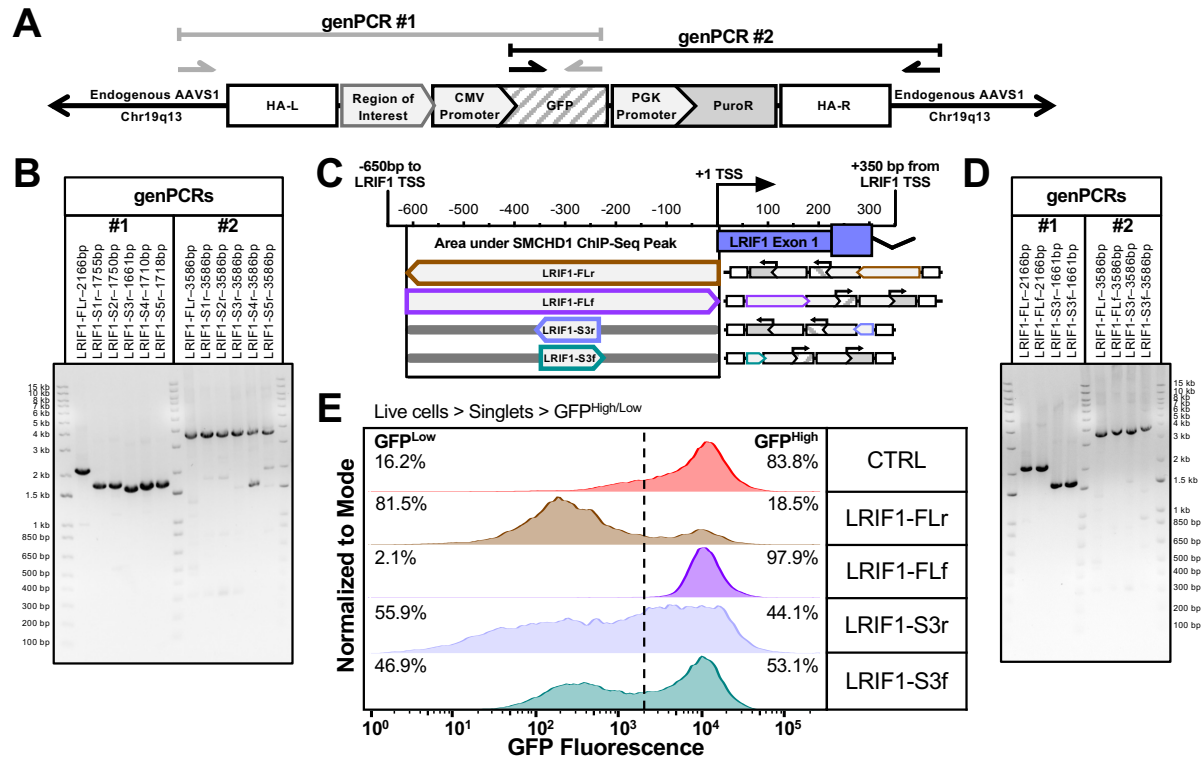


Figure 2.6: (A, B) Representative singleton GFP fluorescence histograms of CTRL, D4Z4-S5 and LRIF1-FLr cells treated for 3 days with DMSO or escalating doses of p38 pathway inhibitors Losmapimod **(A)** or SB203580 **(B)**. Histograms are normalized to the mode of the population. **(C, D)** Left: Fold change in GFP MFI upon 3-day treatment with escalating doses of Losmapimod **(C)**, or SB203580 **(D)** compared to DMSO treatment in CTRL, D4Z4-S5, and LRIF1-FLr cells. Data points represent mean fold change $\pm$ SD of

biological replicates, $n=3$. Statistical significance was determined by Welch's t-test: $*p<0.05$, $**p<0.01$, $***p<0.001$. Right: Percentage of GFP^{Low} cells in the population upon treatment with escalating doses of Losmapimod (**C**) or SB203580 (**D**). Data represent mean $\pm$ SD of biological replicates, $n=3$. (**E**) Schematic of live cell imaging experimental time course. Cells were treated with 5-aza-dC for three days with a 12 hour on/off schedule. 24 hours after washout, GFP^{High} cells were FACS sorted into multi-well plates. 24 hours after sort, cells were treated with p38 inhibitors (p38i) and incubated for 6 days in an Incucyte live cell imaging system, and images of each well were captured every 2 hours for GFP fluorescence quantification. On day 3 of Incucyte incubation, cell growth media (GM) was refreshed and p38i were replenished. (**F**) Plots of GFP+ cell confluency over 6 days upon treatment with Losmapimod (left) or Doramapimod (right). Data shown is the GFP+ cell confluency in p38i-treated wells relative to DMSO-treated wells in CTRL, D4Z4-S5, or LRIF1-FLr cells. Data points are the average relative GFP+ cell confluency $\pm$ SD of biological replicate wells, $n=6$.

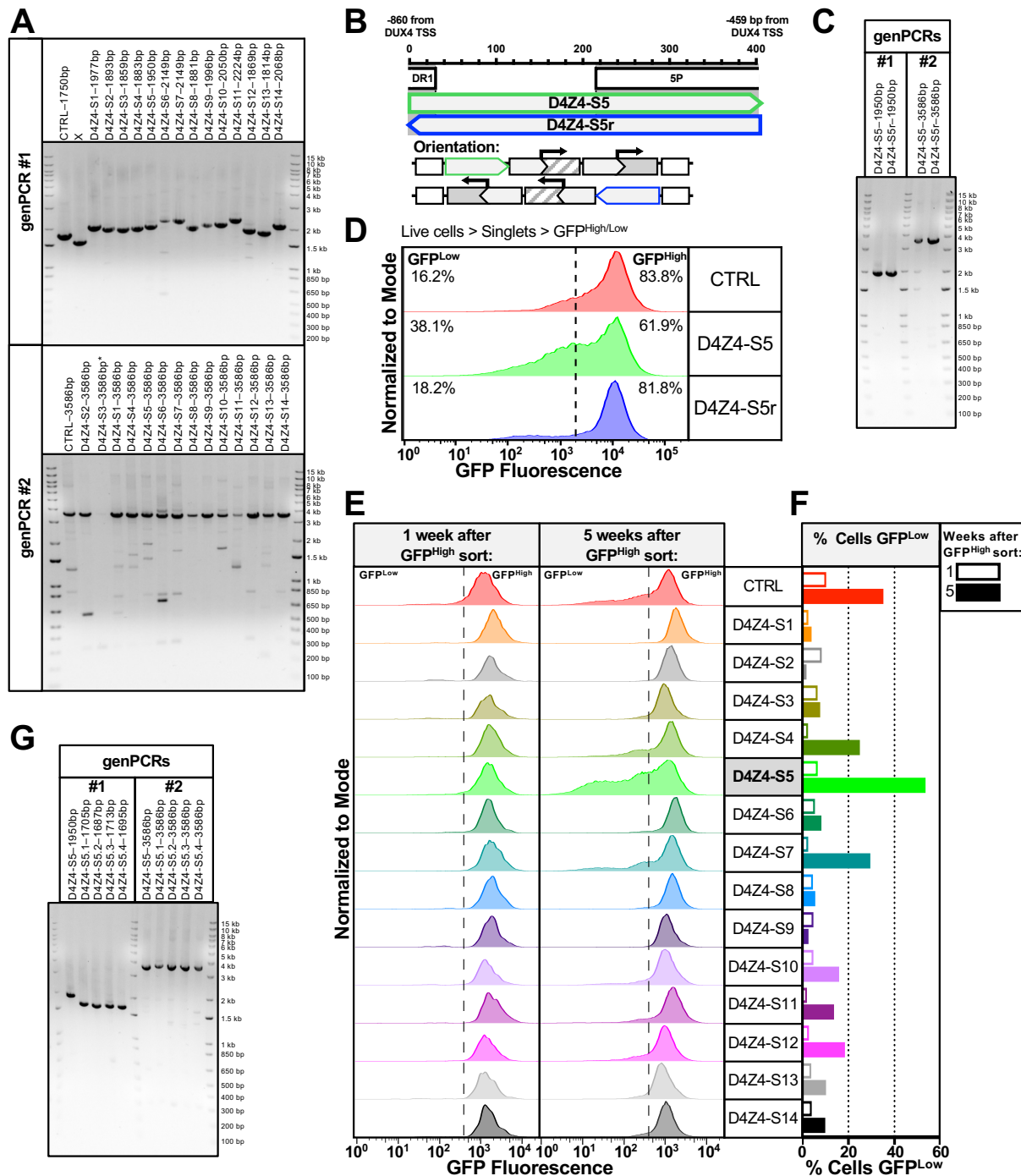
Supplementary Figures

Supplementary Figure S2.1: PCR genotyping reveals proper integration of LRIF1 constructs, and silencing activity of LRIF1-S3, but not LRIF1-FL, is independent of sequence orientation.



Supplementary Figure S2.1: (A) Schematic of genotyping PCR primer locations. Overlapping PCRs each contain one primer outside the homology arms in the AAVS1 genomic region to insure correct integration. **(B, D)** Gel electrophoresis indicating size of genotyping PCR products in LRIF1 construct lines. **(C)** Schematic of LRIF1-FLr, LRIF1-FLf, LRIF1-S3r and LRIF1-S3f orientation relative to the genomic sequence and the integrated silencing reporter construct (TSS = transcription start site, bp = base pair). **(E)** Flow cytometry GFP fluorescence histogram of LRIF1-FLr/f and LRIF1-S3r/f compared to CTRL cells, normalized to mode.

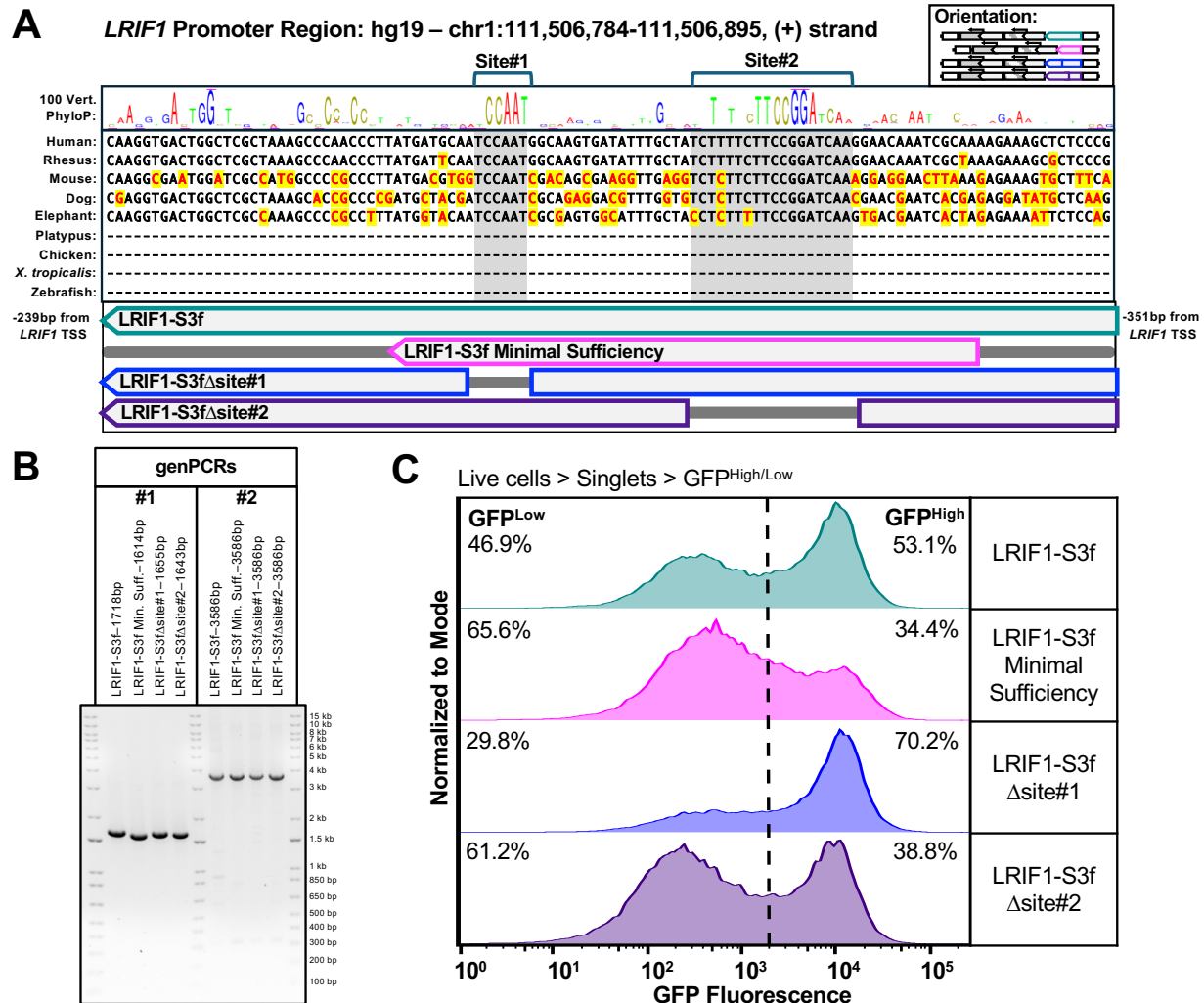
Supplementary Figure S2.2: D4Z4-S5 silencing activity is orientation-dependent and increases over time in culture.



Supplementary Figure S2.2: (A, C, G) Gel electrophoresis indicating size of genotyping PCR products in D4Z4 (A), D4Z4-S5 and D4Z4-S5r (C), and D4Z4-S5 sub-segmentation (G) construct cell lines. * Indicates faint band due to low concentration genomic DNA. (B) Schematic of D4Z4-S5 and D4Z4-S5r orientation relative to the genomic sequence and the integrated silencing reporter construct (TSS = transcription start site, bp = base pair). (D) Flow cytometry GFP fluorescence histogram of D4Z4-S5

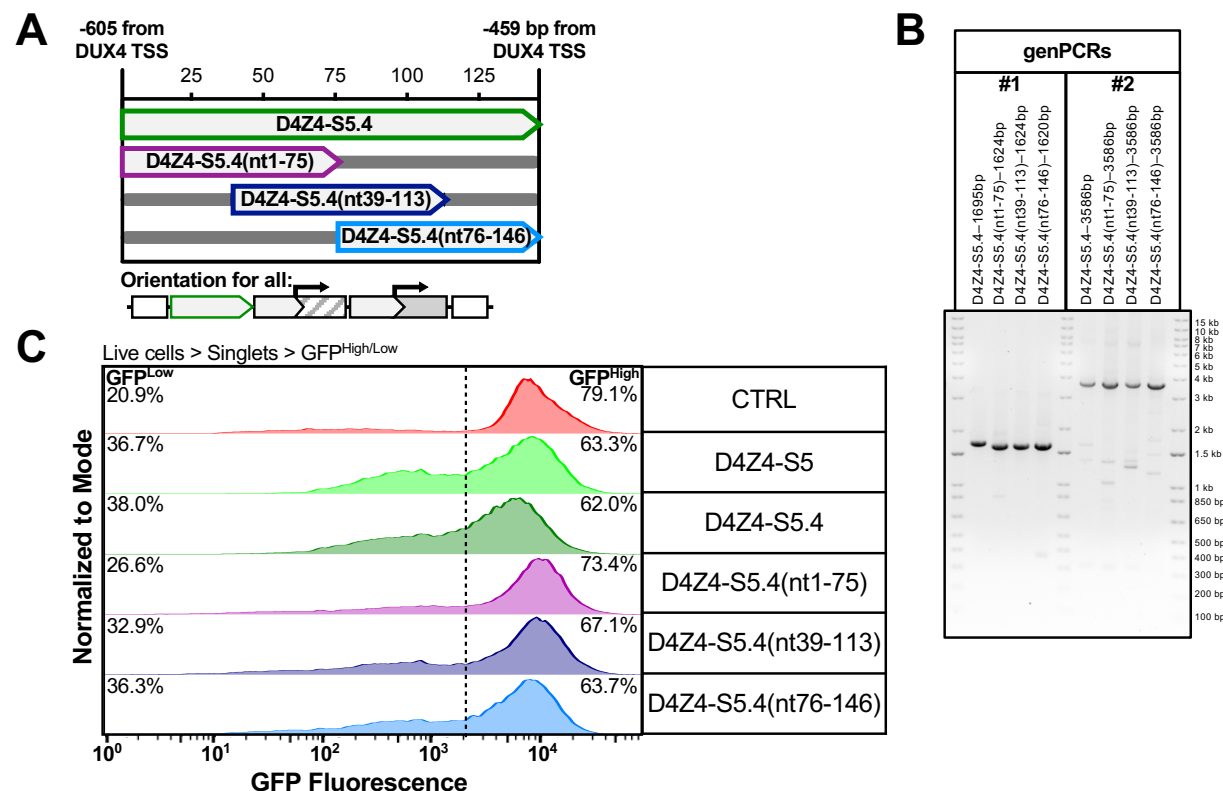
and D4Z4-S5r compared to CTRL cells, normalized to mode. **(E)** Flow cytometry GFP fluorescence histogram of D4Z4 silencing construct segments 1 (left) and 5 (right) weeks after GFP^{High} FACS sort, normalized to mode. **(F)** Percentage of GFP^{Low} cells in each D4Z4 silencing construct line 1 (clear) or 5 (filled) weeks after GFP^{High} FACS sort.

Supplementary Figure S2.3: LRIF1-S3 contains a conserved sequence that conveys silencing activity.



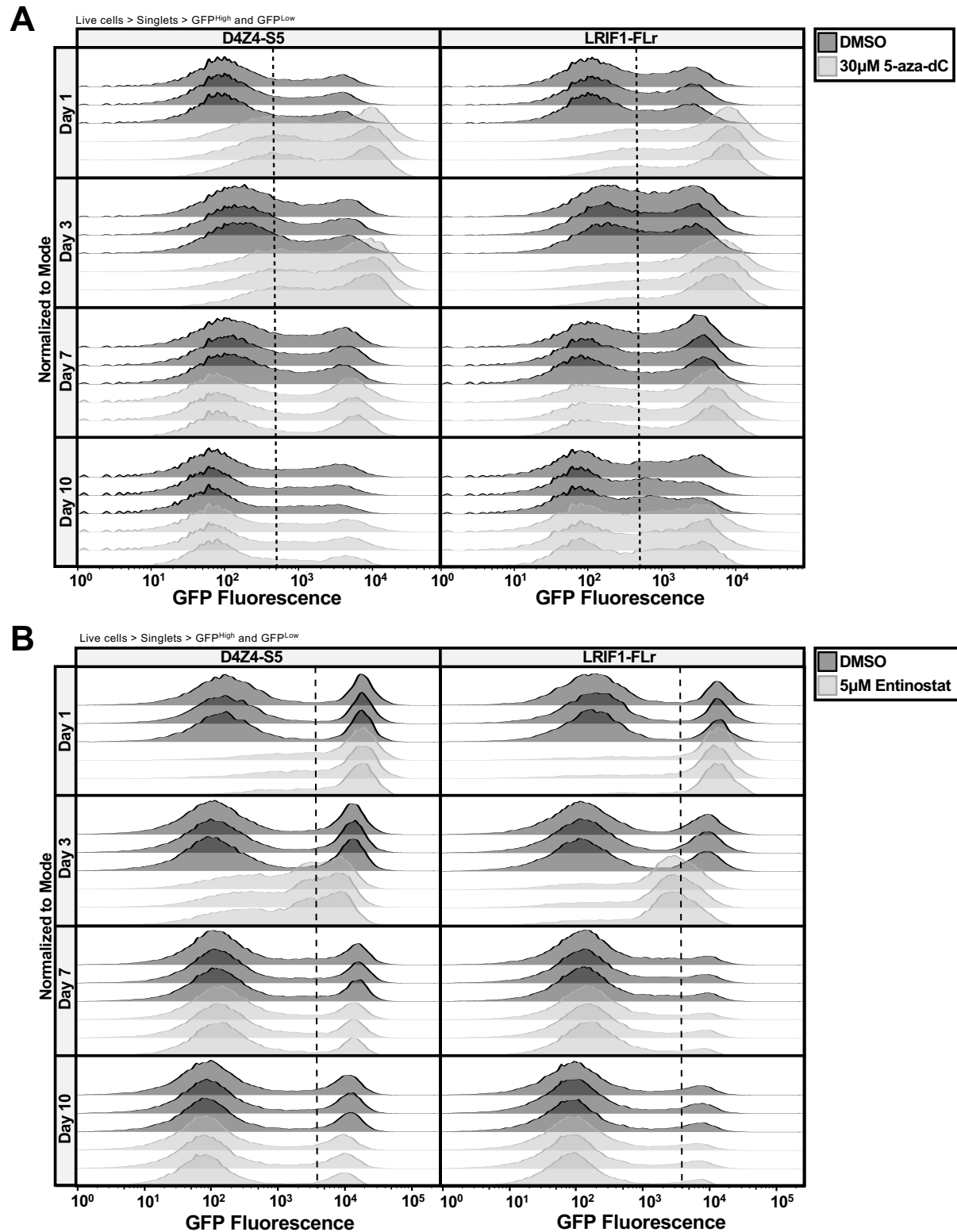
Supplementary Figure S2.3: (A) Top: Logo plot of LRIF1-S3 basewise conservation consensus sequence based on UCSC genome browser phyloP analysis of 100 vertebrates. Middle: LRIF1-S3 sequences in 9 representative species show conservation only in placental mammals. Bottom: Schematic of LRIF1-S3 conserved site deletion inserts with arrowheads indicating orientation relative to genomic sequence and *LRIF1* transcriptional start site (TSS), (bp = base pair). **(B)** Gel electrophoresis indicating size of genotyping PCR products in LRIF1-S3 conserved site deletion construct cell lines. **(C)** Flow cytometry GFP fluorescence histogram of LRIF1-S3 conserved site deletion inserts, normalized to mode. Percentages of cell populations classified GFP^{High} or GFP^{Low} are indicated.

Supplementary Figure S2.4: Further segmentation of D4Z4-S5.4 does not narrow region that confers silencing activity.



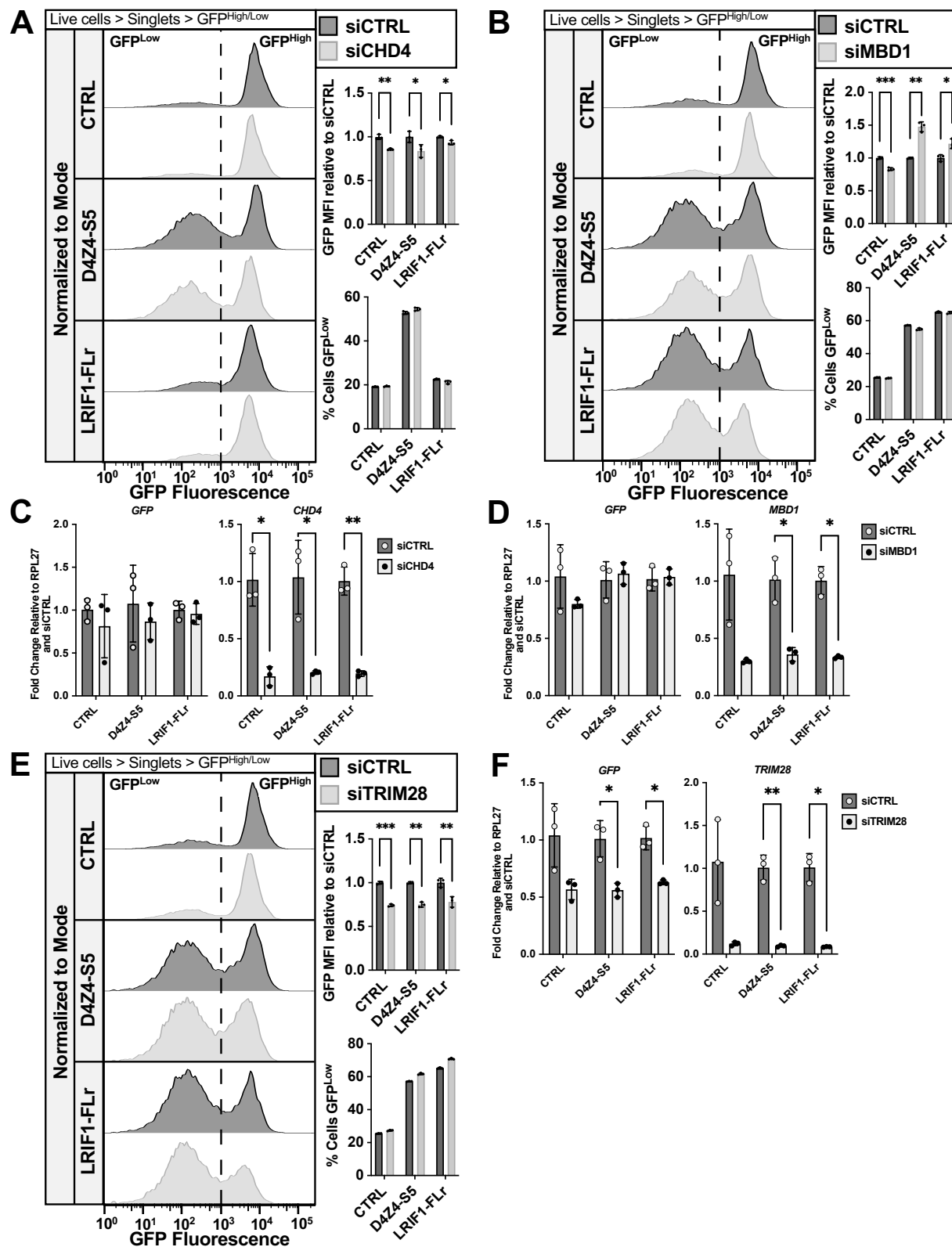
Supplementary Figure S2.4: (A) Schematic of D4Z4-S5.4 sub-fragment inserts relative to *DUX4* TSS (nt=nucleotide). **(B)** Gel electrophoresis indicating size of genotyping PCR products in D4Z4-S5.4 sub-segmentation construct cell lines. **(C)** Flow cytometry GFP fluorescence histogram of D4Z4-S5.4 sub-fragment inserts, normalized to mode. Percentages of cell populations classified GFP^{High} or GFP^{Low} are indicated.

Supplemental Figure S2.5: Treatment of D4Z4-S5 or LRIF1-FLr cells with 5aza-dC or Entinostat initially increases GFP expression, followed by GFP re-silencing over time.



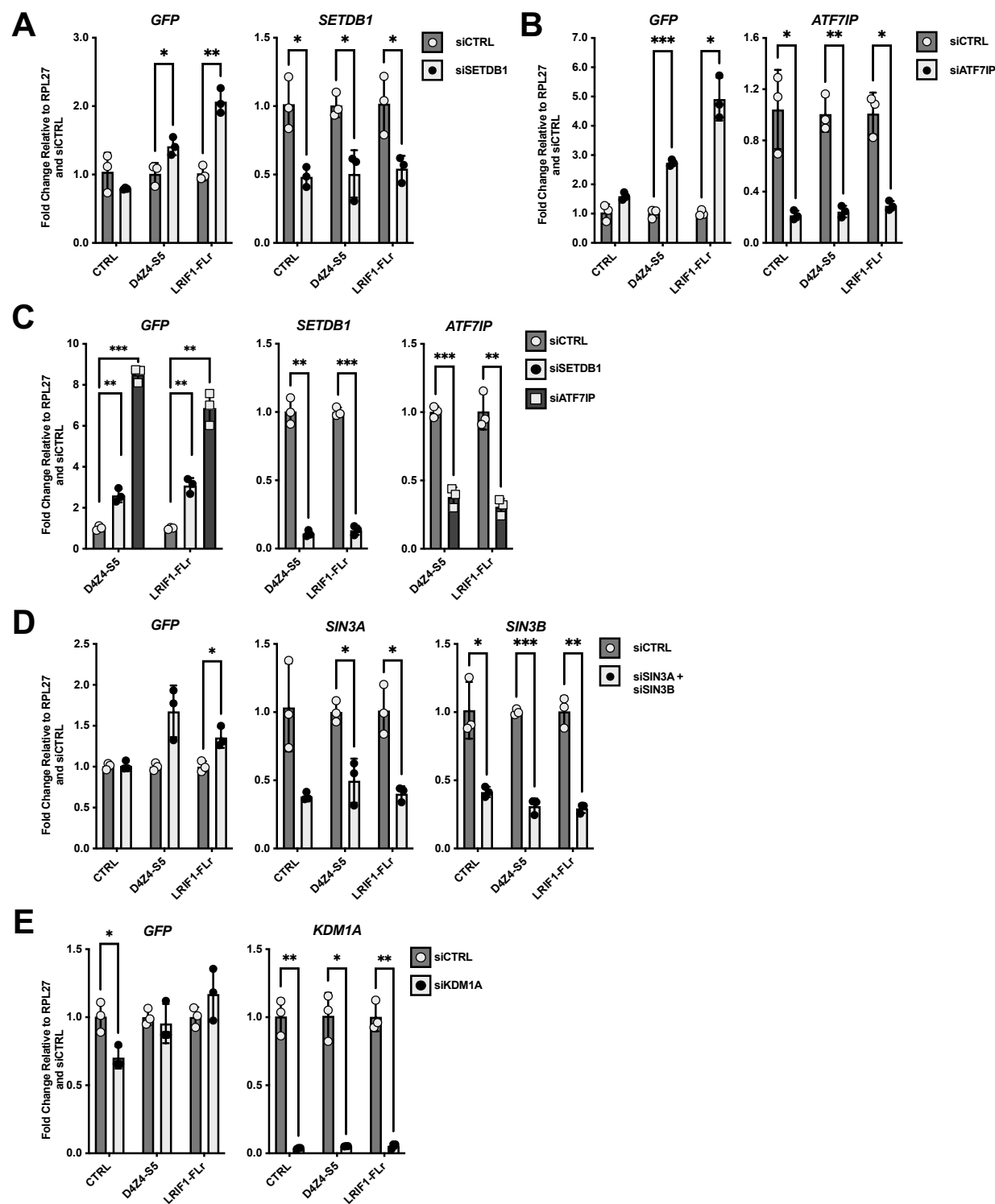
Supplemental Figure S2.5: (A, B) GFP fluorescence histograms of D4Z4-S5 and LRIF1-FLr cells at various timepoints after 5-aza-dC **(A)** or entinostat **(B)** treatment. Histograms are normalized to the mode of the population. Biological triplicate (n=3) histograms are presented for each treatment at each time point.

Supplementary Figure S2.6: Knockdown of MBD1, CHD4, or TRIM28 does not rescue GFP expression in D4Z4-S5 and LRIF1-FLr cells.



Supplementary Figure S2.6: (A, B, E) Left: Representative singleton GFP fluorescence histograms of CTRL, D4Z4-S5 and LRIF1-FLr cells treated with CTRL or CHD4 **(A)**, MBD1 **(B)**, or TRIM28 **(E)** siRNAs. Histograms are normalized to the mode of the population. Right Top: Fold change in GFP median fluorescence intensity (MFI) upon siRNA-mediated knockdown of CHD4 **(A)**, MBD1 **(B)**, or TRIM28 **(E)** compared to siCTRL. Data represent mean $\pm$ SD of biological replicates, $n=3$. Statistical significance was determined by Welch's t-test: * $p<0.05$, ** $p<0.01$, *** $p<0.001$. Right bottom: Percentage of GFP^{Low} cells in the population at various time points after siRNA-mediated knockdown of CHD4 **(A)**, MBD1 **(B)**, or TRIM28 **(E)** compared to siCTRL. Data represent mean $\pm$ SD of biological replicates, $n=3$. **(C, D, F)** RT-qPCR analysis of *GFP* (left) and *CHD4* **(C)**, *MBD1* **(D)** or *TRIM28* **(F)** (right) expression compared to housekeeping gene *RPL27* expression in siRNA-treated CTRL, D4Z4-S5, and LRIF1-FLr cells. Data plotted is fold change of expression in siCHD4- **(C)**, siMBD1- **(D)**, or siTRIM28- **(F)** treated cells compared to siCTRL-treated cells. Data represent mean fold change $\pm$ SD of biological replicates, $n=3$. Statistical significance was determined by Welch's t-test: * $p<0.05$, ** $p<0.01$.

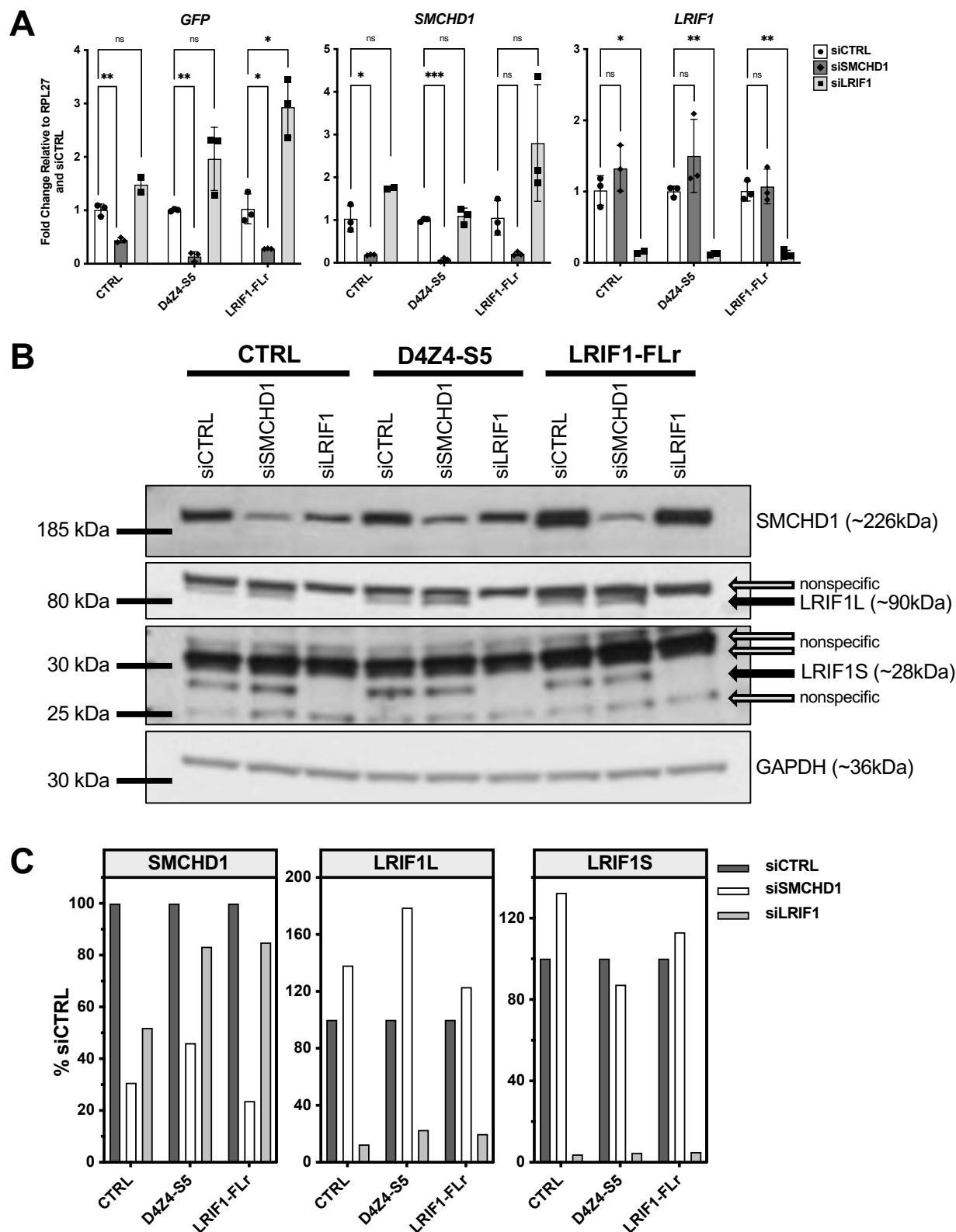
Supplementary Figure S2.7: RT-qPCR analysis confirms GFP expression changes and sufficient knockdown efficiency of identified epigenetic regulators.



Supplementary Figure S2.7: (A-E) RT-qPCR analysis of *GFP* (left) and *SETDB1* (A), *ATF7IP* (B), *SETDB1* and *ATF7IP* from H3K9me3 ChIP samples (C), *SIN3A* and *SIN3B* (D), or *KDM1A* (E) (right) expression compared to housekeeping gene *RPL27*

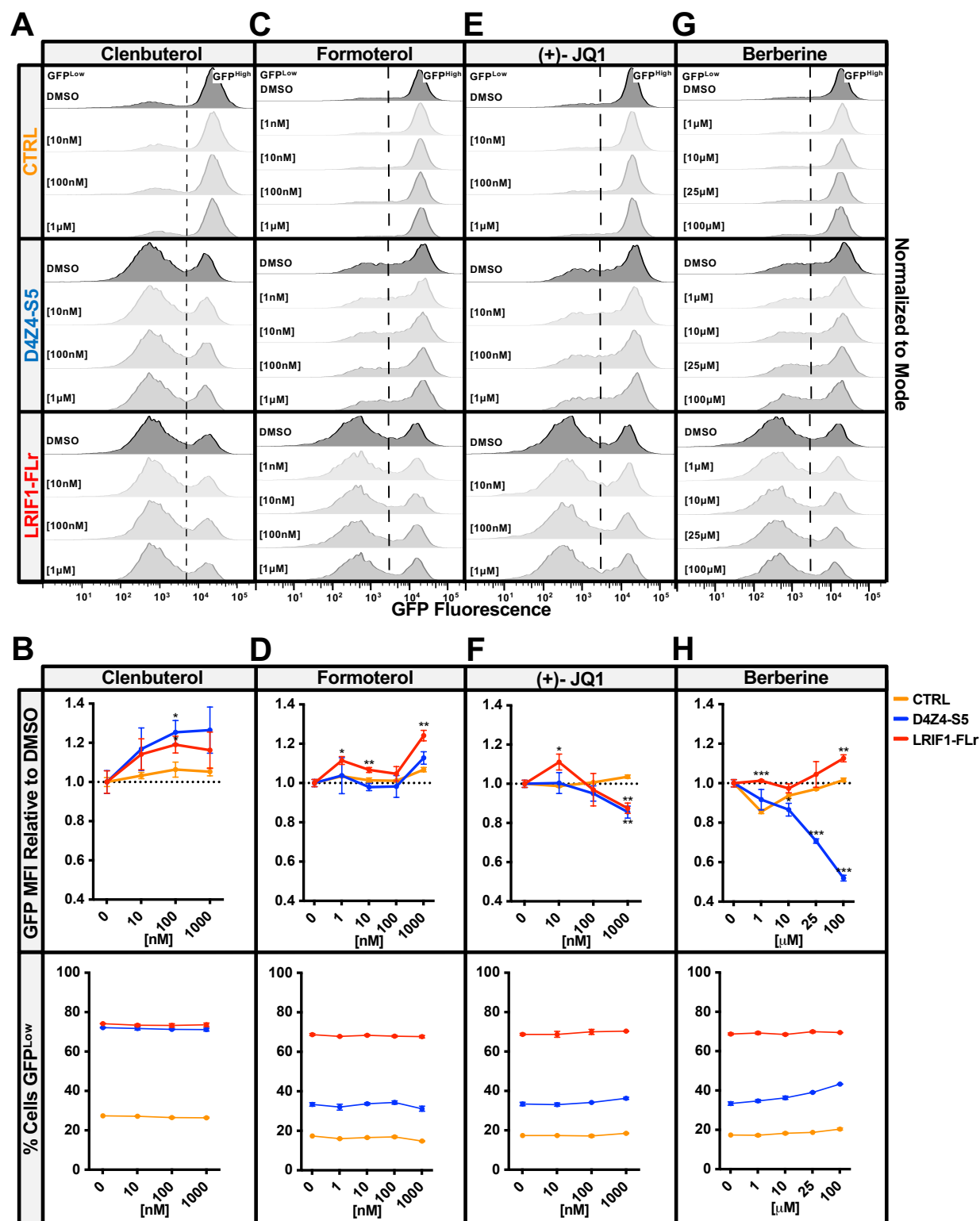
expression in siRNA-treated CTRL, D4Z4-S5, and LRIF1-FLr cells. Data plotted is fold change of expression in siSETDB1- **(A, C)**, siATF7IP- **(B, C)**, siSIN3A/B- **(D)**, or siKDM1A- **(E)** treated cells compared to siCTRL-treated cells. Data represent mean fold change $\pm$ SD of biological replicates, n=3. Statistical significance was determined by Welch's t-test: *p<0.05, **p<0.01, ***p<0.001.

Supplementary Figure S2.8: RT-qPCR and western blot analyses confirm changes in GFP expression and sufficient knockdown of SMCHD1 and LRIF1.



Supplementary Figure S2.8: (A) RT-qPCR analysis of *GFP* (left), *SMCHD1* (middle) and *LRIF1* (right) expression compared to housekeeping gene *RPL27* expression in siRNA-treated CTRL, D4Z4-S5, and LRIF1-FLr cells. Data plotted is fold change of expression in siSMCHD1- or siLRIF1-treated cells compared to siCTRL-treated cells. Data represent mean fold change $\pm$ SD of biological replicates, n=3. Statistical significance was determined by Welch's t-test: *p<0.05, **p<0.01, ***p<0.001. **(B)** Immunoblot analysis of SMCHD1 and LRIF1 protein levels in CTRL, D4Z4-S5, and LRIF1-FLr cells treated with siCTRL, siSMCHD1, or siLRIF1. Nonspecific band patterns mirror those shown in LRIF1 KO line validation in Šikrová et al., 2023. GAPDH serves as a loading control. **(C)** Quantification of SMCHD1 (left), LRIF1L (middle), and LRIF1 (right) levels in **(B)** using densitometry analysis normalized to GAPDH and graphed as percentage relative to siCTRL-treated samples.

Supplementary Figure S2.9: D4Z4-S5 and LRIF1-FLr show differential sensitivity to candidate FSHD therapeutics.



escalating dosages of candidate FSHD therapeutics clenbuterol (**A**), Formoterol (**C**), (+)- JQ1 (**E**), or Berberine (**G**). Histograms are normalized to the mode of the population. (**B, D, F, H**) Top: Fold change in GFP median fluorescence intensity (MFI) upon 3-day treatment with escalating dosages of clenbuterol (**B**), Formoterol (**D**), (+)- JQ1 (**F**), or Berberine (**H**) compared to DMSO treatment in CTRL, D4Z4-S5, and LRIF1-FLr cells. Data points represent mean fold change $\pm$ SD of biological replicates, n=3. Statistical significance was determined by Welch's t-test: *p<0.05, **p<0.01, ***p<0.001. Bottom: Percentage of GFP^{Low} cells in the population upon treatment with escalating dosages of clenbuterol (**B**), Formoterol (**D**), (+)- JQ1 (**F**), or Berberine (**H**). Data represent mean $\pm$ SD of biological replicates, n=3.

Tables

Table 2.1: Characteristics of AAVS1 construct insert sequences.

Sequence	#CpGs	%GC	Length (nt)	% CpG Bases
CTRL	7	43%	201	6.96%
LRIF1-FLr	37	54%	617	11.99%
LRIF1-FLf	37	54%	617	11.99%
LRIF1-S1r	14	66%	206	13.59%
LRIF1-S2r	16	66%	201	15.92%
LRIF1-S3r	4	46%	112	7.14%
LRIF1-S3f	4	46%	112	7.14%
LRIF1-S3f Minimal Sufficiency	2	38%	65	6.15%
LRIF1-S3fΔSite#1	4	46%	106	7.54%
LRIF1-S3fΔSite#2	3	47%	94	6.38%
LRIF1-S4r	12	48%	161	14.91%
LRIF1-S5r	5	46%	169	5.92%
D4Z4-S1	46	76%	428	21.50%
D4Z4-S2	34	74%	344	19.77%
D4Z4-S3	37	74%	310	23.87%
D4Z4-S4	35	68%	334	20.96%
D4Z4-S5	37	66%	401	18.45%
D4Z4-S5r	37	66%	401	18.45%
D4Z4-S6	46	70%	600	15.33%
D4Z4-S7	58	69%	600	19.33%
D4Z4-S8	33	70%	332	19.88%
D4Z4-S9	47	74%	447	21.03%
D4Z4-S10	56	78%	501	22.36%
D4Z4-S11	67	75%	675	19.85%
D4Z4-S12	25	68%	320	15.63%
D4Z4-S13	19	70%	265	14.34%
D4Z4-S14	54	75%	519	20.81%
D4Z4-S5.1	17	68%	156	21.79%
D4Z4-S5.2	12	64%	138	17.39%
D4Z4-S5.3	13	65%	164	15.85%
D4Z4-S5.4	11	62%	146	15.07%
D4Z4-S5.4(nt1-75)	6	64%	75	16.00%
D4Z4-S5.4(nt39-113)	8	68%	75	21.33%
D4Z4-S5.4(nt76-146)	5	59%	71	14.08%

Chapter 3: Using D4Z4-S5 and LRIF1-FLr reporter constructs for unbiased discovery of novel FSHD regulators

This chapter contains unpublished work.

Introduction

Screening for FSHD therapeutics is challenging, as disease pathogenesis is sporadic and variable in individuals and endogenous *DUX4* is expressed in only 0.1-1% of cells in patient myoblast lines, making RNA and protein detection difficult [9,62,63]. Though some success has been found in using artificial systems with a *DUX4* target gene promoter driving a reporter [67,111,138,139], the reliance on sporadic and variable patient myoblast expression of endogenous *DUX4* or artificial overexpression of *DUX4*, which has many downstream consequences on cellular physiology [110], may impact the efficacy of therapeutic screens. In **Chapter 2**, a silencing reporter system identified a key D4Z4 regulatory sequence, D4Z4-S5, that was able to recruit silencers to inhibit reporter expression in the absence of chromatin context. This reporter assay was successfully performed in HeLa cells to avoid the dynamic regulation of the D4Z4 during myogenic differentiation [118], indicating it can be used in other cell types. The ability of this D4Z4-S5 reporter system to assess drug effects on inhibition of *DUX4* activation in any cell type, without necessitating *DUX4* expression for a readout, makes it a valuable tool for large scale screening for FSHD therapeutics. This chapter uses the D4Z4-S5 reporter system for two different unbiased discovery screening approaches: enChIP for discovery of proteins interacting with the D4Z4-S5 locus and a kinase inhibitor screen in D4Z4-S5 reporter cells for discovery of novel regulatory pathways.

Results

enChIP-MS as an unbiased discovery tool to identify direct regulators of D4Z4-S5 and LRIF1-FLr

The identification in **Chapter 2** of several known D4Z4 regulators, such as SETDB1, ATF7IP, SIN3A/3B and LRIF1 in modulation of expression through the D4Z4-S5 reporter system led us to use an unbiased discovery approach, used previously to discover proteins bound to the whole D4Z4 [46], to uncover proteins bound to D4Z4-S5. We designed lentiviral constructs to perform engineered DNA-binding molecule-

mediated chromatin immunoprecipitation mass spectrometry (enChIP-MS) in the HeLa AAVS1 reporter system. A lentiviral construct containing a U6 promoter driving an sgRNA targeting the exogenous sequence just upstream of the AAVS1 region of interest (sgAAVS1) or a scrambled control (sgScram), a constitutive promoter driving 3xFLAG-tagged catalytically dead Cas9 (dCas9), and blasticidin resistance gene (BlastR) was transduced into HeLa CTRL, D4Z4-S5, and LRIF1-FL cells (**Fig. 3.1A**). The AAVS1 sgRNA should target dCas9 to any AAVS1 reporter construct insert, while the sgScram should not target any specific human genomic sequence and serves as a control for nonspecific binding (**Fig. 3.1A**). After transduction into AAVS1 reporter cells, ChIP-qPCR for FLAG (dCas9) was performed to ensure proper targeting. Off target binding was limited, as a gene desert region (GDR) showed little to no binding of dCas9 in the sgScram or sgAAVS1-targeted cells (**Fig. 3.1B**). ChIP-qPCR of the AAVS1 region revealed strong enrichment above IgG in two biological replicates of 3xFLAG-dCas9 in sgAAVS1-transduced CTRL, D4Z4-S5, and LRIF1-FLr cell lines, while sgScram-transduced cells did not show any enrichment above IgG (**Fig. 3.1C**). Western blotting of 3xFLAG-dCas9 revealed proper sizing (data not shown) and immunofluorescence staining for FLAG revealed nuclear localization of the 3xFLAG-dCas9 protein (**Fig. 3.1D**), indicating the system was operating properly for downstream analysis.

EnChIP was then performed using CTRL, D4Z4-S5, and LRIF1-FLr cells transduced with sgAAVS1-3xFLAG-dCas9. The targeted FLAG-tagged dCas9 was crosslinked to locus associated factors, DNA was fragmented, FLAG was immunoprecipitated to pull down AAVS1-locus associated factors, proteins were de-crosslinked from DNA, and peptides were identified via mass spectrometry (**Fig. 3.2A**). Generally, peptide counts via mass spectrometry were low, though all identified in this chapter met the 1% false discovery rate (FDR) cutoff. Approximately 20% of proteins identified in the D4Z4-S5 enChIP-MS overlapped with previous whole D4Z4 enChIP [46] and/or direct interactors with SMCHD1 via co-immunoprecipitation [53] (**Fig. 3.2B**). Gene ontology (GO) biological process analysis revealed many processes associated with RNA splicing and processing, though DNA conformational change and chromosome organization were among the top hits (**Fig. 3.2C**). Several proteins involved in gene silencing were identified in the D4Z4-S5 enChIP-MS, such as protein SUMOylation factors, silencer complex recruiter TRIM28, heterochromatin regulators, chromatin remodeling factors, and NuRD complex factors (**Fig. 3.2D**). The LRIF1-FL enChIP also had extremely low peptide counts but revealed high overlap with D4Z4-bound proteins and SMCHD1 interactors, with ~15% of MS hits overlapping with one or both (**Fig. 3.3A**). GO analysis of LRIF1-FL-bound proteins also revealed a majority of terms relating to mRNA processing, with only one of the top 15 being negative regulation of gene expression (**Fig. 3.3B**). Looking at all proteins with regards to gene silencing, several proteins overlap with those found in the D4Z4-S5 enChIP, and other factors involved in transcription corepressor activity, suppressive ubiquitin ligase

activity, and heterochromatin regulation were identified (**Fig 3.3C**). Ultimately, low peptide counts limited the efficacy of enChIP-MS analysis for D4Z4-S5 and LRIF1-FL, though the overlap in factors between several previous studies of D4Z4 and SMCHD1 interactors is promising for the notion of D4Z4-S5 and the *LRIF1* promoter as important regulatory elements in FSHD biology.

A live cell imaging kinase inhibitor screen confirms p38 inhibitors as regulators of DUX4 expression.

The encouraging response of the D4Z4-S5 silencing assay system to known DUX4 inhibitory drugs, especially p38 inhibitors (see **Chapter 2**) led us to use D4Z4-S5 reporter cells as a screening platform to identify new regulatory pathways and FSHD therapeutics. With the help of our collaborator at FHCC, Taran Gujral, we used a panel of kinase inhibitors, designed to span the cellular kinome and exploit the known promiscuity, or polypharmacology, of kinase inhibitors on multiple targets [140–142] for regulatory pathway deconvolution. CTRL, D4Z4-S5, and LRIF1-FL cells were treated with previously identified DUX4 inhibitory drugs (**Suppl. Fig. S2.9**) and an optimized kinase panel (**Table 3.1**), following the same live cell imaging scheme as in the earlier losmapimod trial (**Fig. 2.6E, Fig. 3.4A**). Cells were treated with the kinase inhibitor panel for 6 days, with one media change at day 3. The relative GFP confluency was then calculated and compared to DMSO and the t0 time point (**Fig. 3.4A**). A heat map of the ratio of GFP+/Total cell confluency relative to t0 and DMSO shows that DUX4-suppressive p38 inhibitors and berberine are effective in further inhibition of GFP in D4Z4-S5 (**Fig 3.4B**). In analysis of DUX4 inhibitors, p38 inhibitors showed the largest reduction in relative GFP+ cell confluency over 144 hours (**Fig. 3.5 A-D**). Other candidate DUX4 inhibitors showed similar phenotypes to those shown in **Suppl. Fig. S2.9**), with berberine showing suppression in D4Z4-S5 and LRIF1-FL and other drugs not displaying a significant impact on GFP expression (**Fig 3.5 E-H**).

Other compounds in the kinase inhibitor screen showed inhibitory effects on relative GFP+ confluency. BCR-ABL inhibitor GZD824, RTK and P1-3K inhibitor PP121, and PDGFR, VEGFR, and FGFR inhibitor Vargatef displayed specific inhibition of D4Z4-S5, while ALK and IGFR inhibitor AZD3463 inhibited relative GFP+ cell confluency in both D4Z4-S5 and LRIF1-FLr cells (**Fig. 3.6A-D**). After measuring the phenotypic effects of each kinase inhibitor involved in the kinase inhibitor panel, computational deconvolution of the overlapping targets and pathways were used to identify operative kinase targets for further validation, known as Kinase Inhibitor Regularization (KiR) [140,141]. KiR analysis, using the values from the 48- and 120-hour GFP+ cell confluency Incucyte data, identified several key kinases specific to D4Z4-S5 cells, shown in purple in **Fig 3.6E**. These KiR generated networks can be used to identify and explore novel regulatory pathways of D4Z4-mediated expression of DUX4.

Discussion

Though great strides have been made in the identification of epigenetic regulators of the D4Z4 [9,34,35,46,47,49–53,110,112–115], the mechanisms of DUX4 suppression in somatic cells are not fully elucidated. A number of pathways and drugs have been identified as potentially important regulators of DUX4 expression [64–70], though there is still no cure identified for FSHD [109]. The D4Z4-S5 reporter (see **Chapter 2**) provides a simple tool to screen new FSHD therapeutics that epigenetically silence *DUX4* as well as screen for new modifiers that interact with that important regulatory element.

Previous work in our lab successfully performed enChIP-MS by immunoprecipitation of the entire D4Z4 and subsequent identification of bound proteins [46]. Applying this technique to D4Z4-S5 was meant to narrow the focus into proteins that directly interact with this key regulatory sequence, to mixed success. The enChIP system did appear to work correctly in quality control checks (**Fig. 3.1**), but peptide counts were extremely low in the mass spectrometry results, with many identified proteins only having coverage of one peptide in just one of the two replicates, though all hits were significant enough to qualify for the >1% FDR threshold. Though many of these identified proteins did, promisingly, overlap with previous studies, most proteins were also present in the CTRL sample, indicating a lack of specificity for D4Z4-S5 due to nonspecific recruitment to the exogenous AAVS1 construct. One possible improvement if this method that could increase specificity and yield would be usage of nanobodies instead of anti-FLAG antibodies, which bind their targets more strongly and allow for more stringent washing of nonspecific interactions [143]. Further optimization of the enChIP protocol and better peptide coverage will allow for proper identification of the important interactors seeding silencing effects through D4Z4-S5.

To assess the efficacy of this reporter system as a platform for future drug screens, we sought to determine if other readouts besides flow cytometry could be used for detecting subtle changes in GFP expression. Live cell imaging is particularly useful for drug screening, as it allows for screening of many chemicals simultaneously in a multi-well plate format in addition to consistent monitoring over multiple days, rather than endpoints requiring harvest and processing. In **Chapter 2**, live cell imaging of D4Z4-S5 and LRIF1-FL cells treated with p38 inhibitors reflected flow cytometry experiments, where GFP levels were further decreased upon treatment. To ensure that the effects of candidate DUX4-inhibitor drugs on GFP expression measured via flow cytometry in **Suppl. Fig. S2.9** were also consistent in the live cell imaging system, clenbuterol, formoterol, (+)- JQ1, and berberine were assessed via live cell imaging. Promisingly, live cell imaging of these candidate therapeutics showed similar trends seen in the flow cytometry experiments, indicating the live cell imaging system is robust enough for larger scale screens.

Many different regulatory pathways have been implicated in regulation of DUX4 expression, though the mechanisms of action for many, including the p38 pathway, remain an area of active study. By uncovering pathways that regulate expression specifically through D4Z4-S5, we hope to narrow the focus for therapeutic discovery as well as begin to elucidate regulatory mechanisms of action. Taran Gujral, a collaborator at FHCC, explores the use of polypharmacology for discovery of key kinases and pathways involved in a particular cellular phenotype [140,141]. Kinase inhibitors have varying levels of promiscuity in their inhibitory activity, often inhibiting activity of other kinases to varying degrees in addition to their target kinase(s) [142]. Exploiting this promiscuity, a comprehensive panel of kinase inhibitors was created to inhibit most kinases to varying degrees [141]. By incorporating known *in vitro* inhibition rates of each kinase inhibitor on all known kinases, the phenotypic effect of each panel kinase inhibitor can be deconvoluted to identify key kinases for that phenotype, known as kinase inhibitor regularization (KiR) analysis [140,141]. We treated CTRL, D4Z4-S5, and LRIF1-FLr cells with this panel of kinases and measured effects of each kinase on GFP+ cell confluency. Before KiR analysis, visualization of GFP+ cell confluency levels based on kinase inhibitor treatment revealed several kinase inhibitors that specifically inhibited D4Z4-S5 and/or LRIF1-FL more than CTRL: GZD824, AZD3463, PP121, and Vargatef, which may be good candidates for future FSHD therapeutic study. KiR analysis identified several key kinases for GFP modulation through D4Z4-S5, including MAPK9 (JNK2), FGFR1, VEGFR, RAF1, and TEK. The expanded STRING network from those key kinases also contained p38 (MAPK14), indicating the network aligns with known D4Z4-S5 regulators that have been identified as modulators of D4Z4-S5. Further exploration of these key kinases and their impact on D4Z4-S5 regulation will provide new avenues for FSHD therapeutic exploration.

Materials and Methods

Cell culture

HeLa cells were grown in DMEM High Glucose with L-glutamine (Gibco, #11965092), supplemented with 10% fetal bovine serum (FBS) (Cytiva Hyclone, #SH30396.03) and 1% penicillin/streptomycin (Gibco, #15140122) and incubated at 37°C supplemented with 5% CO₂. Cells were passaged by treatment with 0.25% trypsin- Ethylenediaminetetraacetic acid (EDTA) (Gibco, #25200056) for 5 minutes at 37°C followed by growth media addition and subculturing.

enChIP construct cloning

The pZLCv2-3XFLAG-dCas9-HA-2XNLS (Addgene, #106357) vector was modified to replace puromycin resistance with blasticidin resistance. Overlapping fragments for the P2A, Blasticidin S deaminase gene, and PRE sequence were amplified via PCR from the pZLCv2 plasmid and the pCW57-MCS1-P2A-MCS2 (Blast) vector (Addgene, #80921), purified, and cloned into the PmeI (NEB, #R0560S)- and

BamHI-HF (NEB, #R3136L)- digested backbone via HiFi DNA Assembly according to manufacturer's instructions. The pZLCv2-3XFLAG-dCas9-HA-2XNLS (Blast) plasmid was verified via Sanger sequencing. Five sgRNAs targeting the AAVS1 insert sequence were designed using the IDT Custom Alt-R CRISPR-Cas9 guide RNA design program. The backbone was digested with BsmBI-v2 (NEB, #R0580) and annealed oligos (IDT) containing the sgRNA sequence were cloned into the BsmBI site via T4 ligation. Verified enChIP plasmid, psPAX2 (Addgene, #12260), and pMD2.G (Addgene, #12259) were transfected into 293T cells with Lipofectamine 3000 as described above to produce lentiviral particles. AAVS1 knock-in cells were then transduced with lentivirus in the presence of 10ug/mL polybrene. Cells were selected for lentiviral integration with growth media supplemented with 3ug/mL blasticidin S.

ChIP-qPCR

ChIP-qPCR was adapted from previously published protocols [136,137]. HeLa AAVS1 cells were trypsinized, washed with 1xPBS, and fixed in 1% Formaldehyde Ultrapure EM grade (EMS, #15710) for 10 minutes at room temperature with gentle rocking. Formaldehyde was quenched with 125mM glycine, cells were washed with PBS and then lysed in Lysis Buffer (Protease and phosphatase inhibitors (Pierce, #A32961), 1% Na-deoxycholate, 1% SDS, 5mM EDTA pH 8, and 50mM Tris pH 8 in 1xPBS). Lysate was sonicated in Biorupter Plus TPX Polymethyl pentene tubes (Diagenode, C30010010) using a Biorupter Plus Sonication System (Diagenode, B01020001) until chromatin was sheared to a range of 200-600bp; a small aliquot of sheared chromatin was incubated with RNase A at 37°C for one hour, treated with Proteinase K overnight at 65°C, and purified using a NucleoSpin Gel and PCR cleanup mini kit (Machery-Nagel, #740609) before concentration was determined using a Qubit and chromatin was run on a 2% agarose gel in Tris/Borate/EDTA (TBE) buffer to assess fragment size. Chromatin was centrifuged at 16,000xg for 10 min at 4°C to clear insoluble fraction, and 10% of the chromatin was reserved for input DNA.

For immunoprecipitation, chromatin was diluted in Dilution Buffer (1% Triton X-100, 150mM NaCl, 2mM EDTA pH 8, and 20mM Tris pH 8 in PBS) to reduce SDS concentration to <0.1%. For ChIP, 5µg of antibody (Mouse Anti-FLAG [M2] Sigma #F3165 or Mouse IgG isotype control Invitrogen #31903) was added to 30µg of chromatin and incubated with rotation overnight at 4°C. Agarose Fast Flow Prot A (Millipore Sigma #16-156) beads were pre-blocked overnight at 4°C in Dilution Buffer + 2% bovine serum albumin (BSA). Samples were incubated with beads at 4°C for 2 hours, and washed with Wash Buffer I (protease and phosphatase inhibitors (Pierce, #A32961), 0.1% SDS, 1% Triton X-100, 2mM EDTA pH 8, 20mM Tris pH 8, 150mM NaCl), washed with Wash Buffer II (protease and phosphatase inhibitors, 0.1% SDS, 1% Triton X-100, 2mM EDTA pH 8, 20mM Tris pH 8, 500mM NaCl), washed with Wash Buffer III (protease and phosphatase inhibitors, 0.25M LiCl, 1% NP-40, 1% Na-deoxycholate, 1mM EDTA pH 8, 10mM Tris pH 8, 500mM NaCl), and washed twice with ice cold PBS. Beads and attached chromatin were then de-crosslinked by resuspension in Ultrapure water with 1% SDS and 0.1M NaHCO₃, incubated with RNase A (Thermo Scientific, # R1253) for 1 hour at 37°C, and incubated overnight at 65°C with Proteinase K (Thermo Scientific, #EO0491). IP DNA was then purified using NTB Buffer (Machery-Nagel, #740595.150) and the NucleoSpin Gel and PCR cleanup mini kit (Machery-

Nagel, #740609). Quantitative PCR of IP DNA was run on a QuantStudio 7 Flex (Applied Biosystem) using iTaq SYBR Green Supermix (Bio-Rad, #1725124).

Immunofluorescence microscopy

Cells were fixed with 4% paraformaldehyde for 10 minutes and permeabilized with 0.5% TritonX-100. Samples were incubated with primary antibodies (mouse anti-FLAG Sigma# F1804 and Anti-GFP [EPR14104] rabbit monoclonal antibody, Abcam#ab183734) at 4°C overnight, followed by incubation with fluorescently conjugated secondary antibodies (Rhodamine (TRITC) Donkey Anti-Mouse IgG Jackson ImmunoResearch #715-025-151 and Fluorescein (FITC) Donkey Anti-Rabbit IgG Jackson ImmunoResearch #711-095-152) for 1 hour at room temperature, and counterstained with DAPI (Sigma #D-9542). Plates were imaged using an immersion lens, Zeiss Axiophot fluorescent microscope, AxioCam MRc digital camera, and AxioVision 4.6 software.

enChIP (engineered DNA-binding molecule-mediated chromatin immunoprecipitation)

enChIP was performed according to a modified protocol from Campbell et al., 2018 [46] and Fujita and Fujii, 2013 [144]. HeLa AAVS1 cells were trypsinized, washed with 1xPBS, and fixed in 1% Formaldehyde Ultrapure EM grade (EMS, #15710) for 10 minutes at room temperature with gentle rocking. Formaldehyde was quenched with 125mM glycine, cells were washed with PBS and then lysed in IP Buffer (Protease and phosphatase inhibitors (Pierce, #A32961), 150mM NaCl, 1% Triton X-100, 0.5% NP-40, 5mM EDTA pH 8, and 50mM Tris-HCL pH 7.5 in water).

Lysate was sonicated in Biorupter Plus TPX Polymethyl pentene tubes (Diagenode, C30010010) using a Biorupter Plus Sonication System (Diagenode, B01020001) until chromatin was sheared to a range of 200-1500bp; a small aliquot of sheared chromatin was incubated with RNase A at 37°C for one hour, treated with Proteinase K overnight at 65°C, and purified using a NucleoSpin Gel and PCR cleanup mini kit (Machery-Nagel, #740609) before concentration was determined using a Qubit and chromatin was run on a 2% agarose gel in Tris/Borate/EDTA (TBE) buffer to assess fragment size. Chromatin was centrifuged at 12,000xg for 10 min at 4°C to clear insoluble fraction, and 10% of the chromatin was reserved for input DNA.

For pre-clearing, chromatin was diluted in IP Dilution Buffer (1% Triton X-100, 150mM NaCl, 2mM EDTA pH 8, 0.01% SDS and 20mM Tris pH 8 in water) to reduce SDS concentration to <0.2%. Antibody slurry was generated per sample, mixing 25µg mouse IgG isotype control into 200µL 0.1% PBST. Storage solution was removed from 100uL Dynabeads Protein G per sample and added to antibody slurry before incubation with rotation for 10 minutes at RT. Beads were washed once with 0.1% PBST, resuspended in 100uL 0.1% PBST per sample, and added to sonicated material before incubation with rotation for 2 hours at 4°C.

For immunoprecipitation, storage buffer was removed from 150uL Protein G Dynabeads and washed twice with 0.1M sodium phosphate buffer pH 7.0. 15µL FLAG antibody (Sigma #F3165) and 15µL 0.1M sodium phosphate buffer were mixed for each sample, then total mixture was added to beads before incubation for 25min at RT with rotation. Beads were concentrated and washed twice with 0.1M sodium phosphate buffer with 0.01% Tween-20. Beads were washed twice with 0.2M triethanolamine pH

8.2 and resuspended in 20mM DMP in 0.2M triethanolamine pH 8.2 before incubation with rotation for 30 minutes at RT. Beads were concentrated and incubated with rotation at RT for 15 minutes with 50mM Tris-HCl pH 7.5. Beads were washed three times in 0.1% PBST and resuspended in 150uL 0.1% PBST per sample. Supernatant from pre-clearing IgG beads was mixed with crosslinked beads for each sample and immunoprecipitated at 4°C overnight.

To elute from beads, beads were concentrated and washed 4 times with cold Low Salt Wash buffer (protease and phosphatase inhibitors (Pierce, #A32961), 20mM Tris pH 8.0, 2mM EDTA, 150mM NaCl, 1% Triton X-100, and 0.1% SDS in water), washed twice with cold IP Wash Buffer 2 (protease and phosphatase inhibitors (Pierce, #A32961), 10mM Tris pH 8.0, 1mM EDTA, 0.25M LiCl, 1% NP-40/Igepal, and 1% Na-deoxycholate in water), washed once with cold TE with protease and phosphatase inhibitors (Pierce, #A32961), and washed once with cold TBS + 0.1% NP-40 with protease and phosphatase inhibitors (Pierce, #A32961) before resuspension in Elution Buffer and incubation at 37°C for 20 minutes with agitation. Eluted complexes were precipitated by adding 1mL isopropanol, 150uL 3M sodium acetate, and 6.6uL of 15mg/mL GlycoBlue before storage at -20°C overnight.

To run on SDS-PAGE gel, samples were centrifuged at 17,000xg for 30 minutes at 4°C, pellets were washed with 70% EtOH and centrifuged at 13,000xg for 5 minutes at RT, and pellets were resuspended in 25uL 4X Sample Buffer (500mM Tris-HCl pH 6.8, 8% SDS, 20% BME, bromophenol blue, and 30% glycerol) before incubation at 98°C for 30 minutes. Samples were immediately run one cm into an SDS-PAGE gel and cut out for submission to the Proteomics & Metabolomics Shared Resource of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704). Samples were prepped and desalted via ZipTip before analysis on an Orbitrap Ascend Mass Spectrometer (Thermo Fisher). The raw mass spectrometry data files were searched with Proteome Discoverer version 3.1 against a Uniprot Human database, custom protein sequences from enChIP and AAVS1 constructs, and common contaminants. The search results were filtered for high confidence, 1% FDR at the peptide level.

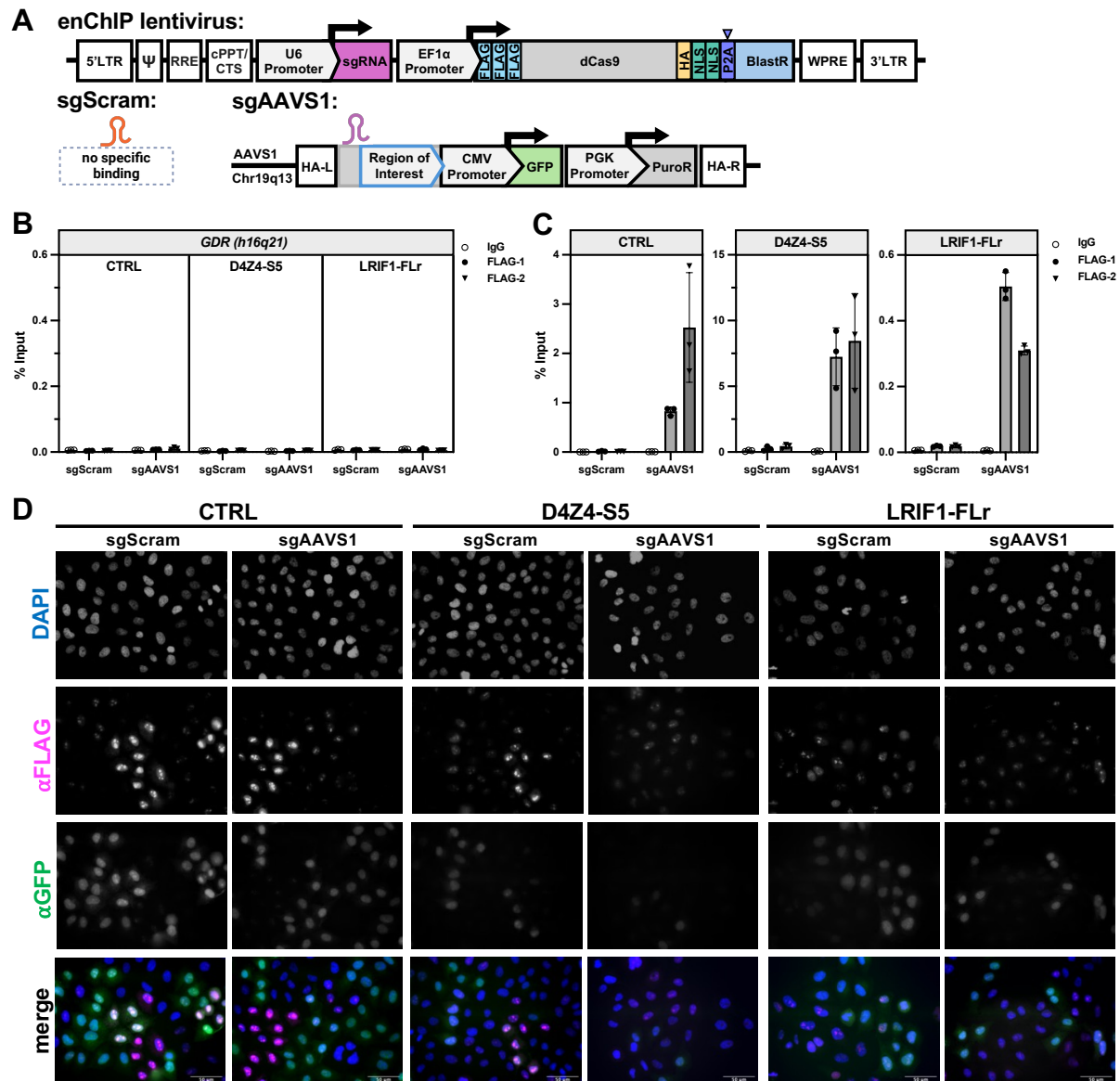
Incucyte tracking of GFP+ cell confluence

HeLa AAVS1 lines were seeded at 80% confluency and 24 hours later were treated with 5-aza-2'-deoxycytidine (Millipore Sigma #189825) for 12 hours. Cells were then incubated with standard media for 12 hours; this was repeated twice more, for a total of 36 hours of 5-aza-2'-deoxycytidine treatment in 72 hour, GFP+ cells were then sorted by FACS at a density of 10,000 cells per well in 96-well plates with 100uL media per well. 24 hours later, 100uL of media with 2x concentration of kinase inhibitors listed in **Table 3.1** or DUX4 inhibitors losmapimod (Sigma Aldrich, #SML3596), SB203580 (Selleckchem, #S1076), clenbuterol (Sigma Aldrich, #C5423), formoterol (Sigma-Aldrich, #F9552), (+)- JQ1 (Cayman Chemical, #11187-1), or berberine (Sigma Aldrich, #B3251). 96-well plates were placed in the Incucyte S3 Live Cell Analysis system, kept at 37°C with 5% CO₂. Wells were imaged for GFP+ and brightfield cell confluency, cell count, and GFP Mean Fluorescence intensity every 2 hours for 6 days. On day 3, plates were removed from the Incucyte between images and media was changed and supplemented with fresh drug. GFP+ cell confluency per well was normalized to t0 for

that well and to the average of DMSO treated wells to analyze differences in inhibitor-treated wells between cell lines. Kinase Inhibitor Regularization (KiR) analysis was performed using previously established R scripts [140,141].

Figures

Figure 3.1: Design and validation of AAVS1-targeting enChIP system.



Supplementary Figure S3.1: (A) Top: Schematic of enChIP lentiviral construct encoding an sgRNA, 3xFLAG-tagged catalytically dead Cas9 (dCas9) and blasticidin resistance. Bottom: Target sites for sgRNAs used. sgScramble (sgScram) should not specifically target any human genomic sequence and sgAAVS1 targets the sequence directly upstream of the sequence of interest in the exogenous AAVS1 reporter construct. LTR= long terminal repeat, Ψ = HIV-1 psi packaging element, RRE = HIV-1 rev response element, cPPT/CTS = central polypurine tract and central termination sequence of HIV-1, sgRNA = specific guide RNA, HA = human influenza hemagglutinin epitope tag, NLS = nuclear localization signal of SV40 large T antigen, P2A = 2A peptide from porcine teschovirus-1 polyprotein, BlastR = blasticidin resistance, WPRE = Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element, HA-L/R = AAVS1

homology arm (left or right). **(B, C)** ChIP-qPCR of 3xFLAG-tagged dCas9 and IgG binding to a non-specific gene desert region (*GDR h16q21*) **(B)** and the AAVS1 site **(C)** in AAVS1 reporter cell lines transduced with either sgScram or sgAAVS1 enChIP constructs. N=3 technical replicates. FLAG-1 and FLAG-2 represent two biological replicates. **(D)** Immunofluorescence microscopy to determine localization of 3xFLAG-dCas9 (α FLAG) and GFP (α GFP) in AAVS1 reporter lines with various enChIP constructs co-stained with a DAPI nuclear stain. Bar = 50 μ m.

Figure 3.2: enChIP-MS analysis confirms binding of many previously identified D4Z4-regulatory proteins to D4Z4-S5.

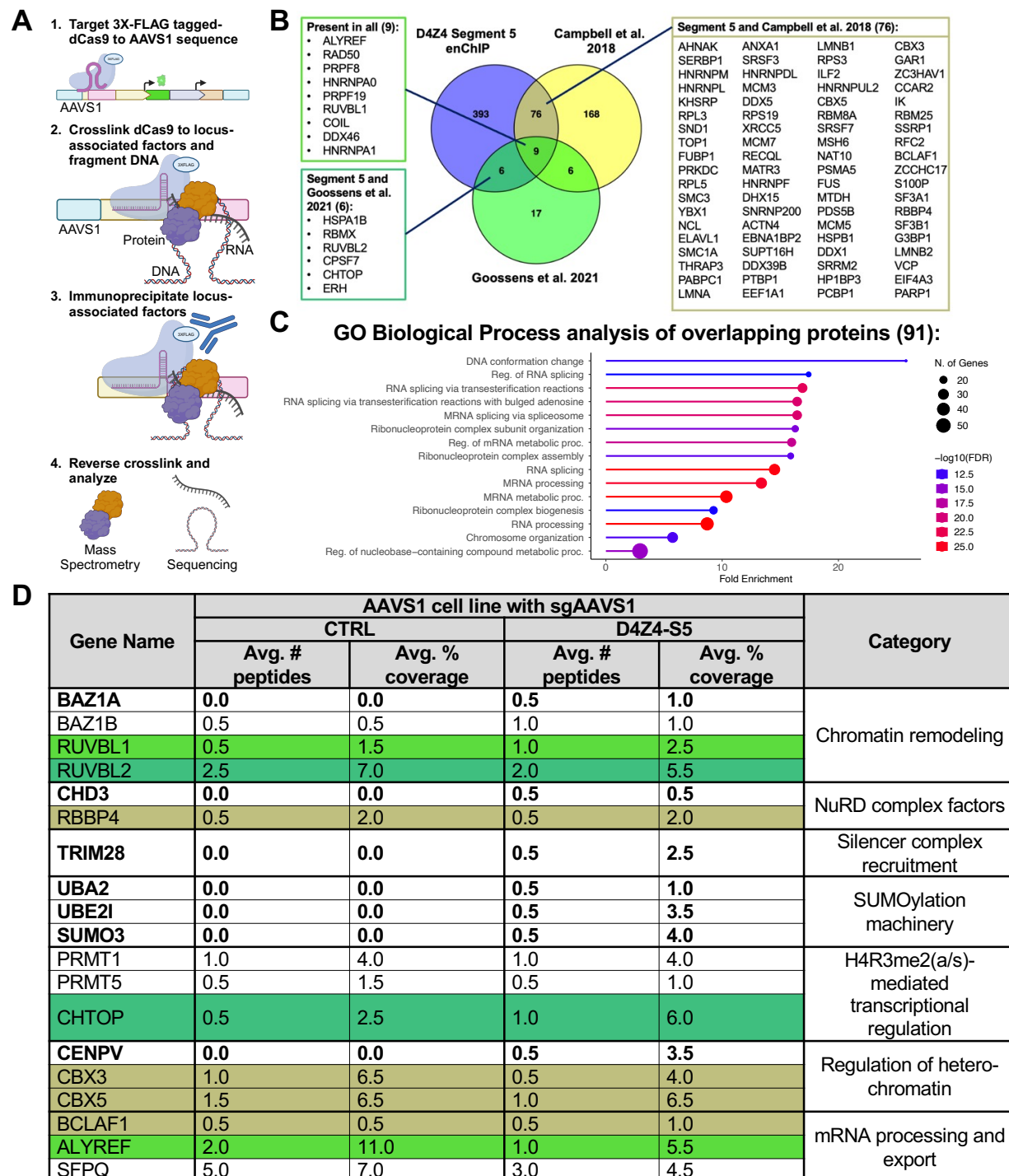


Figure 3.2: (A) Schematic of enChIP workflow for detection of proteins bound to the exogenous AAVS1 sequence. **(B)** Venn diagram of overlapping mass spectrometry proteomics hits from this study in D4Z4-S5, whole D4Z4 enChIP (Campbell et al., 2018) [46], and immunoprecipitated SMCHD1 interacting proteins (Goossens et al., 2021)

[53]. **(C)** Gene ontology (GO) biological process enrichment analysis of overlapping proteins from **(B)**. Graph created using the ShinyGO tool [145]. **(D)** Table of selected proteins, peptide coverage, and functions from enChIP-MS proteomics analysis of AAVS1-bound proteins in CTRL D4Z4-S5 reporter cells. Colors reflect Venn diagram binning from **(B)**. Bold indicates peptides were not detected in CTRL reporter line enChIP-MS analysis.

Figure 3.3: enChIP-MS analysis also confirms binding of many previously identified D4Z4-regulatory proteins to LRIF1-FLr.

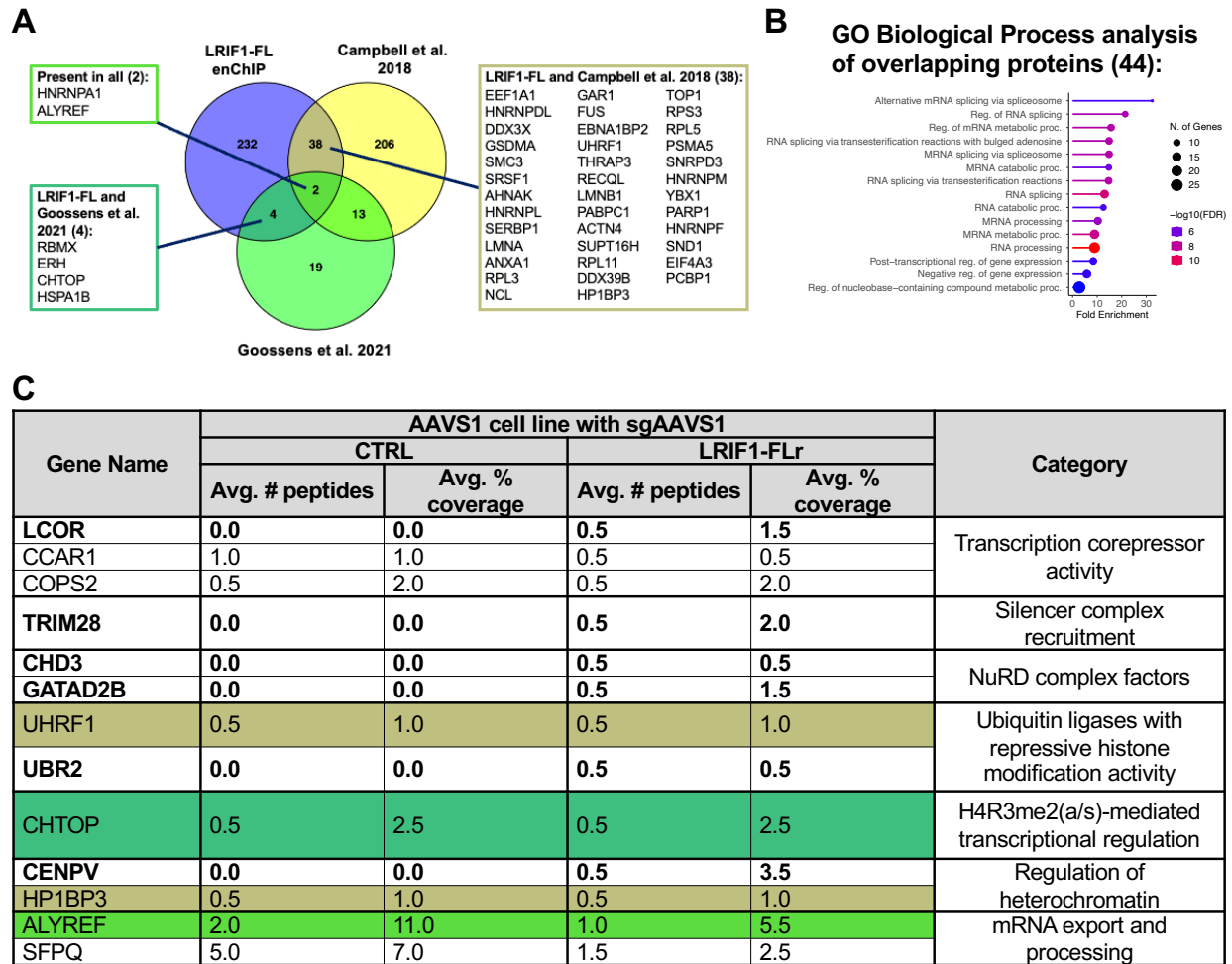


Figure 3.3: (A) Venn diagram of overlapping mass spectrometry proteomics hits from this study in LRIF1-FLr, whole D4Z4 enChIP (Campbell et al., 2018) [46], and immunoprecipitated SMCHD1 interacting proteins (Goossens et al., 2021) [53]. **(B)** Gene ontology (GO) biological process enrichment analysis of overlapping proteins from **(A)**. Graph created using the ShinyGO tool [145]. **(C)** Table of selected proteins, peptide coverage, and functions from enChIP-MS proteomics analysis of AAVS1-bound proteins in CTRL and LRIF1-FLr reporter cells. Colors reflect Venn diagram binning from **(A)**. Bold indicates peptides were not detected in CTRL reporter line enChIP-MS analysis.

Figure 3.4: Live-cell imaging kinase inhibitor screen of GFP fluorescence in CTRL, D4Z4-S5, and LRIF1-FL cells.

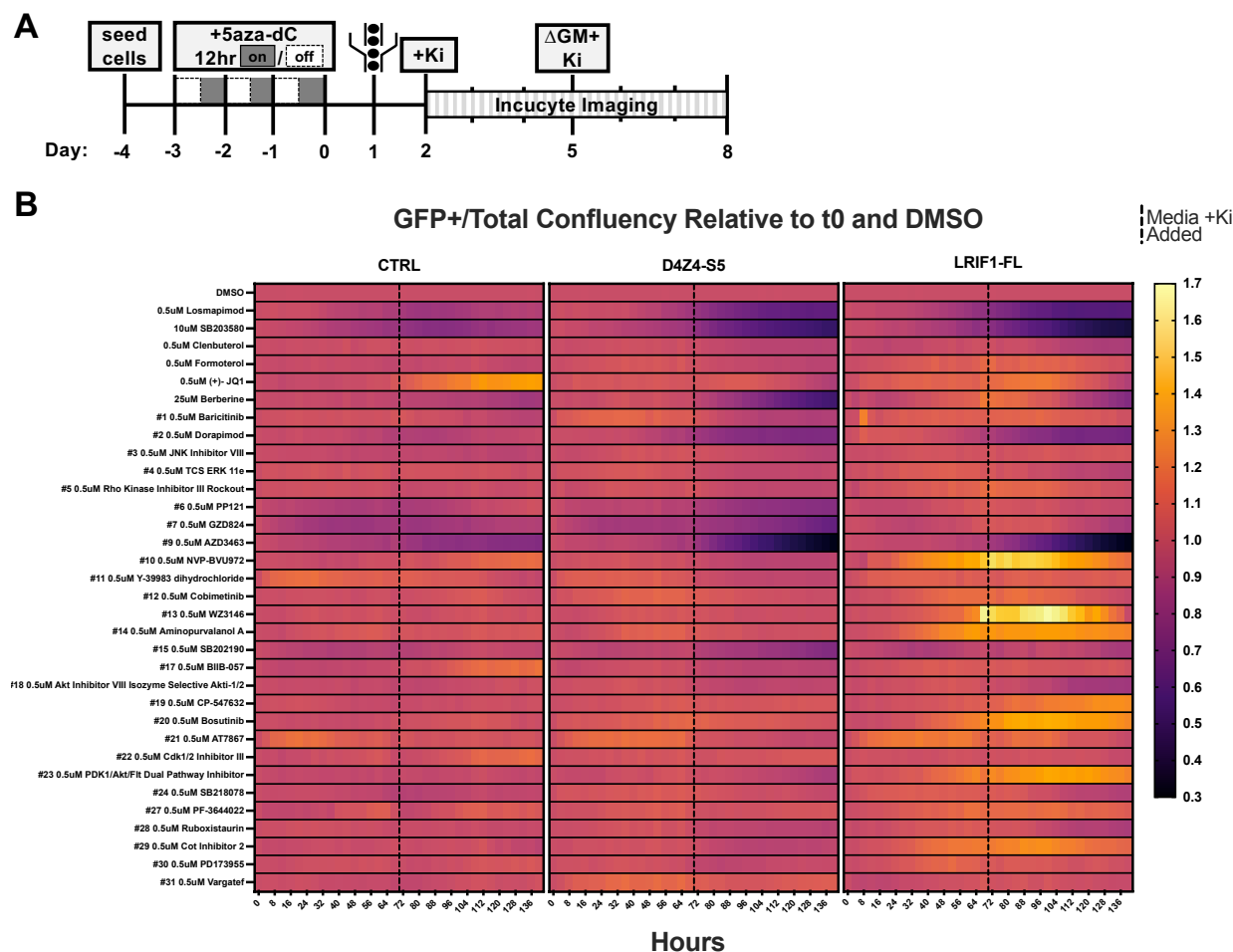


Figure 3.5: Live-cell imaging GFP confluency analysis of candidate *DUX4* inhibitors in AAVS1 reporter lines reflects previous flow cytometry results.

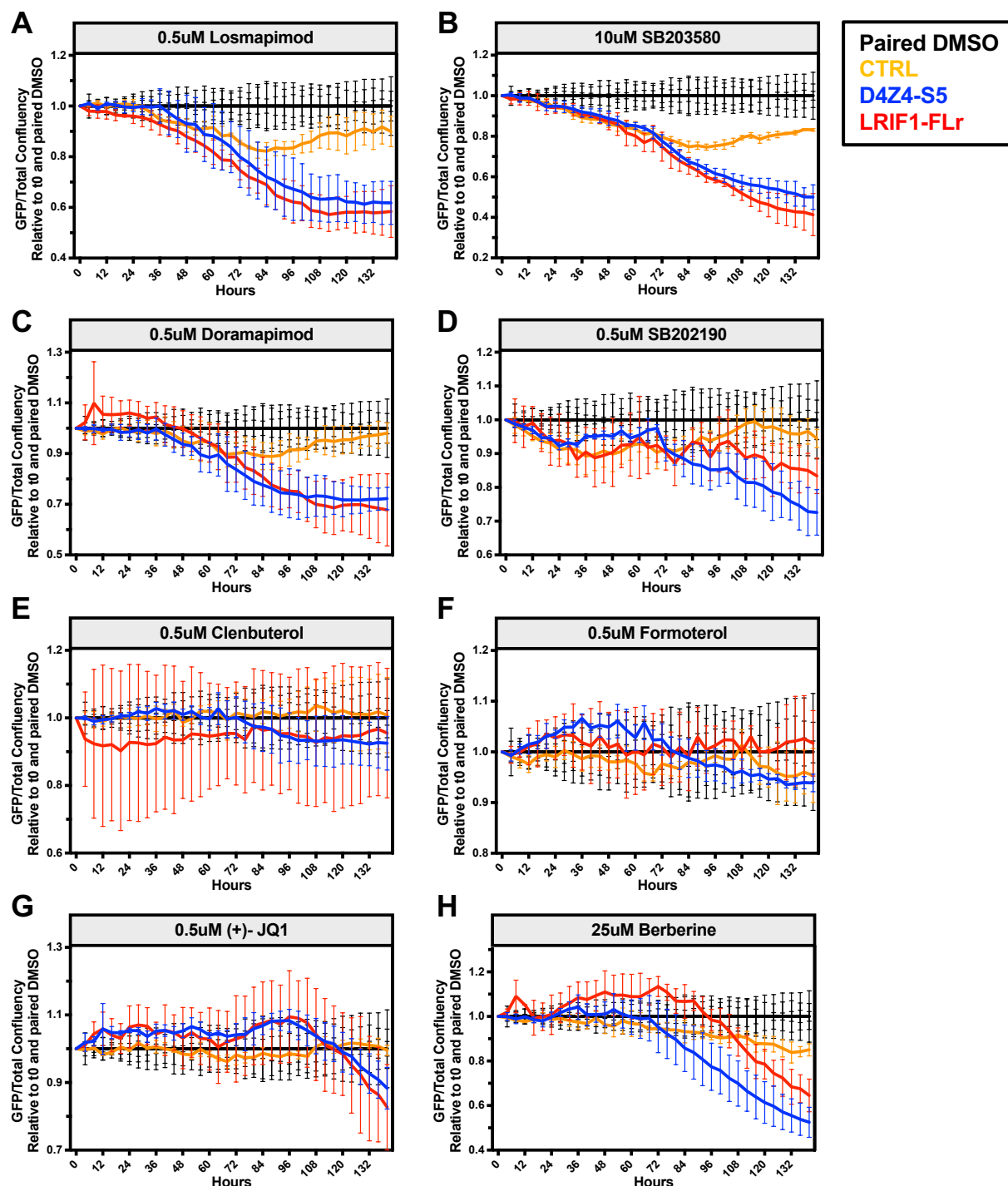


Figure 3.5: (A-H) Plots of relative GFP+ cell confluency measured via live cell imaging over 6 days upon treatment with *DUX4* inhibitor drugs shown in **Supplementary Figure S2.9**. Data shown is the GFP+ cell confluency in drug-treated wells relative to DMSO-treated wells in CTRL, D4Z4-S5, or LRIF1-FLr cells. Data points are the average relative GFP+ cell confluency $\pm$ SD, n=3 biological replicates.

Figure 3.6: Kinase Inhibitor Regularization (KiR) analysis of AAVS1 reporter kinase inhibitor screen.

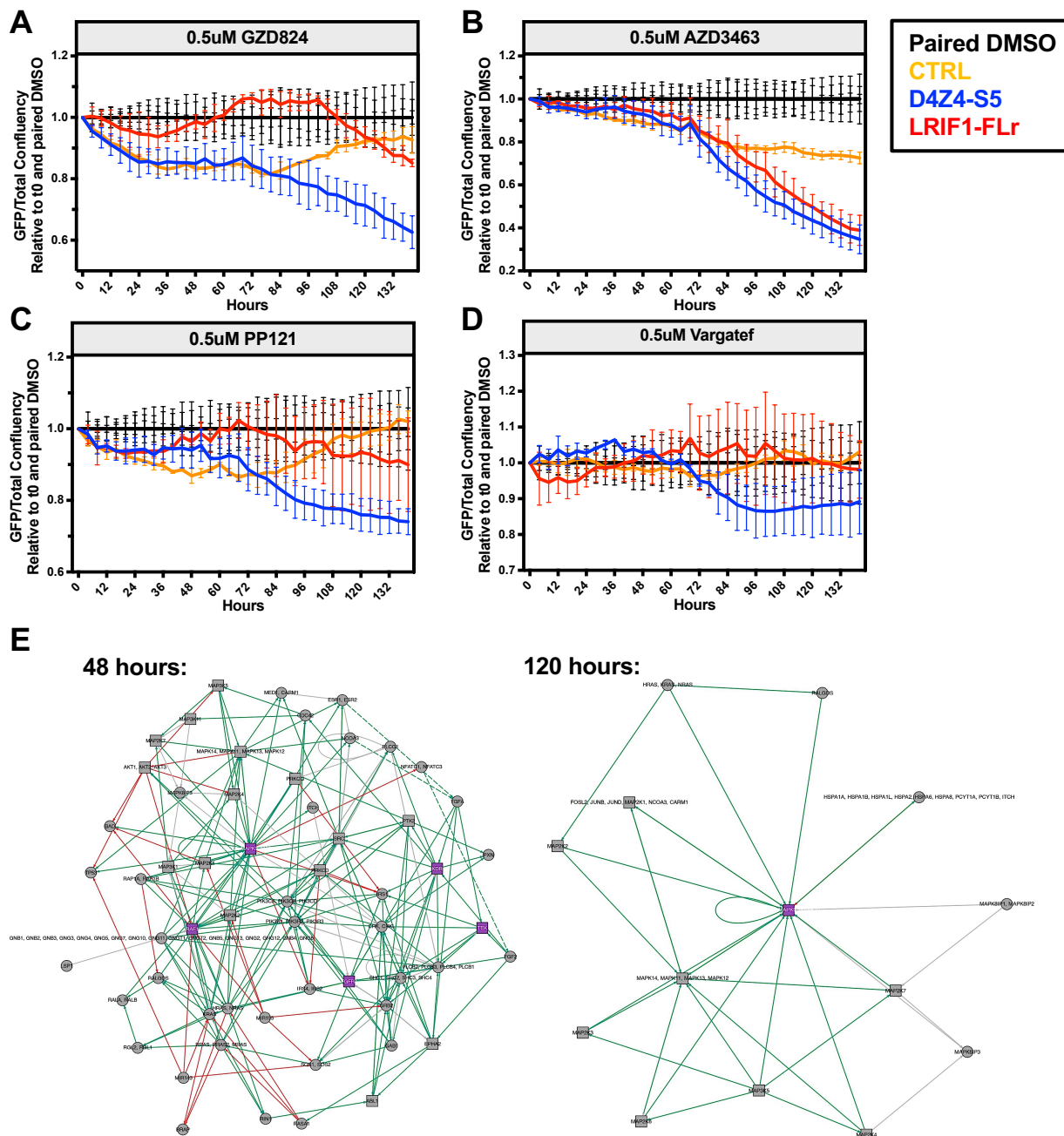


Figure 3.6: (A-D) Selected plots of relative GFP+ cell confluency measured via live cell imaging over 6 days upon treatment with kinase inhibitors. Kinase inhibitors selected show inhibition of relative GFP+ cell confluency in D4Z4-S5. Data shown is the GFP+ cell confluency in drug-treated wells relative to DMSO-treated wells in CTRL, D4Z4-S5, or LRIF1-FLr cells. Data points are the average relative GFP+ cell confluency $\pm$ SD, $n=3$ biological replicates. **(E)** Key kinase networks for factors involved in modulation of GFP+ cell confluency, identified by Kinase Inhibitor Regularization (KiR) analysis [141,140]. Purple color indicates key kinases identified by KiR, squares indicate kinases, dotted

lines indicate indirect effects, solid lines indicate direct effects, green lines are activating effects, and red lines are inhibitory effects.

Tables

Table 3.1: Kinase inhibitors used in live cell imaging screen.

Compound	CAS	Final Conc.	Primary Targets:
Baricitinib (LY3009104, INCB028050)	1187594-09-7	0.5µM	JAK1, JAK2
Doramapimod (BIRB 796, p38 MAP Kinase Inhibitor X)	285983-48-4	0.5µM	P38 MAPK
JNK Inhibitor VIII	894804-07-0	0.5µM	JNK1, JNK2, JNK3
TCS ERK 11e (ERK Inhibitor VIII, VX-11e)	896720-20-0	0.5µM	ERK1, ERK2
Rho Kinase Inhibitor III Rockout	7272-84-6	0.5µM	ROCK
PP121	1092788-83-4	0.5µM	RTKs (ABL, EGFR, EPHB4, HCK, PDGFR, RET, SRC, VEGFR2) and PI3-Ks (p110α/ β/ δ/ γ, mTOR, DNA-PK)
GZD824	1421783-64-3	0.5µM	BCR-ABL, BCR-ABL mutants
Staurosporine*	62996-74-1	0.5µM	PKA, CAMK, MLCK, PKC, PKG
AZD3463	1356962-20-3	0.5µM	ALK, IGF1R
NVP-BVU972	1185763-69-2	0.5µM	MET
Y-39983 dihydrochloride	173897-44-4	0.5µM	ROCK
Cobimetinib (GDC-0973, RG7420)	934660-93-2	0.5µM	MEK1
WZ3146	1214265-56-1	0.5µM	EGFR
Aminopurvalanol A	220792-57-4	0.5µM	CDK1, CDK2, CDK5
SB202190 (FHPI)	152121-30-7	0.5µM	P38 MAPK
PD 166285 dihydrochloride*	212391-63-4	0.5µM	WEE1, MYT1, CHK1
BIIB-057 (PRT062607, P505-15)	1370261-96-3	0.5µM	SYK, LCK, PAK5, FGR, MLK1, YES, FLT3, LYN, SRC, PYK, FAK, ZAP70
Akt Inhibitor VIII Isozyme Selective Akti-1/2	612847-09-3	0.5µM	AKT1, AKT2, AKT3
CP 547,632	252003-65-9	0.5µM	VEGFR2, FGFR
Bosutinib	380843-75-4	0.5µM	SRC, ABL
AT7867	857531-00-1	0.5µM	AKT1, AKT2, AKT3, PKA, P70S6K
Cdk1/2 Inhibitor III	443798-55-8	0.5µM	CDK1, CDK2
PDK1/Akt/Fit Dual Pathway Inhibitor	331253-86-2	0.5µM	PDK1 and AKT activities
SB 218078	135897-06-2	0.5µM	CHK1, CDC2, PKC
PF-3644022	1276121-88-0	0.5µM	MK2, MK3, PRAK
LY 333531 mesylate (Ruboxistaurin)	192050-59-2	0.5µM	PKCβI, PKCβII, PKCη, PKCδ, PKCγ, PKCα, PKCε
Cot inhibitor-2	915363-56-3	0.5µM	MAP3K8
PD 173955	260415-63-2	0.5µM	SRC, ABL, YES, ATP, MAP
Vargatef (Nintedanib, BIBF 1120)	928326-83-4	0.5µM	PDGFR, VEGFR, FGFR
TG02*	937270-47-8	0.5µM	CDK2, FLT3, JAK2

* Exhibited strong cell toxicity, excluded from analysis

Chapter 4: The transcription factor DUX4 orchestrates translational reprogramming by broadly suppressing translation efficiency and promoting expression of DUX4-induced mRNAs

This chapter is published in PLOS Biology as [99]:

The transcription factor DUX4 orchestrates translational reprogramming by broadly suppressing translation efficiency and promoting expression of DUX4-induced mRNAs

Danielle C. Hamm, Ellen M. Paatela, Sean R. Bennett, Chao-Jen Wong, Amy E. Campbell, Cynthia L. Wladyka, Andrew A. Smith, Sujatha Jagannathan, Andrew C. Hsieh, Stephen J. Tapscott

2023. PLOS Biology 21(9): e3002317. <https://doi.org/10.1371/journal.pbio.3002317>

Introduction

The double homeobox protein 4 (DUX4) gene encodes a transcription factor that is expressed in immune-privileged niches such as the preimplantation embryo [2,3], testis [9], and, possibly, thymus [40]. DUX4 is briefly expressed in the 4-cell human embryo and serves as a key transcriptional activator of the zygotic genome, driving expression of hundreds of coding genes and repetitive retroelements [2,3]. In addition to zygotic genome activation (ZGA), regulation of mRNA degradation and translation is essential to rapidly diversify the proteome during early development [146] and has been associated with increased developmental potential of human preimplantation embryos [147]. It is becoming abundantly clear that translational control, both globally and at the level of individual transcripts, helps mediate cell fate transitions. This includes the shift from the maternal to the embryonic developmental program, the balance of stem cell self-renewal and differentiation, and the plasticity of cancer [108,148].

Emerging evidence has shown that broad translational suppression is a hallmark of reprogramming in embryonic and somatic stem cells [149–151], where low rates of translation are thought to promote an undifferentiated state. This is further supported by the finding that a rare population of totipotent mouse embryonic stem cells (ESCs) known as 2-cell-like cells (2CLCs), thought to recapitulate the naïve state of the preimplantation embryo [152], exhibit global repression of nascent protein synthesis [153,154]. Mouse Dux and human DUX4 belong to the conserved DUXC family of proteins found in eutherian mammals [5]. Functionally, expression of mouse Dux reprograms these rare populations of 2CLCs to have expanded developmental potential

[2,3,152], and human DUX4 has been reported to drive a similar totipotent program in human induced pluripotent stem cells (iPSCs) and ESCs [2,6].

DUX4 is also the causative gene of facioscapulohumeral muscular dystrophy (FSHD), a complex genetic disorder that results in epigenetic de-repression of the DUX4 locus in skeletal muscle and progressive muscle atrophy [11,17,33,155]. The aberrant expression of DUX4 in skeletal muscle activates expression of genes associated with germline and stem cell development [111,156], characteristic of the early embryonic ZGA program. Although DUX4 is sporadically expressed in approximately 0.1% of FSHD muscle cells in culture [9,62], DUX4 target gene activation results in a host of pathogenic features including impaired myogenesis [157], oxidative stress and DNA damage [158,159], compromised mRNA quality control [95,96], and inflammation [160,161].

In addition to a role in FSHD, recent analysis of nearly 10,000 cancer transcriptomes from 33 different cancer types revealed DUX4 to be one of the most commonly expressed cancer-associated genes [100]. Full-length DUX4 expression in diverse cancer types is strongly correlated with increased expression of high-confidence DUX4 targets activated in the embryo [100,105]. DUX4 expression in cancers was associated with decreased major histocompatibility complex class I (MHC-I) expression, resistance to checkpoint inhibitors, and decreased patient survival rates [100]. DUX4 expression in several cancer cell lines was sufficient to prevent the induction of MHC-I expression in response to interferon gamma (IFN γ); however, it was unclear whether this represented an activity restricted to MHC-I or a broader activity of DUX4 in regulating protein expression.

In this study, we have implicated the DUX4 transcriptional program as a driver of broad translational suppression that reprograms de novo protein synthesis. We initially focused on the DUX4 suppression of MHC-class I and related interferon-stimulated proteins to characterize the mechanisms of protein suppression. We found that a brief pulse of DUX4 expression disrupts several key regulators of translation initiation and elongation, which are sufficient to suppress protein synthesis of IFN γ -stimulated MHC-I and immunoproteasome (iProteasome) subunits. Moreover, high-throughput translational profiling with ribosome footprinting and polysome gradients showed significant changes in MHC-I translational efficiency (TE) and broad translational suppression of many cellular mRNAs, whereas DUX4-induced mRNAs were robustly translated. Taken together, DUX4 activation of its transcriptional program resulted in the replacement of the prior cellular proteome and lineage identity with the DUX4-induced proteome enriched for ZGA-associated proteins. We propose that coordinated regulation of transcription and translation is employed by DUX4 to reshape the cellular translome in both development and disease.

Results

DUX4 activity induces prolonged suppression of antigen presentation factors

We recently reported that DUX4 blocks IFN γ -stimulated induction of MHC-I and surface antigen presentation [100]. To determine the mechanism of DUX4-induced MHC-I regulation, we used a well-characterized cellular model system of human myoblasts with a doxycycline (DOX)-inducible DUX4 transgene (MB135iDUX4) [97]. DUX4 expression occurs in transient bursts in rare populations of ESCs [2,6] and is sporadically misexpressed in FSHD muscle cells [9,62], making it difficult to characterize downstream mechanisms endogenously. We have previously demonstrated that a short “pulse” of DUX4 in MB135iDUX4 myoblasts induced a transcriptional program representative of FSHD muscle cells and the early cleavage-stage embryo [162]. Pulsed DUX4 expression in this cell culture system enabled reproducible and synchronized DUX4 induction, permitting the investigation of mechanisms downstream of DUX4 that may have otherwise been masked by heterogeneous populations of DUX4-expressing cells.

Using our MB135iDUX4 cell culture system, we tested both the immediate effect of continuous DUX4 expression, as well as the prolonged consequences following a brief pulse of DUX4 on the expression levels of several IFN γ -induced factors involved in immunogenic antigen presentation, including canonical MHC-I subunits (HLA-A, HLA-B, HLA-C) and iProteasome subunits (PSMB8, PSMB9, PSMB10). MB135iDUX4 myoblasts were treated with DOX for 20 hours or for a 4-hour period followed by washout to induce a “continuous” versus a “pulse” of DUX4 expression, respectively. Cells were exposed to IFN γ for the final 16 hours prior to harvest, collecting terminal time points at 20 hours for the continuous treatment, and at 44 hours for the pulse (40 hours after the DOX washout) (**Fig. 4.1A**). Cells stimulated with IFN γ showed elevated protein levels of MHC-I and iProteasome subunits, whereas IFN γ induction of MHC-I and the iProteasome was suppressed in myoblasts continually expressing DUX4 (**Fig. 4.1B**, left). Remarkably, a pulse of DUX4 elicited the same degree of suppression despite having diminished levels of DUX4 protein (**Fig. 4.1B**, right).

In a recent study, we found that the DUX4 protein was sufficient to inhibit IFN γ induction of interferon-stimulated genes (ISGs) at the mRNA level by interacting with STAT1 and preventing RNA Pol-II recruitment to STAT1-regulated genes [104]. Our current findings that suppression of IFN γ -stimulated factors persisted following a transient pulse of DUX4 suggested an additional mechanism of regulation downstream of DUX4 protein expression. We compared mRNA levels of MHC-I and iProteasome subunits following continuous or pulsed DUX4 expression and found that transcriptional suppression of IFN γ signaling by continuous DUX4 mostly recovered at 44 hours following the pulse of DUX4 (**Fig. 4.1C**). These data indicated that a pulse of DUX4 induced post-transcriptional suppression of MHC-I and iProteasome proteins through a method distinct from its interaction with STAT1 to suppress ISG mRNA induction.

To determine the necessity of DUX4 transcriptional activity for long-term protein suppression of MHC-I and the iProteasome, we performed a time course with active DUX4 and transcriptionally inactive DUX4 mutants carrying either a mutation in the DUX4 DNA binding domain (F67A) or mutations within the (L)LxxL(L) motifs of the C-terminal activation domain (mL1dL2). A prolonged time course revealed that protein suppression persisted for several days following a pulse of active DUX4, whereas transcriptionally inactive DUX4 mutants were insufficient to suppress MHC-I and PSMB9 at later time points (**Fig 4.1D**). Suppression of MHC-I and PSMB9 induction at the 20-hour time point was observed with the DUX4 F67A mutant, but not the mL1dL2 mutant (**Fig. 4.1D**). This is likely mediated by inhibition of interferon signaling and ISG transcription by the DUX4 protein that requires the C-terminal activation domain and (L)LxxL(L) motifs [104], whereas long-term suppression of MHC-I and PSMB9 that persists after the loss of DUX4 protein required the transcriptional activity of DUX4. Collectively, these data suggest that DUX4 acts as a repressor of antigen presentation factors through 2 distinct mechanisms. Here, we focus on the finding that transient DUX4 expression activates a transcriptional program required for prolonged protein suppression.

Subcellular fractionation of mRNAs and proteins showed that a pulse of DUX4 did not disrupt HLA-A, HLA-B, HLA-C, PSMB8, PSMB9, or PSMB10 mRNA nuclear export or protein localization to the cytoplasm following IFN γ stimulation (**Suppl. Fig. S4.1**). Additionally, treatment with proteasome inhibitor MG132 or autophagy inhibitor Bafilomycin did not rescue suppression of IFN γ -stimulated MHC-I, PSMB8, PSMB9, or PSMB10 following a pulse of DUX4 (**Suppl. Fig. S4.2**), eliminating protein degradation as a causal mechanism. iProteasome production of immunogenic antigens has also been linked to MHC-I protein stability; however, siRNA-mediated knockdown of iProteasome catalytic subunits PSMB8 and PSMB9 in parental MB135 myoblasts did not impact IFN γ -induced MHC-I levels (**Suppl. Fig. S4.3**). Furthermore, treatment with ONX-0914, a selective inhibitor of the iProteasome, did not reduce IFN γ -induced MHC-I levels (**Suppl. Fig. S4.3B**). Thus, MHC-I stability does not require iProteasome-dependent proteolysis. Together, these data suggest that DUX4 post-transcriptionally suppresses MHC-I and the iProteasome independently through methods of translational inhibition.

DUX4 modulates multiple pathways involved in translational regulation

We previously reported that DUX4 induces nuclear double-stranded RNA (dsRNA) accumulation resulting in phosphorylation of PKR and the eukaryotic initiation factor eIF2-alpha [163]. In a time-course experiment, phosphorylation of eIF2-alpha persisted for several days following a pulse of DUX4 (**Fig. 4.2A**). PKR plays a key role in blocking translation through the phosphorylation of eIF2-alpha at serine 51, which results in destabilized eIF2-GTP/Met-tRNAi ternary complexes [164]. We generated a

PKR knockout (KO) in MB135iDUX4 myoblasts and observed no differences in DUX4 suppression of MHC-I or iProteasome subunits between wild-type (WT) and PKR KO cells with either continuous or pulsed DUX4 induction despite the rescue of DUX4-induced eIF2-alpha phosphorylation (**Suppl. Fig. S4.3C-E**). These results demonstrated that inhibition of eIF2-alpha was not necessary for DUX4-mediated protein suppression of antigen presentation factors.

Notably, a pulse of DUX4 in MB135iDUX4 myoblasts also caused prolonged dephosphorylation of 4E-binding protein 1 (4EBP1) (**Fig. 4.2A**), a negative regulator of cap-dependent translation. In its hypo-phosphorylated state (denoted as alpha and beta), 4EBP1 functions as a scaffold protein that sequesters eukaryotic initiation factor 4E (eIF4E), a major mRNA cap-binding protein, to prevent engagement in translation initiation. Active eIF4E is marked by phosphorylation at serine 209 [165], and indeed, a pulse of DUX4 resulted in a loss of phospho-eIF4E (**Fig. 4.2A**). Several protein kinases are assumed to phosphorylate 4EBP1, the best characterized being mechanistic target of rapamycin complex 1 (mTORC1). Therefore, we investigated other effectors downstream of mTORC1 signaling pathways known to regulate translation. We found that a pulse of DUX4 also increased phosphorylation of eukaryotic elongation factor 2 (eEF2) on threonine 56 (**Fig. 4.2A**), a functional modification that disables eEF2 in mediating ribosome elongation during mRNA translation. Collectively, these data strongly suggest that DUX4 has adverse effects on both translation initiation and elongation.

Hypo-phosphorylated forms of 4EBP1 antagonize eIF4E by sequestering it from eIF4F protein complex formation with other translation initiation complex subunits eIF4A and eIF4G. Levels of active eIF4E are rate limiting for eIF4F complex assembly and cap-dependent translation initiation [166]. To directly measure if DUX4 disrupts the cap-binding capacity of the eIF4F complex, we conducted a cap pull-down assay by incubating sepharose beads coupled to 7-methylguanosine 5'-triphosphate (m7GTP) with whole cell lysate extracted from MB135iDUX4 myoblasts 3 days after a DUX4 pulse. The m7GTP beads failed to coprecipitate eIF4G and eIF4A following a pulse of DUX4 compared to untreated cell lysate, despite equivalent precipitation of eIF4E. Instead, active complex subunits were displaced by inhibitory 4EBP1 as indicated by increased precipitation of hypophosphorylated 4EBP1 with eIF4E (**Fig. 4.2B**). These data demonstrate that DUX4 disrupts the eIF4F complex, likely having substantial effects on cellular translation.

To evaluate the sufficiency of mTOR inactivation to block MHC-I and iProteasome protein levels in the absence of DUX4, we treated MB135 myoblasts with mTOR inhibitors everolimus or Torin2. While everolimus, a second-generation rapamycin analog, blocked S6K phosphorylation of RPS6, it did not efficiently target 4EBP1 hypo-phosphorylation or reduce eIF4E phosphorylation. Furthermore, treatment with everolimus followed by IFN γ did not inhibit MHC-I or iProteasome expression (**Fig.**

4.2C, left; **Suppl. Fig. S4.4A**), indicating that the S6K branch of mTORC1 signaling is not sufficient on its own to suppress protein synthesis of MHC-I and the iProteasome subunits. Conversely, treatment with Torin2, an ATP-competitive inhibitor of mTOR, blocked phosphorylation of RPS6, resulted in robust phosphorylation of eEF2, hypophosphorylation of 4EBP1, and reduced eIF4E phosphorylation. Treatment with Torin2 suppressed MHC-I, PSMB8, PSMB9, and PSMB10 protein levels (**Fig. 4.2C**, right) while having no significant effect on IFN γ induction of MHC-I and iProteasome subunits at the mRNA level (**Suppl. Fig. S4.4A**). These findings support previous observations the 4EBP-eIF4E axis, independent of RPS6, can significantly control cap-dependent translation [167], and demonstrated that combined eIF4E and eEF2 inactivation was sufficient to suppress antigen presentation factors. Additionally, MB135 cells treated with 4EGI-1, a small molecule that pharmacologically mimics 4EBP function and inhibits eIF4E [168], suppressed IFN γ -induced MHC-I and iProteasome subunits at the protein level with no significant effect on mRNA levels (**Suppl. Fig. S4.4B, C**). Together, these findings indicate that DUX4 expression alters the activity of several key regulators of translation initiation and elongation, which are sufficient to suppress protein expression of antigen presentation factors.

Response to cellular hypoxia, metabolic signaling, oxidative stress, and DNA damage are reported to be involved in early embryonic development [106,169–174] and contribute to the pathogenicity of DUX4 in FSHD [158–160,175,176]. These pathways have also been implicated in protein synthesis inhibition. Therefore, we investigated whether agents that activate each of these stress pathways would be sufficient to recapitulate the suppression of MHC-I and PSMB9 protein levels independent of DUX4. MB135 myoblasts treated with hydrogen peroxide to induce oxidative damage, DNA-damaging agent etoposide, or cobalt chloride to mimic hypoxic stress showed only slight suppression of IFN γ -induced MHC-I and PSMB9 (**Suppl. Fig. S4.5A**), much less than that observed following a pulse of DUX4. Therefore, it is likely that multiple pathways modulate protein suppression downstream of DUX4.

Transient DUX4 activity broadly suppresses nascent protein synthesis

Uncoupling of the transcriptome and proteome downstream of DUX4 is not limited to antigen presentation factors. We previously reported a discordant relationship between RNA and protein levels for a multitude of mRNAs in DUX4-expressing cells [95]. This, in combination with the dysregulation of multiple key translational regulators (**Fig. 4.2A**), suggested that DUX4 might broadly alter cellular translation. Indeed, metabolic labeling with ³⁵S-methionine/cysteine in MB135iDUX4 myoblasts treated with IFN γ showed that de novo protein synthesis was transiently suppressed following a pulse of DUX4, with an approximate 50% reduction at the 68-hour time point (**Fig. 4.2D, E**), whereas induction of transcriptionally inactive DUX4(F67A) did not alter nascent protein synthesis (**Fig. 4.2E**, right).

We confirmed DUX4 inhibition of protein synthesis by labeling cells with methionine analog L-homopropargylglycine (HPG) followed by fixation and Click-iT chemistry, wherein fluorescence microscopy and flow cytometry showed a dramatic reduction in HPG-labeled peptides after a pulse of DUX4 (**Fig. 4.2F**). As seen with 35S-methionine/cysteine labeling, HPG signal was lowest 44 hours after a pulse of DUX4, comparable to the degree of translational suppression induced by cycloheximide (CHX) treatment alone, and nascent protein synthesis started to recover in a subset of cells by 68 to 92 hours. These findings establish long-lived, yet transitory, suppression of protein synthesis downstream of the DUX4 transcriptional program that encompassed IFN γ -induced MHC-I and PSMB9 expression.

Ribosome footprinting reveals a DUX4-induced loss of 5' ribosome occupancy

To globally profile translational regulation downstream of DUX4, we performed ribosome footprinting (Ribo-seq) paired with RNA sequencing (RNA-seq) on MB135iDUX4 myoblasts with 4 treatment conditions: untreated, IFN γ alone, DUX4 pulse harvested at 68 hours, and DUX4 pulse+IFN γ harvested at 68 hours (**Fig. 4.3A**). Our sequencing reads representing ribosome-protected fragments (RPFs) displayed 3-nucleotide (nt) periodicity and exhibited lengths of 26 to 29 nt (**Suppl. Fig. S4.6**). Metagene analysis showed the majority of RPFs mapped to the coding region (CDS) and revealed a profound depletion of Ribo-seq reads mapping to the mRNA region surrounding the translation initiation site (TIS) in samples pulsed with DUX4 (**Fig. 4.3B**).

We calculated differential TE of steady-state mRNAs by measuring Ribo-seq reads relative to RNA-seq reads. As observed in our metaplot analysis, ribosome occupancy was significantly reduced within the 5' UTR, at the TIS, and across the first coding exon in samples pulsed with DUX4 relative to untreated or IFN γ -treated cells, whereas fewer differential changes occurred across the CDS or 3' UTR (**Fig. 4.3C, D**). These changes are consistent with a decrease in translation initiation affecting the 5-prime mRNA region with relative preservation of ribosome footprints over the CDS, as suggested by the molecular mechanisms identified in **Fig. 4.2**.

In cells treated with a DUX4 pulse+IFN γ versus IFN γ , 26.7% of transcripts showed significantly reduced TE at the 5' UTR, 13.9% at the TIS, and 4.3% within the first exon (**Suppl. Fig. S4.7**). These transcripts shared a large degree of overlap in enriched Gene Ontology (GO) terms, including processes involved in mRNA regulation, metabolism, translation, antigen processing, and MHC-I presentation (**Fig. 4.3E**). Thus, Ribo-seq analysis suggests that a subset of RNAs and biological processes, such as antigen presentation pathways, are particularly sensitive to DUX4 translational suppression.

Genome-wide polysome profiling identifies defects in translation initiation and elongation

Because ribosome footprinting reports the distribution of ribosome protected fragments rather than the ribosome abundance per transcript, it can be less effective at identifying differences in translation that parallel changes in protein expression. Therefore, we turned to classical polysome profiling to directly measure changes in ribosome density per mRNA using sucrose gradient-based isolation from MB135iDUX4 myoblasts treated with a pulse of DUX4+IFN γ harvested at 68 hours versus IFN γ alone. We pooled RNA fractions representing sub-polysome (40S-60S-80S), low polysome (1 to 3 ribosomes), and high polysome (>3 ribosomes) populations and performed RNA-seq analysis (**Suppl. Fig. S4.8A**). We initially measured changes in steady-state mRNA translation, excluding DUX4-altered gene expression. RNA abundance in each polysome fraction was determined relative to total input mRNA reads normalized to an internal spike-in (**Suppl. Fig. S4.8B**). The high-to-sub polysome (high/sub) ratio identified 5,800 genes that had decreased polysome association following a pulse of DUX4, consistent with a broad suppression of translation initiation, and 323 genes with increased polysome abundance ($|\log_2FC| > 1$; $p\text{-adj} < 0.01$) (**Fig. 4.3F**).

GO analysis of the 323 genes enriched in the high polysome fraction showed an abundance of mRNAs characterized as ribosomal proteins and translation factors (**Fig. 4.3G**). Many of these mRNAs contain 5-prime terminal oligopyrimidine (TOP) motifs [177] and are particularly sensitive to mTORC1 regulation of initiation factors [178,179] and eEF2K-eEF2 control of translation elongation [180]. Analysis of mRNAs with characterized TOP motifs [177] showed that TOP mRNAs remain associated with polysomes following a pulse of DUX4, while most other transcripts are depleted in the high polysome fraction (**Fig. 4.3H**). This enrichment could reflect enhanced ribosome biogenesis used to poise cells for a rapid shift in translation rate upon recovery from DUX4 suppression. Conversely, mRNA enrichment in the high polysome fraction could result from stalled and accumulating ribosomes correlated with inhibited elongation and reduced protein expression. TOP mRNA-encoded ribosomal proteins RPL10A, RPL4, RPS6, and RPS15A were suppressed by DUX4 even though the mRNAs remain bound by polysomes (**Fig. 4.3I, J**, and **Suppl. Fig. S4.8C**), consistent with inhibited translation elongation.

To elucidate DUX4 posttranscriptional suppression of MHC-I and iProteasome subunits specifically, we similarly assessed their mRNA high-to-sub polysome ratios. HLA-A, HLA-B, and HLA-C mRNAs showed reduced polysome association indicative of impaired translation initiation; however, PSMB8, PSMB9, and PSMB10 mRNAs remained associated with polysomes, like TOP mRNAs (**Fig. 4.3I**). Therefore, we propose that translation of MHC-I mRNAs is particularly sensitive to initiation defects, while a subset of mRNAs, including a subset of TOP mRNAs and iProteasome mRNAs,

appear suppressed by DUX4 possibly through combinatorial inhibition of translation initiation and stalled elongation.

Previous analysis of the DUX4-induced transcriptome in muscle cells showed that DUX4 can both activate and inhibit genes to prevent myogenic differentiation [157,181,182]. This led us to question whether a pulse of DUX4 also had a prolonged suppressive effect on myogenesis. Interestingly, several early myogenic markers, including ITGA7, PAX3, MEF2A, MEF2C, and MEF2D mRNAs, showed a loss of polysome abundance and decreased protein levels indicative of reduced translation (**Fig. 4.3I, J**). While the master regulator of skeletal myogenesis, MYOD1, showed no change in polysome abundance, and mRNAs encoding the muscle-specific gene Desmin had an increase in polysomes (**Fig. 4.3I**), both were suppressed at the protein level following a pulse of DUX4 (**Fig. 4.3J**). Thus, consistent with our molecular and Ribo-seq analysis, polysome profiling supports DUX4-induced translational suppression of broad classes of mRNAs, including those involved in antigen presentation and lineage determination.

Translation of DUX4-induced mRNAs

Our previous studies showed that mRNAs transcriptionally induced by DUX4 are translated into protein [95]. Comparing the high polysome fractions of DUX4 pulse+IFN γ samples to IFN γ -treated samples provided a measure of how DUX4 changes the overall translome and showed that DUX4-induced mRNAs were indeed associated with polysomes (**Fig. 4.4A**). Thousands of mRNAs were reduced in the DUX4 pulse+IFN γ high polysome fraction (7,765 genes; $\log_2FC < -1$, $p\text{-adj} < 0.01$); whereas 256 polysome-bound mRNAs were significantly up-regulated ($\log_2FC > 1$, $p\text{-adj} < 0.01$), many of which are well-characterized DUX4-target genes [183]. Immunoblot analysis confirmed that several DUX4-induced mRNAs in the high polysome fraction correlated with translation of these mRNAs (**Fig. 4.4B**).

The enrichment of DUX4-induced mRNAs in the high polysome fraction might reflect their increased abundance following DUX4 expression or a relative resistance to the DUX4-mediated translational inhibition, or both. Thermodynamic stability and RNA secondary structures within the 5' UTR of an mRNA can influence translation initiation rates of distinct transcripts in cis (**Fig. 4.4C**), with higher predicted minimum free energy (MFE) showing increased translation efficiency [184,185]. The annotated 5' UTRs of the 256 genes induced by DUX4 had a significantly higher predicted MFE per 100 nt relative to the average of all annotated 5' UTRs, whereas the average MFE of 5' UTRs belonging to the 7,765 genes repressed by DUX4 was significantly lower (**Fig. 4.4D**). Furthermore, selective usage of alternative transcription start sites (TSS) have been shown to alter 5' UTR sequences to influence cell type-specific protein synthesis [186]. We have previously observed that some DUX4-bound repetitive elements are co-opted to form alternative promoters for DUX4 target genes [156,183]. To account for

noncanonical TSS and splicing events, we annotated the functional 5' UTRs of 70 direct DUX4 targets based on RNA-seq alignment and published DUX4 ChIP peaks (**Suppl. Fig. S4.8D**). Indeed, DUX4 targets were predicted to have less structured 5' UTRs on average (**Fig. 4.4D**). Therefore, DUX4-induced mRNAs are predicted to be less susceptible to inhibition of translation initiation, which correlates with their observed increase in protein expression and polysome association.

Expression of endogenous DUX4 correlates with translational suppression and reduced MHC-I surface antigens in cancer cells

To determine whether endogenous DUX4 expression alters mRNA translation similar to our DOX-inducible DUX4 model system, we used FSHD muscle cells and SuSa germinoma cells, 2 cell types previously shown to express DUX4 [9,111]. In both FSHD and SuSa cells, a rare population of cells stochastically express endogenous DUX4 at any given time. These cell types, like ESCs, exhibit cell-to-cell heterogeneity in culture, with only about 0.1% to 5% of cells expressing the DUX4 transcriptional program [6,9,62]. Metabolic labeling of de novo protein synthesis with HPG in FSHD myotubes and SuSa cells showed a dramatic reduction in HPG Click-iT signal in cells expressing DUX4-target genes H3.X and H3.Y (**Fig. 4.4E**), demonstrating translation of DUX4-induced mRNAs in cells with otherwise broadly suppressed protein synthesis.

To study the downstream long-lived consequences of endogenous DUX4 expression in SuSa cells, we synchronized cells using a release from confluence culture protocol (see Methods). This resulted in a burst of endogenous DUX4 expression and prolonged expression of DUX4 target genes H3.X/Y, ZSCAN4, and LEUTX, and we showed that gapmer-mediated knockdown of DUX4 abrogated target gene expression (**Fig. 4.4F**). Fluorescence-activated cell sorting (FACS) of IFN γ -stimulated SuSa cells revealed that cells with low levels of MHC-I had elevated levels of DUX4 target gene expression relative to MHC-I high cells (**Fig. 4.4G, H**). Knockdown of endogenous DUX4 in SuSa cells rescued 5% of the MHC-I low population relative to control (**Fig. 4.4I, J**), corresponding with the percentage of cells that expressed a brief pulse of DUX4 as evidenced by prolonged expression of DUX4 targets H3.X/Y (**Fig. 4.4E, F**). Together, these results show that endogenous DUX4 expression in 2 independent cell types, FSHD muscle cells and SuSa germinoma cells, suppress protein synthesis and that this correlates with suppression of IFN γ -induced MHC-I expression in cancer cells. Our data demonstrate that DUX4 orchestrates broad inhibition of mRNA translation concurrent with de novo translation of DUX4-induced mRNAs (**Fig. 4.4K**), effectively reprogramming the cellular translationalome.

Discussion

In this study, we have shown that brief expression of the developmental transcriptional factor DUX4 results in prolonged translational reprogramming of the cell.

DUX4 expression drives an early embryonic gene program that facilitates a totipotent-like state [2,3,6]. Reprogramming cells to totipotency requires both the activation of a new gene expression program and suppression of existing gene products to erase the previous cellular identity. Broad suppression of protein synthesis is often accompanied by selective translation of mRNA networks during instances of cell stress and cellular reprogramming [187–190]. Critically, we found that DUX4 induced relatively broad translational regulation, where translation of many classes of mRNAs—including factors involved in antigen presentation, translation, and somatic cell lineage specification—were suppressed while DUX4-induced transcripts were translated. Thus, DUX4-induced transcription and downstream translational regulatory mechanisms ultimately result in reprogramming of protein synthesis, underscoring the importance of understanding DUX4 biology beyond its role as a transcriptional activator.

Many signaling pathways converge on the same key translation factors to rapidly control protein expression at both the initiation and the elongation stages. Building upon our previous observation of eIF2-alpha phosphorylation and changes to the proteome in DUX4-expressing myoblasts [95,163], we report novel effects of DUX4 involving the key regulators of translation initiation 4EBP1, eIF4E, and elongation factor eEF2. These factors are often modulated by intracellular and environmental cues, including nutrient and energy deprivation, cellular stress, hypoxia, and DNA damage. The exact intermediate mechanisms disrupting these translation factors downstream of DUX4 remain unknown. However, we show that multiple pathways are involved that may act redundantly. Complimentary to what we report here, recent studies have looked at early time points following continuous DUX4 expression in muscle cells and identified posttranscriptional regulatory mechanisms impacting protein phosphorylation, protein stability, nonsense mediated decay, and mRNA splicing [95,96,191,192].

Mechanisms underlying broad repression of translation and selective translation are often tightly coupled, and we argue that translational reprogramming is innate to the role of DUX4 as a developmental regulator. Recent work in mammalian preimplantation embryos optimized high-throughput ribosome profiling techniques for low-input applications [193–195]. These studies demonstrated a marked shift in the translome associated with zygotic genome activation. As protein synthesis is an energetically expensive cellular process, we propose a model by which DUX4 initiates broad suppression of translation to reduce the cellular burden of protein synthesis while DUX4-induced mRNAs undergo translation to produce proteins vital for development. In line with this, we found that DUX4-induced mRNAs are predicted to have less structured 5' UTRs, are associated with polysomes, and are up-regulated at the protein level. We propose that translational reprogramming by DUX4 facilitates translation of mRNAs promoting an early embryonic program when general protein synthesis is compromised.

We also provide evidence of decreased de novo protein synthesis in FSHD and cancer cells endogenously expressing DUX4. Similar expression profiles are found in a

rare cell population of mouse ESCs that exhibit a 2C-like signature driven by the expression of mouse Dux, a functional ortholog of human DUX4 [152,196]. These 2CLCs have been associated with increased potency [152,197] and a global reduction of translation [153], including suppression of pluripotency proteins [154] and ribosomal proteins [198]. Like the transient attenuation of protein synthesis following a pulse of DUX4 in human cells, translational suppression in this rare population of mouse ESCs is transitory and protein synthesis resumes upon exit from the 2C-like state [154,198], highlighting the dynamic nature of molecular events as cells transition between states.

Like stem cells, tumors exhibit heterogeneity with certain cells undergoing dedifferentiation and reactivating developmental genes, giving rise to cells with expanded potential [199]. Protein synthesis is frequently dysregulated in cancer [200–204], and translational reprogramming is increasingly recognized as a contributor to tumor heterogeneity and adaptive plasticity [108,205]. Indeed, examples of translational control involving the ternary complex member eIF2-alpha [206,207] and subunits of the translation initiation complex eIF4F, particularly eIF4E [208], have been shown to regulate tumor immune escape. Additionally, the eukaryotic elongation factor 2 kinase (eEF2K), responsible for inhibitory eEF2 phosphorylation that impedes protein synthesis, is overexpressed in several cancers and associated with poor survival outcomes [209]. Our study extends these observations, as we find DUX4 disrupts many of these same translational regulators to modulate immune signaling.

In summary, our study demonstrates that transient expression of DUX4 reprograms the translome through combined transcriptional and post-transcriptional control. Monumental shifts in the transcriptome and translome occur during the initial stages of development as the fertilized egg transitions to the totipotent cells of the early embryo [146,210]. In human preimplantation embryos, ZGA is required for adequate suppression of the previous maternal program [147]. We provide molecular insight into how DUX4, a driver of ZGA, facilitates cellular reprogramming in somatic cells by controlling mRNA translation and suggest that it likely has a similar role during embryogenesis. We speculate that DUX4-expressing cancer cells might hijack these mechanisms to enhance cellular plasticity and create an immunosuppressive milieu, thereby promoting tumorigenesis and therapeutic resistance. This work provides new insights into cellular reprogramming and highlights opportunities for FSHD and cancer treatments targeting DUX4 and its downstream effectors. As a result, combining our knowledge of DUX4 biology in cancer, FSHD, and embryonic development will be critical to understand conserved pathways and to develop innovative therapeutic approaches.

Materials and Methods

Cell culture

MB135 myoblasts were grown in Ham's F-10 supplemented with 10% FBS, 1% penicillin/streptomycin, 10 ng/mL rhFGF, 1 μ M dexamethasone, and 3 μ g/mL puromycin as appropriate to maintain lines carrying the DUX4 transgene. SuSa cells were grown in RPMI 1640 supplemented with 10% FBS and 1% penicillin/streptomycin. Differentiation of FSHD myoblasts into myotubes was achieved by switching myoblast grown to confluence into DMEM, 1% penicillin/streptomycin, 10 μ g/ml insulin, and 10 μ g/ml transferrin for 48 hours. Pulsed MB135iDUX4 myoblasts were treated with 1 to 2 μ g/mL DOX for 4 hours, rinsed with PBS, and fresh growth media added. MB135iDUX4 myoblasts with continuous DUX4 induction were treated with 1 μ g/mL DOX for 20 hours. All cell types were stimulated with 50 ng/mL IFN γ where specified. MB135iDUX4 myoblasts were pulsed with DOX for 4 hours, incubated for 48 hours, supplemented with MG132 (0.5 μ M), Bafilomycin A1 (0.5 μ M), or ONX-0914 (200 nM) with the addition of 50 ng/mL IFN γ for 16 hours. MB135 myoblasts treated with mTOR inhibitors Everolimus or Torin2 (Sigma) were incubated in media supplemented with 10 to 100 nM inhibitor for 24 hours, followed by an additional 16-hour incubation in media resupplemented with inhibitor plus 50 ng/mL IFN γ . MB135 myoblasts were treated with 25 to 50 μ M 4EGI-1 (Sigma) for 48 hours, followed by an additional 16-hour incubation in media resupplemented with 4EGI-1 plus 50 ng/mL IFN γ . MB135 myoblasts treated with cell stress-inducing reagents etoposide (1 to 5 μ M), hydrogen peroxide (200 to 400 μ M), or cobalt chloride (250 to 500 μ M) were incubated for 24 hours, followed by an additional 16-hour incubation in media resupplemented with cell stress-inducing reagent plus 50 ng/mL IFN γ . All cell lines were cultured at 37°C in a humidified incubator supplied with 5% CO₂.

CRISPR-Cas9 knockout generation

Generation of MB135iDUX4 PKR KO myoblasts was achieved using a guide RNA (gRNA) sequence targeting EIF2AK2 [211] cloned into the Cas9(BB)-2A-GFP plasmid (Addgene). MB135iDUX4 myoblasts were transfected with this construct using Lipofectamine 3000 Reagent (Invitrogen) according to the manufacturer protocol and incubated for 1.5 days. Clonal cell lines were isolated using fluorescence activated cell sorting. Individual PKR KO clones were screened using immunoblot analysis, and mutant alleles were validated with Sanger sequencing.

Starvation-induced cell cycle synchronization

SuSa cells were seeded at 90% confluence on 0.1% gelatin-coated plates and incubated for 1 week at 37°C and 5% CO₂. Cells were supplemented with fresh growth media and incubated for 3 to 4 hours to release from synchronization, lifted using trypsin, seeded onto gelatin-coated plates at 30% confluence, treated with or without IFN γ for the last 16 hours, and harvested at terminal time points of 24, 48, or 72 hours for downstream quantitative reverse transcription PCR (RT-qPCR), IF, or flow cytometry analysis. For flow cytometry, cells were stained with BV605 anti-human HLA-A,B,C Antibody (BioLegend #311432), sorted using BD FACS Aria II, or analyzed using BD LSRFortessa X-50, paired with BD FACSDiva software. Data were analyzed using FlowJo V10.5.3.

siRNA and gapmer transfections

Transfections of siRNAs into myoblasts were carried out using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's protocol. A double transfection protocol was followed with MB135iDUX4ca pulse experimental conditions to ensure prolonged depletion of target proteins, where cells were transfected 20 hours before and 20 hours after a 4-hour pulse of DOX. SuSa cells were synchronized and released as described above and reverse transfected with 1 μ l/mL Lipofectamine RNAiMAX and 25 pmol/mL of either a pool of 2 DUX4-targeting gapmers [51] or a nontargeting control gapmer. Cells were incubated with gapmers for 24 hours.

Immunoblotting

Protein samples were harvested in RIPA buffer [150 mM NaCl, 1% NP-40, 0.5% Na-deoxycholate, 1% SDS, 25 mM Tris-HCl (pH7.4)] supplemented with protease and phosphatase inhibitor tablets, followed by sonication in Diagenode Bioruptor. Lysate was cleared by centrifugation at 16,000 $\times$ g and quantified using a Pierce BCA assay. Samples were run on NuPAGE precast polyacrylamide gels and transferred to PVDF membrane. Membranes were blocked in PBS containing 0.1% Tween-20 and 5% nonfat dry milk before overnight incubation at 4°C with primary antibodies. Membranes were incubated with horseradish peroxidase-conjugated secondary antibodies for 1 to 2 hours at room temperature, and SuperSignal chemiluminescent substrate was used for detection on film. Membranes were stripped with Restore Western Blot Stripping Buffer.

Subcellular fractionation

MB135iDUX4 myoblasts were pulsed with DOX for 4 hours, incubated for 24 hours, supplemented with 50 ng/mL IFN γ for 16 hours, and harvested at a terminal time point of 44 hours. Samples divided for whole-cell lysate (WCL) and subcellular fractionation were suspended in 300 μ L ice-cold Cyto-lysis buffer (10 mM Tris (pH 7.4), 10 mM NaCl, 0.2% NP-40, 1 mM DTT in nuclease-free water). Subcellular fractionation samples were centrifuged at 650 RCF to pellet nuclei, while cytoplasmic lysate remained in the supernatant. WCL, cytoplasmic, and nuclear RNA and protein were harvested for RT-qPCR and immunoblotting, respectively.

RT-qPCR

Total RNA was isolated using the NucleoSpin RNA kit according to manufacturer instructions. Isolated RNA was treated with Amp-grade DNase I, heat inactivated in the presence of EDTA, and reverse transcribed into cDNA using SuperScript IV First-Strand Synthesis System following the manufacturer's protocol. Quantitative PCR was carried out on a QuantStudio 7 Flex using iTaq SYBR Green Supermix.

m7GTP cap-binding assay

MB135iDUX4 myoblasts were treated with and without a 4-hour pulse of DOX and harvested at a terminal time point of 68 hours. Cells were lysed in cap binding buffer [10 mM Tris-HCl (pH 7.5), 140 mM KCl, 4 mM MgCl₂, 1 mM DTT, 1 mM EDTA, 1% NP-40] supplemented with protease and phosphatase inhibitor cocktails, incubated on ice for 30 minutes, and lysate cleared at 12,000 rpm for 30 minutes at 4°C. Soluble lysate was quantified with Pierce BCA assay and diluted in cap binding buffer without NP-40 to bring final concentration to 0.5 mg/mL in 0.5% NP-40. Around 50 μ l of

prewashed 7-methyl-GTP-Sepharose bead slurry was added to 400 µg protein and incubated at 4°C for 1 hour. Samples were centrifuged at 5,000 rpm for 5 minutes at 4°C, washed twice with cap binding buffer containing 0.5% NP-40 and twice with PBS. Beads were suspended in NuPAGE LDS Buffer and incubated at 95°C for 10 minutes to elute associated proteins.

[35S] Radiolabeling

Cells were treated with and without a 4-hour pulse of DOX. Eight hours prior to harvest, cells were incubated in DMEM depleted for methionine and cysteine supplemented with 90 microcurie 35S-methionine/cysteine and 50 ng/mL IFN γ . Protein samples were harvested and run on NuPAGE gels as previously described. Gels were stained with InstantBlue Coomassie, dried on whatman paper, exposed to phosphor screen, imaged on Typhoon Trio imager, and analyzed with ImageQuant.

HPG Click-iT and immunofluorescence

Cells were incubated for 30 minutes in DMEM or RPMI media depleted for methionine and cysteine, followed by a 1-hour incubation in methionine-depleted media supplemented with 200 µM HPG. Cells were fixed with 4% paraformaldehyde for 10 minutes, permeabilized 0.5% TritonX-100, and stained with Click-iT HPG Alexa Fluor 488 Protein Synthesis Assay Kit according to manufacturer's protocol. Samples were incubated with primary antibodies at 4°C overnight, followed by incubation with fluorescently conjugated secondary antibodies for 1 hour at room temperature, and counterstained with DAPI. Plates were imaged using an immersion lens, Zeiss Axiophot fluorescent microscope, AxioCam MRc digital camera, and AxioVision 4.6 software. For flow cytometry, cells were analyzed using BD LSRFortessa X-50 with BD FACSDiva software. Data were analyzed using FlowJo v10.5.3.

Ribosome footprinting

Ribo-seq was performed as described previously by Calviello and colleagues [212] using two 70% confluent 15 cm plates of MB135iDUX4 myoblasts per treatment condition (n = 3). Ribosome complexes were isolated using MicroSpin S-400 HR Columns and RNA extracted using the Direct-zol RNA Miniprep Kit. The rRNA Removal Mix–Gold component of Illumina's TruSeq Stranded Total RNA Library Prep Gold kit was used to deplete rRNAs. RPFs were isolated by running samples on a 15% TBE-Urea gel with 10 bp DNA Ladder and Marker-27nt and Marker-30nt. Gels were stained with SYBR Gold and RNA fragments 27 to 30 nt were isolated. RNA samples were diluted to equal input concentrations; libraries were prepared using the NEXTflex Small RNA-Seq Kit v3 following the manufacturer's instructions and sequenced using 50 bp paired-end sequencing on the Illumina NextSeq platform by the Fred Hutchinson Cancer Center Genomics Core.

Polysome fractionation

Polysome profiling was performed using three 70% confluent 15 cm plates of MB135iDUX4 myoblasts per treatment condition (n = 3). MB135iDUX4 myoblasts were pulsed with or without DOX for 4 hours, incubated for 48 hours, supplemented with 50 ng/mL IFN γ for 16 hours, and harvested at a terminal time point of 68 hours. To harvest,

culture medium was supplemented with 100 µg/mL CHX and cells were incubated at 37°C for 10 minutes. Media was aspirated and each 15 cm plate of adherent cells was rinsed with 25 mL ice-cold PBS supplemented with 100 µg/mL CHX, lifted with 0.25% Trypsin–EDTA supplemented with CHX, and resuspended in growth media supplemented with CHX. Cells were pelleted and flash frozen in liquid nitrogen. Cells were lysed with Polysome Lysis Buffer (10 mM Tris (pH 8), 140 mM NaCl, 7.5 mM MgCl₂, 0.25% NP-40, 0.1% Triton X-100, 150 µg/mL CHX, 20 mM DTT, 640 U/mL SUPERase-In RNase Inhibitor) and clarified before quantification with the Bio-Rad Protein Assay. About 1.5 mg of clarified lysate was loaded onto a 10% to 50% sucrose gradient prepared in DEPC-treated water with 25 mM Tris (pH 7.4), 25 mM NaCl, 5 mM MgCl₂, 100 µg/mL heparin, and 2 mM DTT. Gradients were fractionated using a Biocomp Piston Gradient Fractionator. Samples were resuspended in Trizol and *Drosophila* S2 cells were added as an internal spike-in control. RNA was extracted using the Direct-zol RNA Miniprep Kit. Relative mRNA abundance for each sample was normalized to *Drosophila* spike-in to account for differences in RNA extraction efficiency. RNA-seq libraries were prepared using the Illumina TruSeq RNA Sample Prep v2 Kit and sequenced using 50 bp paired-end sequencing on the Illumina NextSeq platform by the Fred Hutchinson Cancer Center Genomics Core.

RNA-seq data analysis

To preprocess the RNA-seq data, we used Trimmomatic to trim the 3' adapter sequence and aligned the trimmed reads to GRCh38 (p13) using Rsubread. We counted mapped reads that overlapped with exons of gene features using the summarizeOverlaps function of Bioconductor's GenomicAlignments package [133] with the IntersectionStrict mode. The gene features were annotated by Gencode v35. We applied DESeq2 for gene expression normalization, log transformation, and differential analysis on different comparisons between treatment conditions.

Ribo-seq data analysis

We performed preprocessing and quality assessment for Ribo-seq data in the following manner: (1) Clipped the 3' adapter sequence and trim the first and last 4 bases from the adapter-clipped reads (cutadapt). (2) Removed rRNA and other small RNA such as tRNA and snoRNA. We customized the rRNA reference genome from the RNA central database (<https://rnacentral.org/>) and then used Bowtie2 to align the trimmed reads against it; the unmapped reads are the desirable RPFs. (3) Aligned RPFs to the GRCh38 (p.13) genome built by STAR. (4) Assessed the quality and size of RPFs using Bioconductor's ribosomeProfilingQC package. (5) Profiled ribosome footprints using ribosomeProfilingQC and the Gencode (v35) annotation. We computed the p-sites counts of RPFs of the dominant length (26 to 29 nt) for 5 different genomic features including the 5' UTRs, translation start sites (TIS; 13 nt extended up/downstream from the start codon), CDS, and 3' UTRs (**Suppl. Fig. S4.6**). We formulated the treatment effects on translation efficiency as changes in Ribo-seq reads between 2 different treatments relative to RNA-seq reads. The null hypothesis is defined as

$$H_0 : \left\{ \log_2 \frac{\text{treated}}{\text{untreated}} \Big|_{\text{Ribo-seq}} - \log_2 \frac{\text{treated}}{\text{untreated}} \Big|_{\text{RNA-seq}} \right\} = 0.$$

We performed hypothesis tests on 5 different genomic features using DESeq2. S7 Fig depicts the flow of the differential analysis and specifies the dynamic nonspecific filtering prior to DESeq2 for different genomic features. To only analyze steady-state genes, DUX4-altered genes (defined by $|\log_2(\text{DUX pulse}/\text{untreated})| > 1$ and adjusted p-value < 0.05 in RNA-seq samples) and IFN γ -altered genes (defined by $|\log_2(\text{IFN}\gamma/\text{untreated})| > 1$ and adjusted p-value < 0.05 in RNA-seq samples) were excluded from differential analysis. We used Bioconductor's goseq package to perform GO analysis on the transcripts showing down-regulated translational changes (FDR = 0.05).

Polysome-seq data analysis

Polysome RNA-Seq reads were quality checked with FastQC, trimmed using Trim Galore!, and aligned to the human reference genome GRCh38 p13 as well as the DM6 Drosophila reference genome using Rsubread. Reads were mapped to the Gencode v35 reference annotation using FeatureCounts. The human counts were normalized using both the DM6 and GRCh38 library sizes, log transformed, and analyzed for differential gene expression using DESeq2. Differential mRNA polysome association was analyzed in 3 different ways:

1. Polysome gradient relative to total mRNA: We formulated the changes of mRNA abundance in sub-polysome, low polysome, and high polysome fractions between 2 different treatments as a ratio of ratios, calculating the abundance of mRNA in each fraction relative to total RNA-seq. DUX4-altered genes (defined as genes with $|\log_2\text{FC}| > 2$ and adjusted p-value < 0.01 in DUX4 pulse+IFN γ versus IFN γ sample inputs) were excluded from the differential analysis. The null hypothesis is defined as

$$H_0 : \left\{ \log_2 \frac{\text{fraction}}{\text{total mRNA}} \Big|_{\text{DUX4+IFN}\gamma} - \log_2 \frac{\text{fraction}}{\text{total mRNA}} \Big|_{\text{IFN}\gamma} \right\} = 0.$$

2. High-to-sub polysome ratio: We formulated the changes of mRNA abundance in the high polysome fractions relative to the sub-polysome fraction between 2 different treatments as a ratio of ratios. DUX4-altered genes were excluded from the differential analysis. The null hypothesis is defined as

$$H_0 : \left\{ \log_2 \frac{\text{high polysome}}{\text{sub - polysome}} \Big|_{\text{DUX4+IFN}\gamma} - \log_2 \frac{\text{high polysome}}{\text{sub - polysome}} \Big|_{\text{IFN}\gamma} \right\} = 0.$$

3. Differential gene expression in high polysome fraction: We formulated the changes of mRNA abundance in polysome fractions between 2 different treatment conditions. DUX4-altered genes include in analysis. The null hypothesis is defined as

$$H_0 : \left\{ \log_2 \frac{\text{DUX4 + IFN}\gamma \text{ high polysome}}{\text{IFN}\gamma \text{ high polysome}} \right\} = 0.$$

5' UTR analysis

Annotated 5' UTR coordinates were obtained from Gencode v35, and the underlying sequences were extracted from the GRCh38 patch 13 genome using AGAT. Genes were filtered to only include transcripts expressed in MB135iDUX4 myoblasts expressing DUX4 for downstream 5' UTR analysis, and duplicate sequences and 5' UTRs with less than 4 nt were excluded. Alternative 5' UTRs of direct DUX4 target genes were annotated using RNA-seq alignments from the human myoblast inducible DUX4 model (GSE85461). (RNAfold from ViennaRNA package [213] was used to predict MFE.

Figure 4.2: A pulse of DUX4 negatively regulates the status of multiple translational regulators and broadly suppresses nascent protein synthesis.

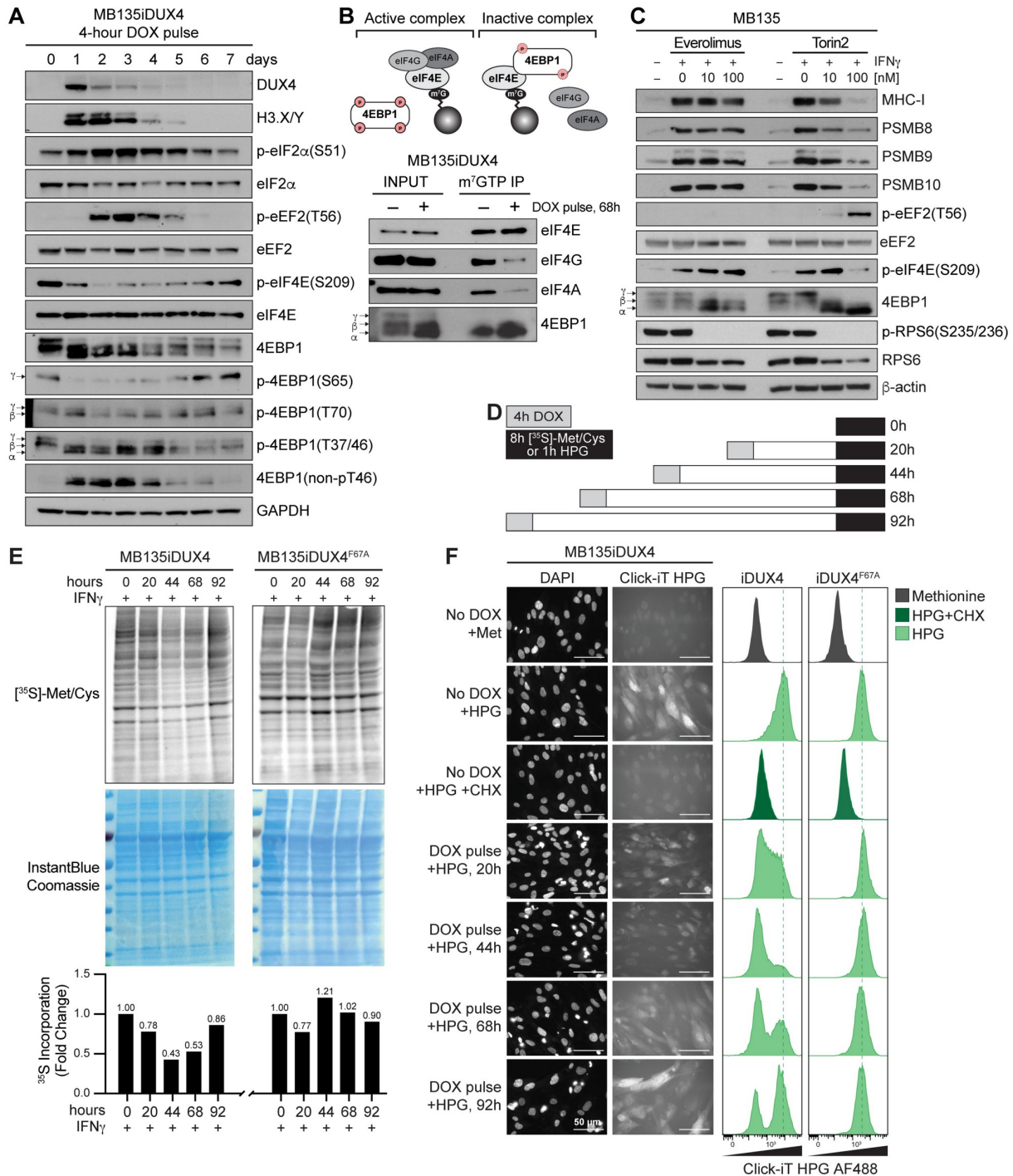


Figure 4.2: (A) Immunoblot analysis 0–7 days following a 4-hour pulse of DOX in MB135iDUX4 myoblasts. Alpha, beta, and gamma correspond to the phosphorylated forms of 4EBP1. GAPDH serves as loading control. **(B)** Schematic of cap-dependent translation initiation complex (top) and immunoblot analysis of m⁷GTP pull-downs (bottom). **(C)** Immunoblot analysis of MB135 myoblasts treated with mTORC1 inhibitors

Everolimus or Torin2. **(D)** Schematic of experimental time course for metabolic labeling with ³⁵S or HPG. **(E)** Autoradiograph of samples pulsed with active DUX4 (top, left) or DNA-binding mutant F67A (top, right); Coomassie stain of total protein (middle); quantification of relative ³⁵S signal normalized to paired 0-hour condition (bottom). **(F)** Immunofluorescence (left) and flow cytometry (right) of HPG/Click-iT labeled proteins (scale bars, 50 μ m). DOX, doxycycline; DUX4, double homeobox protein 4; HPG, L-homopropargylglycine; mTORC1, mechanistic target of rapamycin complex 1; m7GTP, 7-methylguanosine 5'-triphosphate; 4EBP1, 4E-binding protein 1.

Figure 4.3: Ribosome footprinting and polysome profiling reveal DUX4 suppressive effects on translation efficiency.

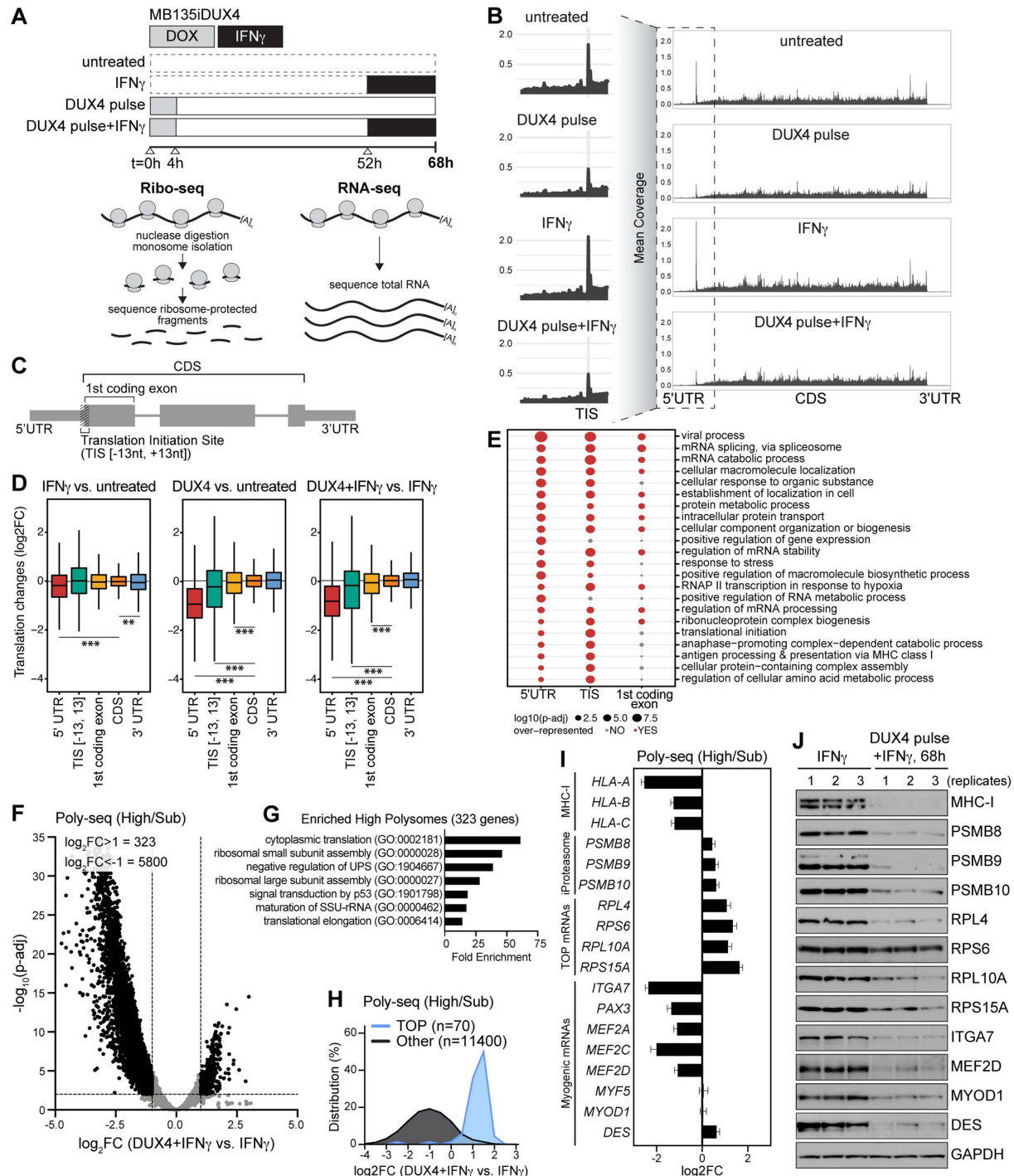


Figure 4.3: (A) Experimental schematic illustrating genome-wide quantification of mRNA counts and RPFs using RNA-seq and ribosome profiling (Ribo-seq). **(B)** Metagene analysis of RPFs. **(C)** Schematic of annotated mRNA features 5' UTR, TIS (-/+ 13 nt around start codon), first coding exon, CDS, and 3' UTR. **(D)** Box plots of translational changes occurring at the level of mRNA features ($|\log_2FC| > 1$, $p\text{-adj} <$

0.05); see **Suppl. Fig. S4.7**. Statistical comparisons were conducted using one-way ANOVA, * $p < 0.001$, ** $p < 0.0001$, *** $p < 2.2 \times 10^{-16}$. **(E)** Gene Ontology analysis for transcripts with significantly decreased ribosome occupancy in 5' UTR, TIS, and first coding exon in MB135iDUX4 myoblasts treated with a DUX4 pulse+IFN γ versus IFN γ . **(F)** Differential RNA-seq analysis of mRNA levels in high polysome fractions relative to sub-polysome fractions (high/sub). Volcano plot showing log2 fold-change differential polysome abundance in cells treated with DUX4 pulse+IFN γ harvested at 68 hours versus IFN γ (significance defined as basemean>50, |log2FC>1|, p-adj<0.01). **(G)** Gene Ontology analysis for transcripts with significantly increased polysome abundance. **(H)** Relative polysome abundance for 5' TOP mRNAs compared to all other mRNAs. **(I)** Log2 fold-change in polysome abundance of MHC-I mRNAs, iProteasome subunits, select TOP mRNAs, and myogenic factors. Data represent mean $\pm$ SD of biological replicates, $n = 3$. **(J)** Immunoblot analysis of total protein lysate from polysome profiling samples representing biological replicates, $n = 3$. GAPDH serves as loading control. CDS, coding sequence; DUX4, double homeobox protein 4; IFN γ , interferon gamma; MHC-I, major histocompatibility complex class I; nt, nucleotide; RNA-seq, RNA sequencing; RPF, ribosome-protected fragment; TIS, translation initiation site; TOP, terminal oligopyrimidine; UTR, untranslated region.

Figure 4.4: DUX4 orchestrates translational reprogramming through broad inhibition of translation concurrent with de novo translation of DUX4 target mRNAs.

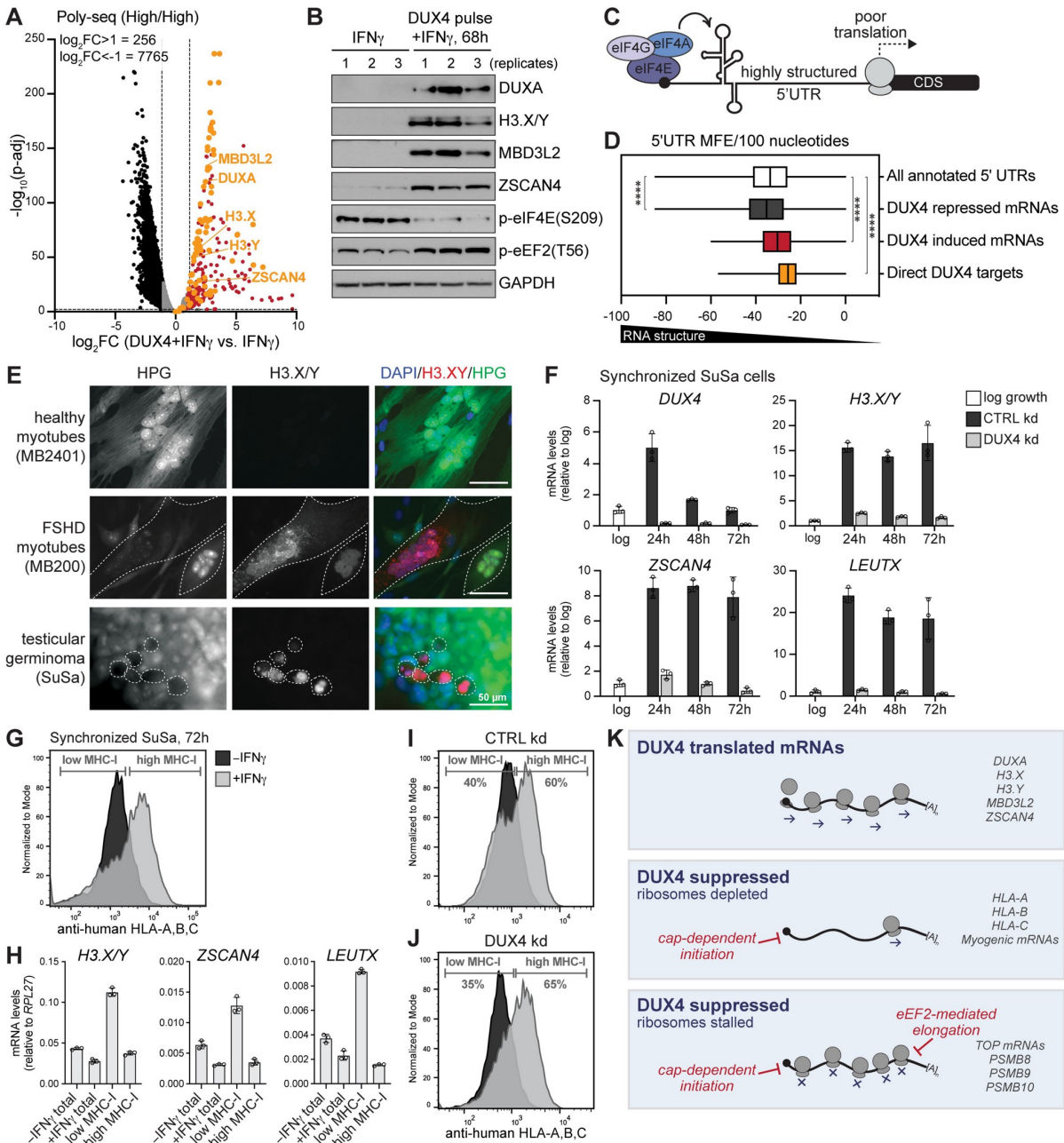
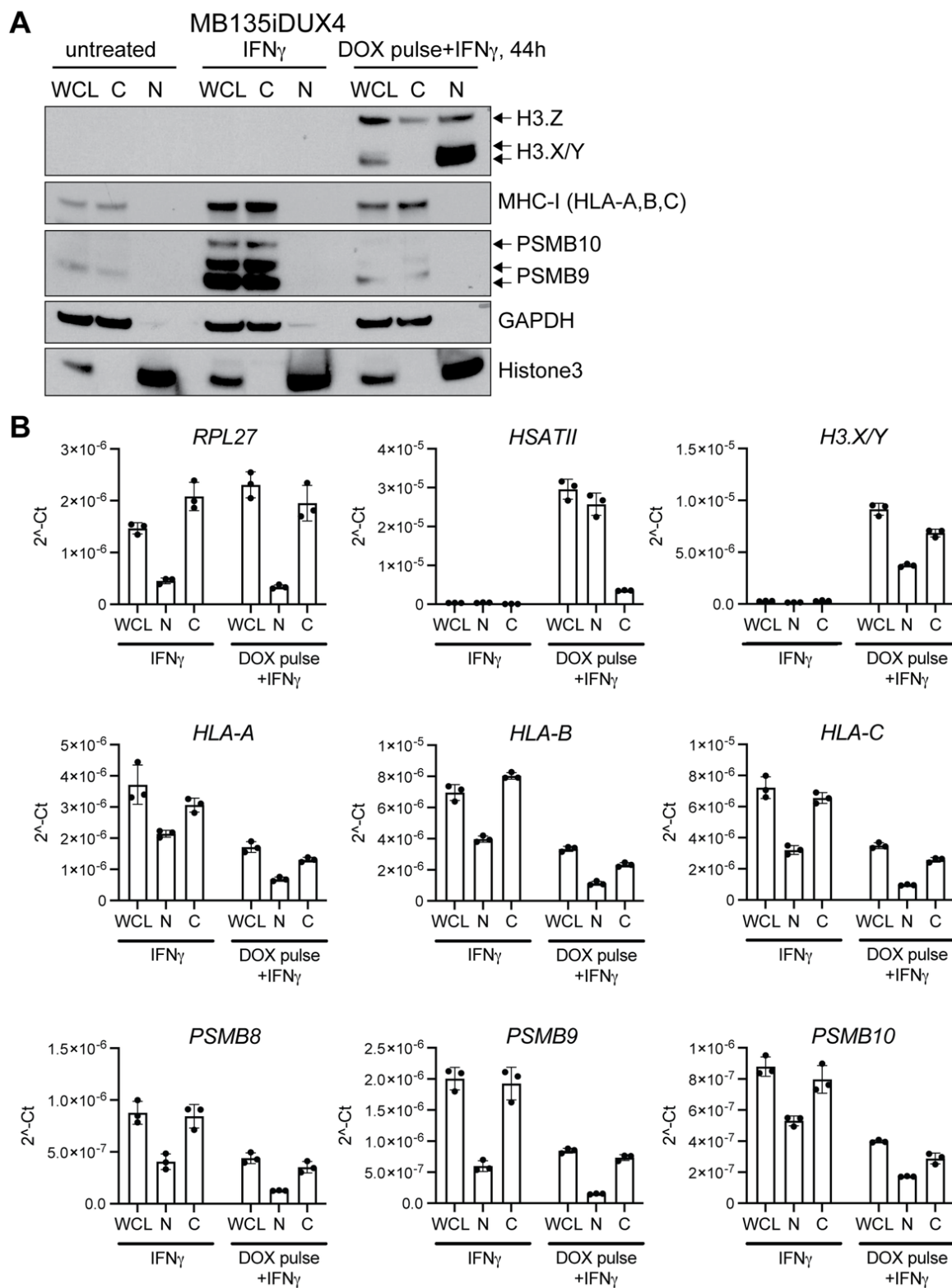


Figure 4.4: (A) Differential RNA-seq analysis of high polysome fractions (high/high). Volcano plot showing $\log_2$ fold-change of DUX4 pulse+IFN γ harvested at 68 hours versus IFN γ (significance defined as $\log_2 FC > 1$, $|\log_2 FC| > 1$, $p\text{-adj} < 0.01$). Polysome-bound mRNAs up-regulated by DUX4 highlighted in red ($n = 256$ genes). Direct DUX4 target genes highlighted in orange ($n = 70$ genes). **(B)** Immunoblot analysis of total protein lysate harvested for polysome profiling samples representing biological replicates, $n = 3$. **(C)** Schematic of mRNA with structured 5' UTR that impedes

translation. **(D)** Box plot showing 5' UTR analysis of predicted MFE per 100 nt, including all annotated mRNAs, a subset of direct DUX4 targets, mRNAs up- or down-regulated following a pulse of DUX4 (defined as mRNAs differentially expressed in **(A)**). Statistical comparisons were conducted using Mann–Whitney *U* test, **** $p < 1 \times 10^{-8}$. **(E)** Immunofluorescence of HPG Click-iT labeled nascent proteins in differentiated FSHD myotubes and SuSa cells co-stained for DUX4-target genes H3.X/Y. MB2401 myotubes serve as a control cell line that does not express DUX4 (scale bars, 50 μ m). **(F)** RT-qPCR analysis of unsynchronized SuSa cells in log growth phase (log) relative to a time course following release from synchronization with gapmer-mediated CTRL or DUX4 kd. Data represent mean $\pm$ SD of biological replicates, $n = 3$ **(G)** FACS analysis of MHC-I surface levels on SuSa cells 72 hours after synchronization and treated with and without IFN γ . **(H)** RT-qPCR analysis of SuSa cells treated with or without IFN γ and sorted based on high versus low MHC-I surface levels highlighted in **(G)**. **(I, J)** Flow cytometry analysis of MHC-I surface levels of SuSa cells 72 hours after synchronization and gapmer-mediated CTRL or DUX4 kd with and without IFN γ treatment. **(K)** Model of ribosome abundance and translation efficiency resulting from DUX4-induced inhibition of translation initiation and elongation. CTRL, control; DUX4, double homeobox protein 4; FACS, fluorescence-activated cell sorting; FSHD, facioscapulohumeral muscular dystrophy; HPG, L-homopropargylglycine; IFN γ , interferon gamma; kd, knockdown; MFE, minimum free energy; MHC-I, major histocompatibility complex class I; nt, nucleotide; RNA-seq, RNA sequencing; RT-qPCR, quantitative reverse transcription PCR; UTR, untranslated region.

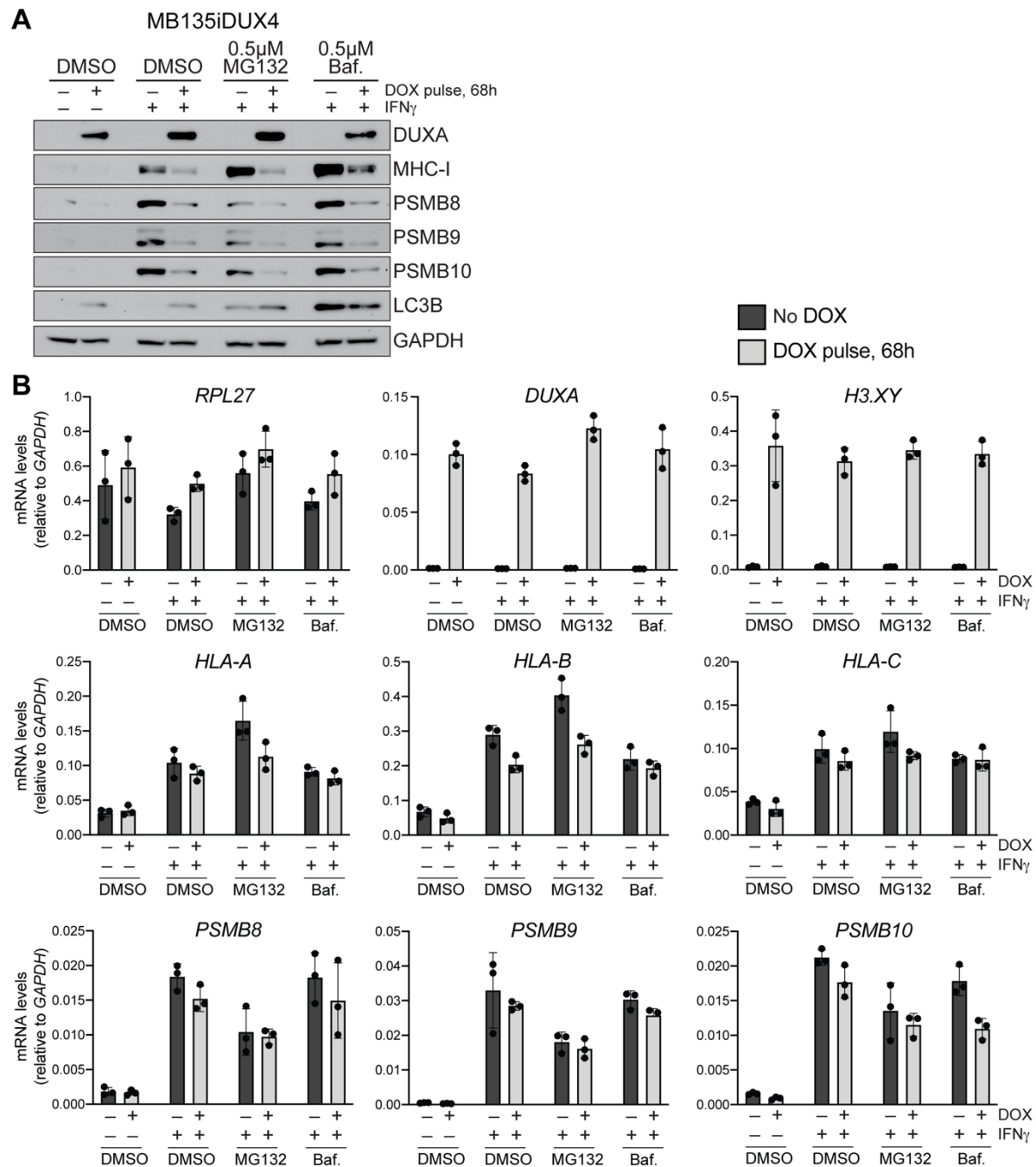
Supplementary Figures

Supplementary Figure S4.1: Subcellular mRNA localization via cell fractionation.



Supplementary Figure S4.1: (A) Immunoblot analysis of DUX4 targets H3.X/Y/Z, MHC-I, iProteasome subunits PSMB9 and PSMB10, and localization controls GAPDH (cytoplasmic) and Histone H3 (nuclear) after nuclear and cytoplasmic fractionation. MB135iDUX4 myoblasts were treated with or without DOX for 4 hours, and 24 hours later stimulated with or without IFN γ 16 hours; harvested cells 44 hours post-DOX treatment (DOX pulse, 44 hours). **(B)** RT-qPCR analysis shows no difference in mRNA localization after a pulse of DOX. Proper localization of *RPL27* (cytoplasmic) and *HSATII* (nuclear) was observed. Data represent mean $\pm$ SD. C, cytoplasmic fraction; DOX, doxycycline; DUX4, double homeobox protein 4; IFN γ , interferon gamma; MHC-I, major histocompatibility complex class I; N, nuclear fraction; RT-qPCR, quantitative reverse transcription PCR; WCL, whole-cell lysate.

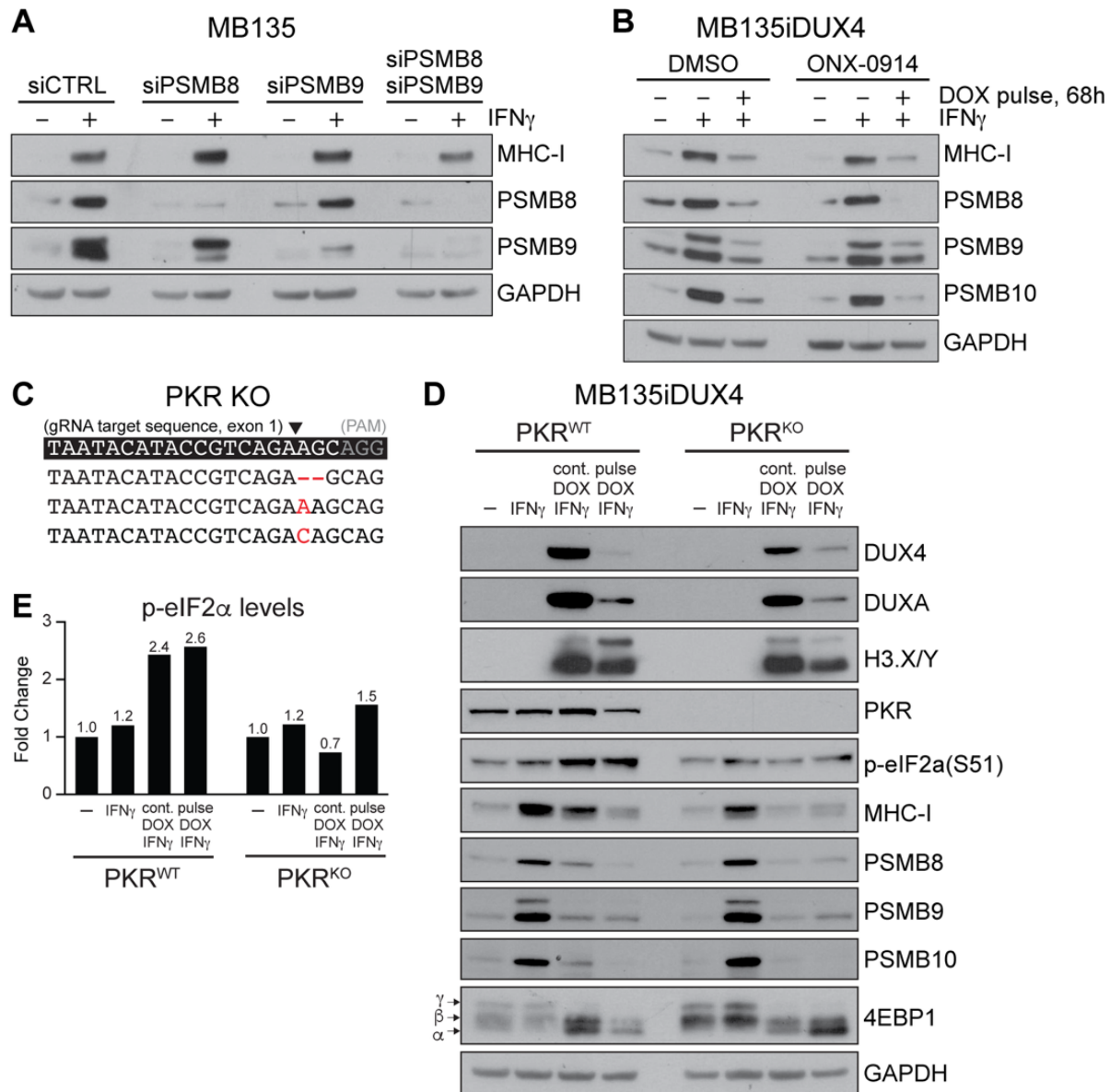
Supplementary Figure S4.2: Treatment with MG132 or Bafilomycin does not rescue DUX4-induced protein suppression of MHC-I and the iProteasome.



Supplementary Figure S4.2: (A) Immunoblot analysis of MB135iDUX4 myoblasts treated with or without a 4-hour pulse of DOX, incubated for 48 hours, stimulated with or without IFN γ , and treated with DMSO, 0.5 μ M MG132, or 0.5 μ M Baf for an additional 16 hours; harvested cells 68 hours post-DOX treatment (DOX pulse, 68 hours). DUXA is a DUX4-activated target gene. GAPDH serves as loading control. **(B)** RT-qPCR analysis shows no effect of MG132 or Baf on relative mRNA levels of housekeeping gene RPL27, DUX4-target genes DUXA and H3.X/Y, MHC-I subunits HLA-A, HLA-B,

HLA-C, or iProteasome subunits PSMB8, PSMB9, and PSMB10. Data represent mean $\pm$ SD. Baf, Bafilomycin; DOX, doxycycline; DUX4, double homeobox protein 4; IFN γ , interferon gamma; MHC-I, major histocompatibility complex class I; RT-qPCR, quantitative reverse transcription PCR.

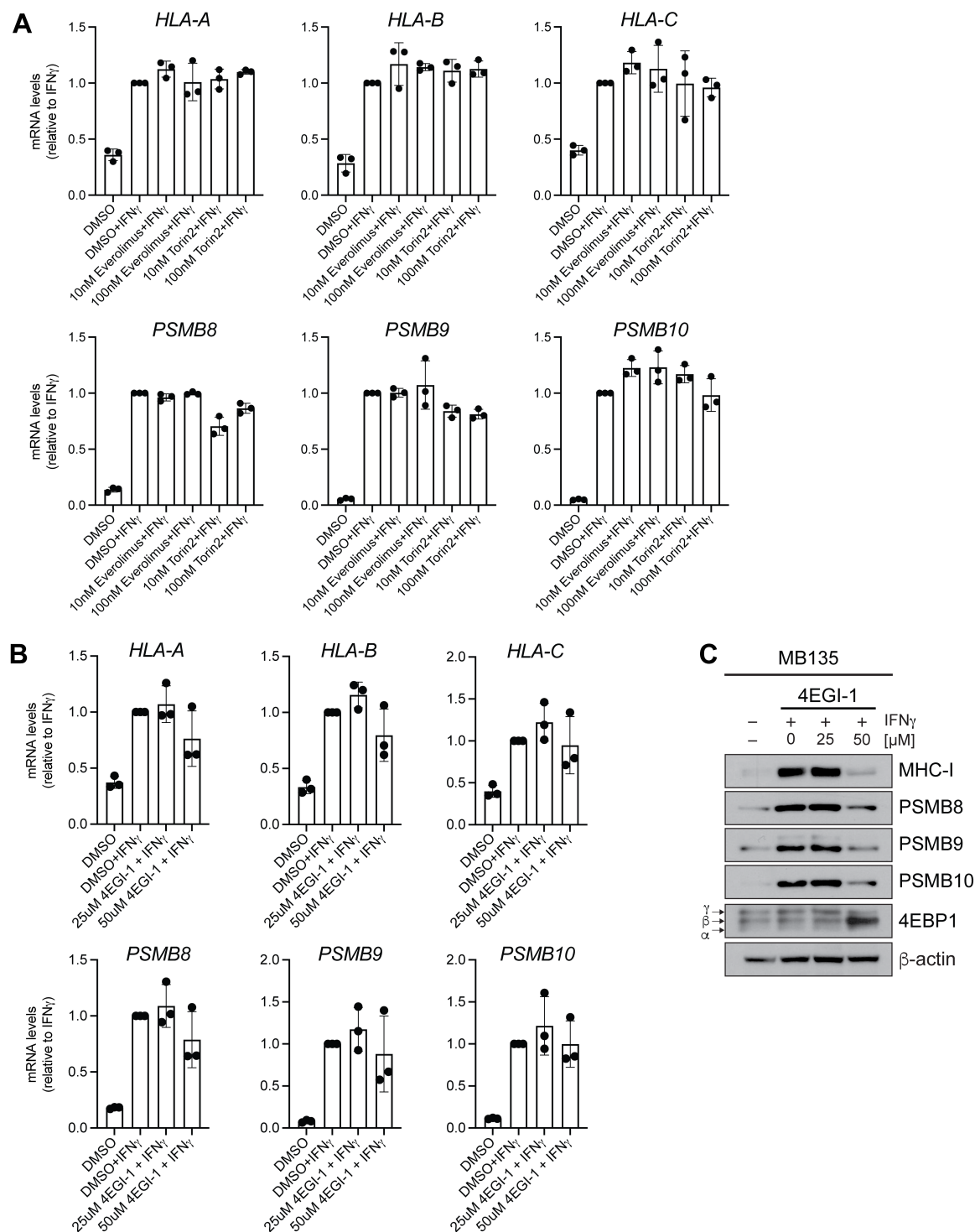
Supplementary Figure S4.3: MHC-I protein stability does not require iProteasome catalytic subunits PSMB8 or PSMB9.



Supplementary Figure S4.3: (A) Immunoblot analysis of MB135 myoblasts treated with siRNAs targeting PSMB8, PSMB9, or nontargeting siCTRL, with and without 16-hour IFN γ treatment. GAPDH serves as loading control. **(B)** Immunoblot analysis of MB135iDUX4 myoblasts treated with DMSO or PSMB8 inhibitor ONX-0914. Cells were treated with a 4-hour pulse of DOX, incubated for 48 hours, followed by treatment with 200 nM ONX-0914 and IFN γ for the terminal 16 hours; harvested cells 68 hours post-DOX treatment (DOX pulse, 68 hours). GAPDH serves as loading control. **(C)** Genotype of polyclonal CRISPR-Cas9 engineered PKR KO in MB135iDUX cells. **(D)** Immunoblot analysis of MB135iDUX4 myoblasts expressing WT PKR (left) or PKR KO (right) stimulated with IFN γ following continuous DUX4 induction or a pulse of DUX4 harvested at 68 hours. GAPDH serves as loading control. **(E)** Quantification of eIF2-alpha

phosphorylation levels in **(D)** using densitometric analysis normalized to GAPDH and graphed as fold change relative untreated samples. DOX, doxycycline; IFN γ , interferon gamma; KO, knockout; MHC-I, major histocompatibility complex class I; siCTRL, siRNA Control; WT, wild-type.

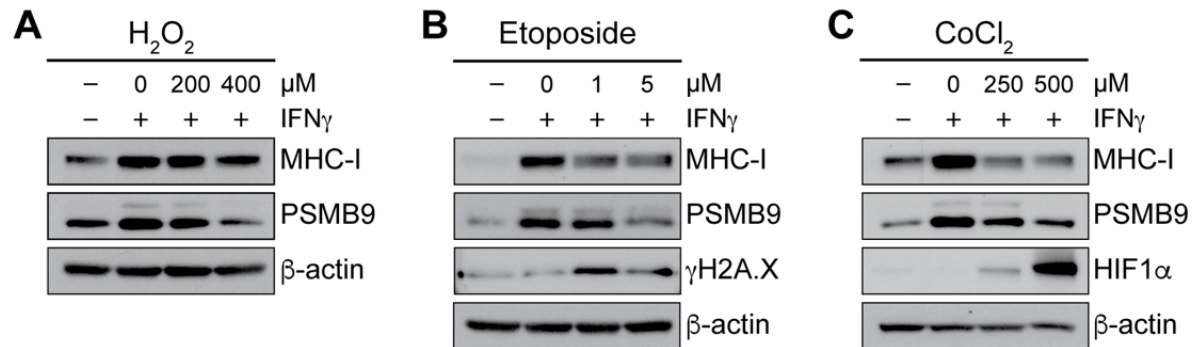
Supplementary Figure S4.4: Validation of IFN γ -induced mRNA expression levels with Everolimus, Torin2, and 4EGI-1 treatments.



Supplementary Figure S4.4: (A) RT-qPCR analysis shows no statically significant effect of Everolimus or Torin2, or **(B)** 4EGI-1 treatment, on the relative mRNA levels of MHC-I subunits HLA-A, HLA-B, HLA-C, or iProteasome subunits PSMB8, PSMB9, and

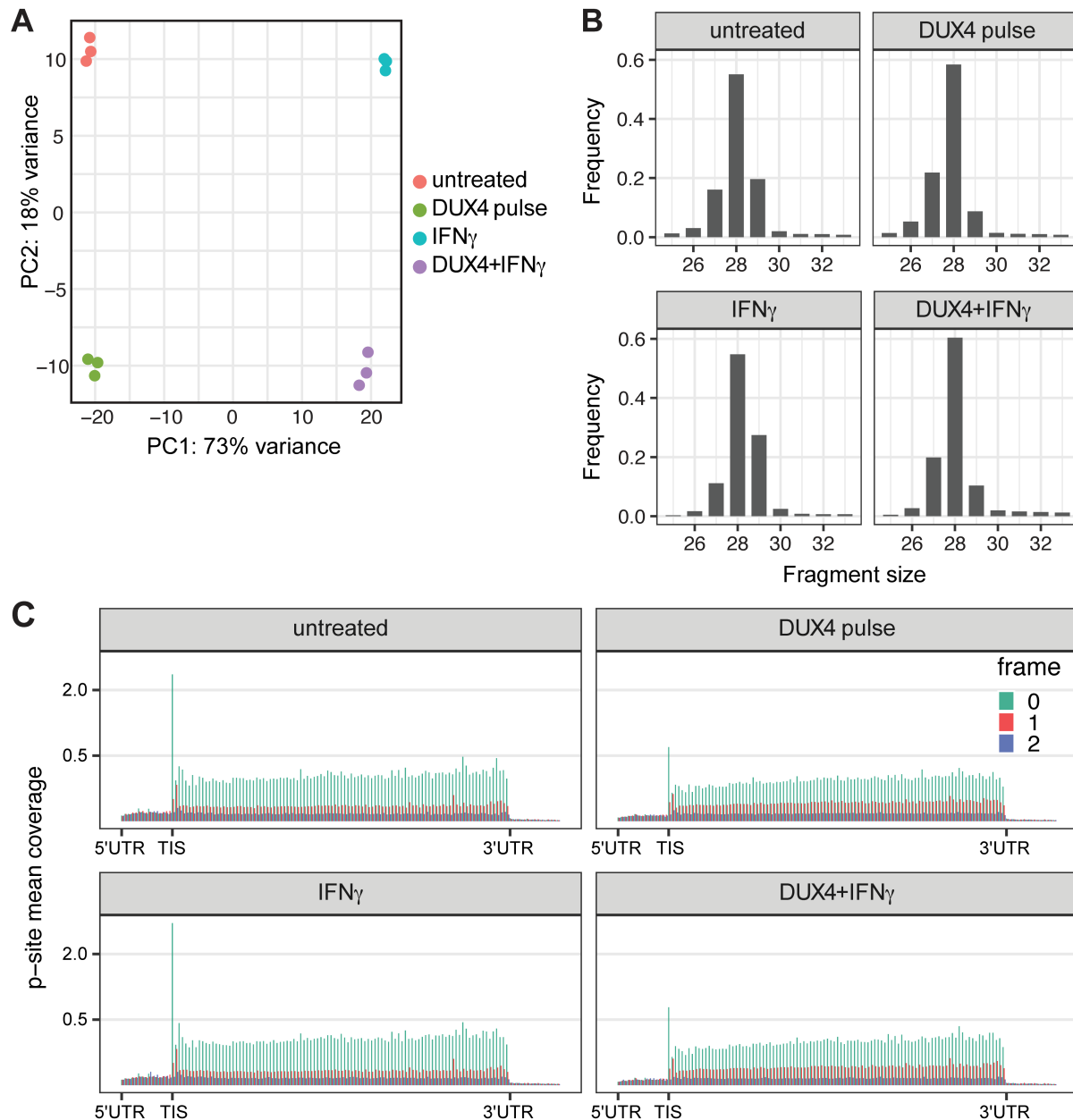
PSMB10 induced by IFN γ treatment in MB135 myoblasts. Data were normalized to RPL27 then graphed relative to IFN γ -treated samples and represent mean $\pm$ SD. **(C)** Immunoblot analysis of MB135 myoblasts treated with the eIF4E/eIF4G inhibitor 4EGI-1, with and without IFN γ . Beta-actin serves as loading control. IFN γ , interferon gamma; MHC-I, major histocompatibility complex class I; RT-qPCR, quantitative reverse transcription PCR.

Supplementary Figure S4.5: Oxidative stress, DNA damage, and hypoxia-induced cell stress pathways moderately suppress MHC-I and PSMB9.



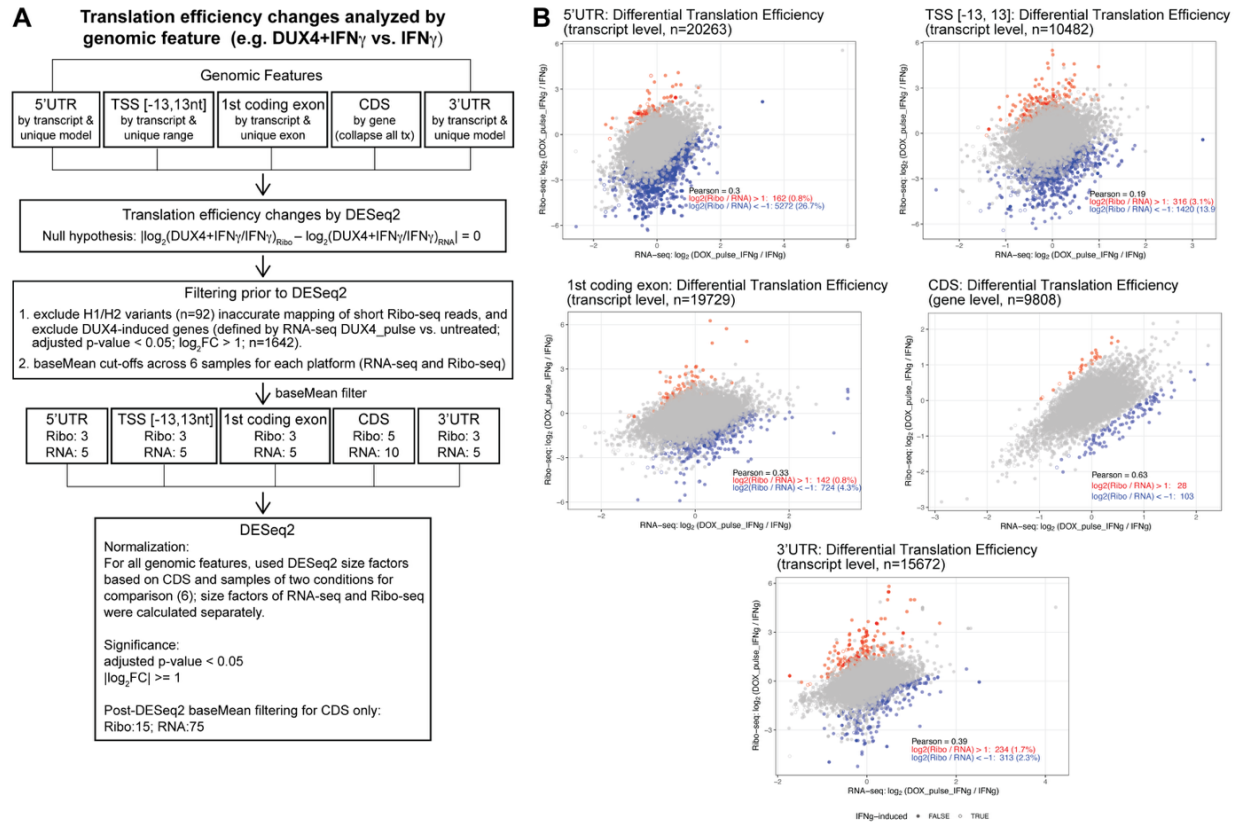
Supplementary Figure S4.5: (A) Immunoblot analysis of MB135 myoblasts treated with hydrogen peroxide (H₂O₂) to induce oxidative stress, **(B)** etoposide to induce DNA damage, or **(C)** cobalt chloride (CoCl₂) to mimic hypoxia. Cells were treated with stress-inducing reagents for 24 hours, followed by a 16-hour incubation in media re-supplemented with cell stress-inducing reagent plus IFN_γ. Beta-actin serves as loading control. IFN_γ, interferon gamma; MHC-I, major histocompatibility complex class I.

Supplementary Figure S4.6: Quality control analysis of Ribo-seq data



Supplementary Figure S4.6: (A) PCA of gene expression in Ribo-seq biological replicates, $n = 3$. **(B)** Length distribution of RPFs. All downstream analysis was restricted to the dominant fragment size of 26–29 nt. **(C)** Metagene coverage of P-sites in 3 different reading frames (green, red, blue) over 5' UTR, TIS, CDS, and 3' UTR regions; data represent the average of biological replicates for each treatment condition, $n = 3$. CDS, coding sequence; nt, nucleotide; PCA, principal component analysis; RPF, ribosome-protected fragment; TIS, translation initiation site.

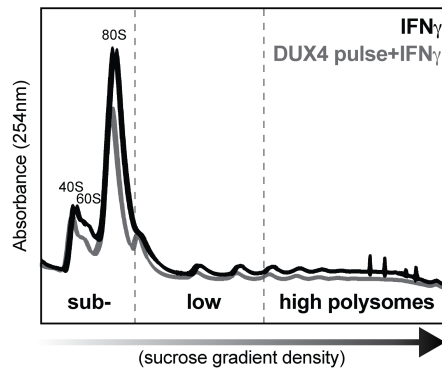
Supplementary Figure S4.7: Genome-wide quantification of ribosome-protected mRNA fragments using RNA sequencing (RNA-seq) and ribosome profiling (Ribo-seq).



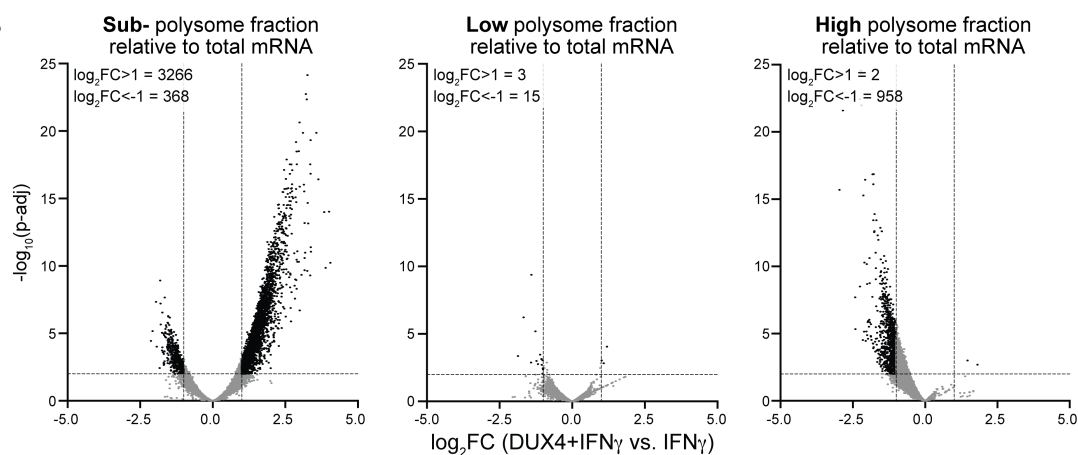
Supplementary Figure S4.7: (A) Workflow of differential analysis and filtering prior to DESeq2 analysis to determine translation efficiency at genomic features. **(B)** Scatter plots of reads aligning to annotated genomic features at the transcript or gene level as indicated. Log2 fold-change values represent the average of biological replicates for DUX4 pulse+IFN γ condition relative to IFN γ treatment alone in MB135iDUX4 myoblasts, n = 3. Translationally up-regulated mRNAs (red) and translationally down-regulated mRNAs (blue) are highlighted ($|\log_2\text{FC}| > 1$, p-adj<0.05, n = 3). DUX4, double homeobox protein 4; IFN γ , interferon gamma; RNA-seq, RNA sequencing.

Supplementary Figure S4.8: Polysome profiling and 5' UTR analysis.

A MB135iDUX4



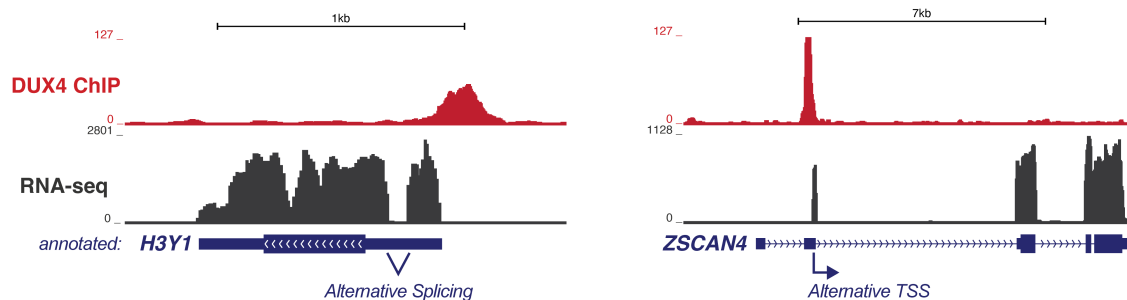
B



C Gene name 5'UTR annotated sequences

RPL10A CUCUUUCCGGUUAGCGCGCGUGAGAAGCCAUGAGC
RPL4 CUUUUCCUGUGGCAGCAGCCGGGUGAGAGGAGCGUGGCUGUCUCCUCUC
RPS6 CUCUUUCCGUGGCGCCUCGGAGGCGUUCAGCUGCUUCAAGAUGAAG
RPS15A CUCUUUCCGCCAUCUUUCCGCGCCGCCACAAUGGUG

D



Supplementary Figure S4.8: (A) Absorbance at 254 nm across a density gradient fractionation system. Traces represent biological replicates for each treatment condition, $n = 3$ (black = IFN γ ; gray = DUX4 pulse+IFN γ harvested at 68 hours). **(B)** Differential RNA-seq analysis of sub-, low, and high polysome fractions relative to total mRNA levels. Volcano plot showing log₂ fold-change differential abundance in cells treated with DUX4 pulse+IFN γ versus IFN γ (significance defined as $\text{basemean} > 50$, $|\log_2\text{FC}| > 1$, $p\text{-adj} < 0.01$, $n = 3$). **(C)** 5' UTR sequences of select TOP mRNAs (blue = TOP motif). **(D)** Two representative direct DUX4-target genes with alternative 5' UTRs based on RNA-

seq alignment shown relative to the annotated transcripts. DUX4, double homeobox protein 4; IFN γ , interferon gamma; RNA-seq, RNA sequencing; TOP, terminal oligopyrimidine.

Chapter 5: Designing a lineage tracing system to track *DUX4*-expressing cells and their progeny

This chapter contains unpublished work.

Introduction

The transcription factor *DUX4* has been identified as a novel driver of immune evasion in many solid cancers. *DUX4* expression in cancer cells inhibits major histocompatibility complex class I (MHC-I) antigen presentation and is correlated with poor patient survival and response to immune checkpoint blockade therapies [100,103]. *DUX4* and its transcriptional programs, in addition to a number of *DUX4*-fusions, have been identified in cancers [105,214]. Recent single cell RNA-seq (scRNA-seq) analysis from our lab has revealed that in several cancer cell lines, there is sporadic expression of full length *DUX4* in 0.1-10% of cells, leading to expression of a totipotent-like program [105]. This *DUX4* induction can be increased by treatment with several cancer therapeutics, such as etoposide and doxorubicin, and nutrient starvation, a common tumor microenvironment stressor [105,215]. Despite this induction of a totipotent program exacerbated by common cancer interventions, the long-term effects of *DUX4* expression on tumor pathogenesis are unclear.

Activation of *DUX4* and its ZGA program in somatic cells leads to a wide variety of consequences, including oxidative stress, inhibition of RNA quality control, stabilization of alternative mRNAs, translational reprogramming (See **Chapter 4** [99]), modulation of immune responses, and induction of retroelements, protein aggregates, cancer testis antigens, and double-stranded RNAs (dsRNAs) [110]. Prolonged *DUX4* target induction eventually leads to a progressive muscle wasting phenotype in FSHD [28,71,163,216,217], while expression in cancers is correlated with a poorer response to immune therapies and a higher prevalence of remission [100,103]. The fate of *DUX4*-expressing cancer cells, however, is unknown. Because a pulse of *DUX4* is critical for normal development and its expression in cancer correlates with immune evasion, it has been hypothesized that cancer cells also express *DUX4* in a short burst and do not progress to apoptosis. Therefore, characterization of the long-term effects of this transient burst of *DUX4* expression is of utmost importance.

Recent works have focused on long-term, post-transcriptional effects of *DUX4*; *DUX4* induced post-transcriptional suppression of over 200 proteins despite no change in regulation at the RNA level [95]. *DUX4* inhibition of nonsense-mediated decay machinery also leads to production of nonfunctional truncated proteins, leading to

protein loss [96,97,192]. As shown in **Chapter 4**, for several days after a pulse of DUX4, translation is inhibited and MHCI and immunoproteasome (iP) proteins are suppressed upon addition of interferon gamma (IFN γ), while DUX4 target genes escape this suppression [99]. Interestingly, expression of Dux targets in mouse embryonic stem cells (mESCs) also leads to a translationally suppressed state [153,154]. Because translational inhibition is also correlated with increased cancer cell plasticity [107,108], DUX4 inhibition of translation in a subset of cancer cells could lead to increased viability in response to harsh environments, nutrient deprivation, and therapeutic intervention. Characterization of these DUX4-positive cells will provide valuable insight into the long-term effects of induction during development and their role in cancer pathogenesis.

In this chapter, attempts to build a sensitive, yet specific lineage tracing system to track DUX4-expressing cells and their progeny are described. In any given cell culture -cancer, FSHD, or iPSC lines- it is unknown whether cells have cycled through a DUX4-positive state if DUX4 or its target proteins are no longer present, which is likely given the brief, pulse-like nature of *DUX4* expression in embryonic development. Moreover, it has yet to be shown that cells expressing endogenous DUX4 survive. Lineage tracing is a powerful technique that is widely used for permanently tracking a desired subpopulation of cells and their progeny. This technique is often used to study cell fate decisions during development as well as in heterogeneous tumors [218]. Development of a lineage tracing system that can track cells expressing transcriptionally active DUX4 will allow for further study of its downstream effects in a variety of systems, from FSHD and cancer to early embryonic development. Though the attempts in this chapter have not established a robust system, there is hope that these techniques can be refined in future attempts to develop this system.

Results

Design of a classical tamoxifen-inducible Cre recombinase-based system to track DUX4-expressing cells and their progeny.

To date, there has been no method to permanently label DUX4-positive cells so cell progeny can be tracked for multiple cell divisions after expression. Because DUX4 induces transcription of its distinct ZGA target genes through direct DNA binding to their promoter regions [2,3,111], some studies [67,111,138,139] have had success using reporter systems driven by DUX4 target gene promoters, such as *ZSCAN4*. These systems provide a snapshot by expressing the reporter only when DUX4 is present. However, there is currently no means to permanently track what happens to DUX4-positive cells in the long term or identify cells that have previously expressed *DUX4*.

To design a lineage tracing reporter for long-term use, a Cre recombinase-based dual-fluorescence reporter system was designed [219]. The “Stoplight” system uses a constitutive promoter to drive expression of a Cre recognition site (loxP)- flanked

DsRED containing a stop codon directly upstream of a *GFP* sequence [220]. Cells with no Cre recombinase will continue to express DsRED alone, while Cre activation will facilitate excision of DsRED and drive constitutive GFP expression. The DUX4 specificity is provided by the presence of a Cre recombinase driven by DUX4-specific target gene promoters, such as *H3Y1* or *LEUTX*, which are not known to have any other transcriptional activators and have DUX4 ChIP-seq peaks in their promoter regions [111,156,162]. Finally, an element of temporal regulation is necessary to ensure only cells that have expressed *DUX4* within a given window of time are tracked, avoiding muddling data with cells that expressed DUX4 at times or passages not relevant to our study. For these reasons, I employed Cre recombinase fused to two estrogen receptor (ERT2) sequences, which ensure Cre recombinase is only active in the presence of 4-hydroxytamoxifen (4'OH-T) [221]. The ERT2CreERT2 system is preferred over other CreER-based systems due to increased specificity and a reduction in background excision events [222]. In summation, lentiviral incorporation of a “Stoplight” construct and a DUX4 target gene-driven ERT2CreERT2 will provide a temporally regulated lineage tracing system; cells that express transcriptionally active DUX4 will induce expression of ERT2CreERT2, which is only enzymatically active in the presence of 4'OH-T (**Fig. 5.1A**). Expression of ERT2CreERT2 in the presence of 4'OH-T will allow for permanent excision of the DsRED sequence in the Stoplight construct, inducing GFP expression in that cell and its progeny.

As an pilot test of this ERT2CreERT2- based lineage tracing system, *H3Y1 promoter-ERT2CreERT2*, *LEUTX promoter-ERT2CreERT2*, and constitutive positive control for expression *PGK promoter-ERT2CreERT2* lentiviral constructs were transduced into three cell lines shown to express transcriptionally active *DUX4* [105]: rhabdoid kidney tumor line G401, FSHD type 2 myoblast line MB200, and inducible DUX4 unaffected myoblast line MB135iDUX4ca#7. These lines were also transduced with the “Stoplight” reporter construct (**Fig. 5.1A**). Cells were treated for three days with 4'OH-T to track cells that express *DUX4* and induce expression of the target gene-ERT2CreERT2 reporters or constitutive control *PGK-ERT2CreERT2* cells to excise DsRED and induce *GFP* expression.

In G401 cells, positive control PGK-ERT2CreERT2 polyclonal cells had very little GFP expression and did not show more GFP expression upon 4'OH-T addition, though a population of GFP+ cells persisted in culture, indicating poor transduction efficiency and/or poor temporal 4'OH-T control (**Fig 5.1B**). *H3Y1* and *LEUTX* promoter-ERT2CreERT2 cells showed promise as lineage tracing system, as flow cytometry showed treatment with 4'OH-T caused a modest increase in GFP positive cells, with levels reflecting previous studies where 0.1-10% of cells spontaneously express *DUX4* [105] (**Fig. 5.1B**). RT-qPCR analysis of PGK-ERT2CreERT2 and *LEUTX-ERT2CreERT2* cells unsorted or FACS sorted for GFP+ cells revealed an increase in *GFP* expression after 4'OH-T treatment, though there was no increase in *DUX4* or

target gene *H3Y1* or *ZSCAN4* expression in LEUTX-ERT2CreERT2 GFP⁺ sorted cells (**Fig. 5.1C**), indicating nonspecific activation of Cre recombinase.

FSHD myoblast line MB200 cells showed more promise in temporally regulated excision of DsRED and subsequent *GFP* expression. PGK-ERT2CreERT2 lines showed the highest increase in percentage of GFP⁺ cells upon 4'OH-T treatment, though low transduction efficiency likely limited excision rates (**Fig. 5.1D**). In MB135iDUX4#7 polyclonal lineage tracing cell lines treated with doxycycline (DOX) to induce DUX4 for three days, any cell line treated with 4'OH-T showed an increase in GFP⁺ cells, regardless of whether cells were treated with DOX (**Fig. 5.2E**). This indicates temporal regulation of ERT2CreERT2 with 4'OH-T is robust, but cells are leaky and not responding specifically to DUX4-induction of ERT2CreERT2. MB135iDUX4ca cells have been shown to exhibit leaky DUX4 expression [99], so this line may not be suited to assessing activity of the lineage tracing system.

Treatment with doxorubicin or etoposide increases DUX4 expression and the number of GFP⁺ cells in the polyclonal line, but not in monoclonal lines

Because endogenous *DUX4* is expressed in such a small percentage of cells (~0.1% in FSHD cell lines [9,62,63]), determination of effective lineage tracing above levels of leakiness may be difficult without artificial induction of DUX4. To increase the proportion of *DUX4*-expressing cells without the dynamic regulation of *DUX4* upon myoblast differentiation [117,118], G401 polyclonal lineage tracing lines were pulsed with 1μM doxorubicin (Doxor) for six hours before release into media containing 2μM 4'OH-T for three days, with a growth media change and 4'OH-T replenishment every 24 hours (**Fig. 5.2A**). Flow cytometry of G401 lineage tracing cells revealed almost no excision of DsRED and expression of *GFP* in cells not treated with doxorubicin, though there was a minor increase in GFP⁺ cells upon 4'OH-T treatment; this reflects rare *DUX4*-expressing cells in culture and shows promise for 4'OH-T-mediated temporal control (**Fig. 5.2B**). Doxorubicin treatment without 4'OH-T increased the percentage of GFP⁺ cells from 0.02% to 2.55%, while doxorubicin treatment in the presence of 4'OH increased the percentage of GFP⁺ cells from 0.2% to 3.8% (**Fig. 5.2B**). Though the increase in GFP⁺ cells in Doxor + 4'OH-T G401 *H3Y1 promoter*-ERT2CreERT2 polyclonal lines is promising, the level of leakiness of Cre recombinase excision activity without 4'OH-T is concerning. To determine if GFP⁺ cells were truly *DUX4* and *DUX4* target gene positive, GFP⁺ cells were sorted via FACS in Doxor + 4'OH-T G401 *H3Y1 promoter*-ERT2CreERT2 polyclonal lines and RNA expression via RT-qPCR was compared to unsorted populations of other treatments (**Fig. 5.2C**). GFP⁺ sorted cells showed the highest expression of *GFP* and *Cre*, but they also showed the highest expression of *DsRED*. Promisingly, levels of *DUX4 exon 1*, *H3Y1*, and *ZSCAN4* were higher in cells treated with doxorubicin, indicating *DUX4* expression can be increased by doxorubicin treatment for further lineage validation. Temporal regulation mediated by

4'OH-T was leaky, however, as Doxor + EtOH vehicle cells also showed high levels of *GFP* expression. This leakiness could be a product of a polyclonal line.

To determine if leaky expression of *Cre* could be prevented by monoclonal cell line generation, G401 *H3Y1 promoter*-ERT2*Cre*ERT2 cells were single cell sorted via FACS, expanded, and flow cytometry was performed to ensure clones did not have leaky *Cre* activity and remained DsRED+/GFP-. Of all clones screened, many showed DsRED-/GFP+ populations, though a few remained DsRED+/GFP-, such as clone #4 (**Fig. 5.2D**). To test if these clones were responsive to *DUX4* induction and temporally regulated by 4'OH-T, monoclonal cells were treated with etoposide (etop) and 4'OH-T according to the doxor schedule in **Fig. 5.2A**. Etop treatment increased the percentage of GFP+ cells, but there was no temporal control upon addition of 4'OH-T (**Fig. 5.2D**), indicating the 4OH-T was not working properly or ERT2*Cre*ERT2 was still active in its absence.

A doxycycline-inducible Cre recombinase lineage tracing system to tightly control Cre expression and stoplight excision.

Because the ERT2*Cre*ERT2 system seemed to have leaky *Cre* activity in the absence of 4'OH-T, we sought to redesign the lineage tracing system to temporally control *Cre* at the level of expression instead of enzymatic activation using a doxycycline-inducible mechanism [223]. In this system the lentiviral dual fluorescence “Stoplight” reporter was used with the puromycin resistance cassette removed; cells were then selected by FACS sorting for DsRED+/GFP- cells (**Fig 5.3A**). Instead of random lentiviral integration, *Cre* recombinase, under the control of a TetON doxycycline (DOX) inducible promoter, was knocked into the AAVS1 safe harbor site [116] and selected for proper integration with puromycin (**Fig. 5.3A**). Finally, lentiviral constructs with constitutive (*PGK*) or *DUX4* target gene (*H3Y1*) promoters driving reverse tetracycline-controlled transactivator (rtTA) expression were incorporated and selected with blasticidin (**Fig. 5.3A**). The system functions as follows: in the presence of *DUX4*, the *H3Y1* promoter drives rtTA expression, which will not induce activation of *Cre* recombinase unless doxycycline is also present, providing a level of temporal regulation. Doxycycline induces a conformational change in rtTA, allowing for binding and activation of the TetON promoter. Therefore, only cells expressing *DUX4* in the presence of doxycycline will express *Cre recombinase*, leading to excision of DsRED from the Stoplight construct and permanent expression of *GFP* (**Fig. 5.3B**). The constitutive *PGK*-rtTA construct is always expressing rtTA and will express *Cre* and *GFP* only in the presence of doxycycline.

A pilot test of this system was performed in G401 polyclonal lines that were treated with doxor or etop for 6 hours to induce *DUX4* expression and treated with doxycycline for three days before harvest for flow cytometry and RT-qPCR analysis (**Fig. 5.3C**). In the constitutive *PGK*-rtTA cells, there was a large presence of leaky

GFP+ cells without DOX treatment (~24% of the population), but there was a substantial increase in GFP+ cells (~35% of the population) upon DOX treatment (**Fig. 5.3D**). In the *H3Y1 promoter*-rtTA cells, all conditions had GFP+/DsRED- leaky excision in approximately ~20% of the population (**Fig. 5.3D**). Treatment with DOX in all conditions also leads to higher percentages of GFP+ cells, with a ~2% increase in untreated, ~6% increase in etop treated, and a ~15% increase in doxor treated cells, aligning with changes in numbers of *DUX4*+ cells seen previously [105] (**Fig. 5.3D**). RT-qPCR analysis showed a strong increase in *DUX4*, *H3.X/Y*, *Stoplight* and excised *Stoplight* upon etop and doxor treatment, as seen in the ERT2CreERT2 system (**Fig. 5.3E**, **Fig. 5.1C**, **Fig. 5.2C**). Excised *Stoplight* levels are higher in DOX-treated samples, however (**Fig. 5.3E**). Interestingly, *Cre recombinase* levels are substantially higher in all DOX+ samples in the *PGK*-rtTA cells and only in the doxor and etop DOX+ samples in the *H3Y1*-rtTA cells, indicating the promoter-driven expression of rtTA is correct and the DOX is providing temporal regulation (**Fig 5.3E**). Because DsRED and GFP fluorescent proteins are stable, long-lived proteins in culture [224], we sought to determine if GFP+/DsRED- (fully excised *Stoplight*) or GFP+/DsRED+ (newly excised *Stoplight* with residual DsRED) were enriched for *DUX4* and target genes. FACS sorting followed by RT-qPCR analysis of these populations in DOX + doxor treated cells (**Fig. 5.3D**) revealed high levels of *Stoplight excised* in both populations, but substantially higher *Cre Recombinase* expression in the GFP+/DsRED+ population (**Fig. 5.3F**). Interestingly, both GFP+/DsRED- and GFP+/DsRED+ populations showed high levels of *DUX4* expression, while expression of *DUX4* target *H3.X/Y* was higher in the double-positive population (**Fig. 5.3F**).

Monoclonal selection was ineffective in producing constitutive control lines but revealed marginal success in DUX4 lineage tracing lines.

Because there was a large population of GFP+ cells in the -DOX conditions of both the *PGK*-rtTA and *H3Y1 promoter*-rtTA polyclonal G401 lines, a reduction in DOX-independent *Cre* expression was imperative. Creation of monoclonal lines that can be assured to have the correct integration of all three lineage tracing constructs was performed. Polyclonal lines were single cell FACS sorted, expanded, and screened for no aberrant *GFP* expression in the absence of DOX. In the constitutive *PGK*-rtTA monoclonal lines, several showed low levels of aberrant GFP expression (**Fig. 5.4A**). When DOX was added to test activation of *Cre Stoplight* excision, however, the clones were not responsive and did not induce excision (**Fig. 5.4A**). Additionally, only one clone (#5) showed expression levels of rtTA above background, indicating improper functionality of the *PGK*-rtTA construct (**Fig. 5.4B**).

Screening the *H3Y1*-rtTA monoclonal lines revealed two clones with minimal leaky GFP expression, though 9-20% of the population was still GFP+ (**Fig. 5.5A**). To test responsiveness, the clones were treated with the doxor and DOX time course in

Fig. 5.3C. In both clones, there was a drastic increase in Stoplight excision upon doxor + DOX treatment (**Fig. 5.5A**). When GFP-/DsRED+ and GFP+/DsRED+ cells in doxor + DOX samples were FACS sorted and RNA levels were compared to other treatment conditions, the GFP+/DsRED+ populations unsurprisingly had the highest levels of *Stoplight excised* and *Cre recombinase* (**Fig. 5.5B**). Encouragingly, the doxor + DOX GFP+/DsRED+ cell populations had substantially higher levels of *rtTA*, *DUX4*, and *H3Y1* than their GFP-/DsRED+ counterparts, indicating that activation of *DUX4* in individual cells was inducing *rtTA* and *Cre* excision activity, and that cells that did not activate *DUX4* were not excising the Stoplight construct (**Fig. 5.5B**).

Another recently developed method, dubbed “Catch and Release” or “release from confluence”, has been shown to induce endogenous *DUX4* expression in cancer cells [105]. Cells are incubated in the same media for 11 days before trypsinization and seeding into fresh media at low confluency. *H3Y1-rtTA* monoclonal lines were subjected to this protocol and harvested at 24-, 48-, and 72- hours post “release” for flow cytometry and RT-qPCR analysis (**Fig. 5.6A**). Flow cytometry revealed increasing percentages of GFP+ cells specifically in the +DOX condition (**Fig. 5.6B**). At the RNA level in Clone #8, however, *Cre recombinase* levels were equal in both – and + DOX conditions at all time points, indicating a lack of temporal regulation of *rtTA*-mediated expression (**Fig. 5.6C**). In Clone #11, *Cre recombinase* levels were substantially higher in the +DOX condition, though the levels of *rtTA* were higher at 0hr than 24hr, despite higher levels of *DUX4* expression at 24hr (**Fig. 5.6C**), indicating that higher levels of *rtTA* and *Cre* might not be specific to *DUX4*-mediated induction.

Antisense oligonucleotide (ASO)-mediated knockdown of DUX4 partially decreases Cre expression in G401 H3Y1-rtTA lineage tracing monoclonal lines.

To determine the sufficiency of *DUX4* in activation of *Cre* and stoplight excision in G401 *H3Y1-rtTA* monoclonal lines, an antisense oligonucleotide (ASO)-mediated knockdown of *DUX4* was performed in a doxor + DOX lineage tracing time course (**Fig. 5.7A**). In clone #8, flow cytometry of GFP and DsRED fluorescence showed the largest increase in the proportion of GFP+ cells in the DOX + doxor treatment condition, as previously shown. KD of *DUX4* modestly decreased the proportion of GFP+ cells in each condition (**Fig. 5.7B**). In clone #11, however, a sharp increase in the percentage of GFP+ cells was present in the +DOX -doxor condition, unlike previous experiments. *DUX4* KD only decreased the percentage of GFP+ cells in the DOX + doxor condition (**Fig. 5.7B**). In both clones, KD of *DUX4* revealed a substantial reduction in *DUX4* and *H3Y1* transcript levels (**Fig. 5.7C**). Interestingly, though *DUX4* KD reduced *Cre recombinase* expression, it had little to no impact on levels of *rtTA* (except for clone #8 Doxor -DOX CTRL KD, which displayed unusually high amounts of *rtTA*) and *excised Stoplight* (**Fig. 5.7C**). Additionally, the levels of *rtTA* in Clone#11 showed the inverse pattern across treatments when compared to previous experiments, with *rtTA* levels

decreasing upon doxor treatment (**Fig. 5.7C**). Ultimately this experiment shows that even monoclonal lines have expression of *rtTA* and *Cre* that is not entirely due to *DUX4* expression.

Design and pilot trial of endogenous DUX4 target gene knock-in of lineage tracing machinery.

To prevent aberrant activation of *DUX4 target promoter-rtTA* due to random lentiviral integration, we designed a system to knock in the *rtTA* specifically at the endogenous locus of *DUX4* target genes *H3Y1* and *ZSCAN4*, ensuring native-like chromatin signatures and the presence of potential *DUX4*-interaction or regulatory sequences of the endogenous target gene. To create a donor template for homologous recombination, homology arms were designed directly up and downstream of the target gene start codon, with the *rtTA* coding region template aligning with the start codon of the endogenous gene (**Fig 5.8A**). This KI cassette also contained an FRT-flanked PGK-BlastR sequence, allowing for selection of cells with blasticidin and later excision of the active promoter that may induce leaky expression with transient transfection of Flippase Recombinase (FLPe) (**Fig. 5.8A**). The Stoplight and *Cre* recombinase components of the lineage tracing system remain the same as in (**Fig. 5.3A, B**).

After selection for all lineage tracing elements, MB135 (unaffected) and MB073 (FSHD1) myoblast polyclonal cell lines were subjected to a six-hour pulse of etoposide, followed by three days of DOX treatment (**Fig 5.8B**) to test for any responsiveness in the lineage tracing system in cells with little to no (MB135) or high (MB073) endogenous *DUX4* expression. MB135 cells with the *H3Y1-rtTA* KI showed very slight increases in the percentage of GFP⁺ cells in cells treated with DOX regardless of etop treatment, while the *ZSCAN4-rtTA* KI cells showed higher increases in the +DOX cells without etop treatment than any other condition (**Fig. 5.8C**). MB073 cells with *ZSCAN4-rtTA* KI followed this leaky trend, with the +DOX cells in the absence of etop showing the highest increase in GFP⁺ cell populations (**Fig 5.8C**). The MB073 cells with *H3Y1-rtTA*, however, showed promise, with a minor (~2%) increase in GFP⁺ percentages upon DOX treatment in cells without etop, and a ~4% increase upon DOX treatment in etop pulsed cells (**Fig 5.8C**). Ultimately, in this polyclonal system, there is a population of leaky cells that express GFP in the absence of DOX. Screening monoclonal lines to ensure proper integration of every part of the lineage tracing system will allow for thorough assessment of this system.

Discussion

The development of a lineage tracing system to permanently track *DUX4*⁺ cells and their progeny is of tantamount importance to address several open questions in *DUX4* biology: 1) Do somatic cells that have expressed *DUX4* survive and divide in

culture? 2) What proportion of cancer or FSHD cells are cycling through a DUX4 positive state at a given time in culture? 3) What properties, such as altered replication or differentiation capacity, stress tolerance, or induction of stem cell-like transcriptional/post-transcriptional programs do *DUX4*-expressing cells and their progeny maintain? and, 4) What is the competitive fitness of DUX4-expressing cells in cancer, development, and FSHD? In this chapter, iterations of potential lineage tracing systems and their varying success were outlined, including tamoxifen-inducible and doxycycline inducible Cre recombinase systems in lentiviral and knock-in based constructs. Though no system to date is fully robust in terms of temporal regulation and DUX4 specificity, this chapter describes important groundwork for future development of a lineage tracing system.

Classical lineage tracing systems involve specific activation, usually by cell-type specific promoters, of Cre recombinase, leading to a trackable change in exogenous constructs by creation of a molecular “switch” to detect cells that have activated those cell-type specific reporters [219]. In all proposed lineage tracing systems described here, Cre recombinase is used to alter a “Stoplight” construct which in the absence of Cre constitutively expresses DsRED and after Cre activity, excises DsRED and constitutively expresses GFP [220]. To add a level of temporal regulation, a (4-hydroxy) tamoxifen-inducible ERT2CreERT2 was used [221], which is a standard approach in *in vivo* studies in mice [225] as well as in cell culture studies. In the context of DUX4 lineage tracing, this system showed twofold problems in successful lineage tracing: inability to activate in the presence of 4’OH-T and leaky Cre excision and GFP expression without 4’OH-T administration. The first might be due to the insolubility of 4’OH-T in water and EtOH and the low solubility in DMSO. In our hands, the compound crashed out of solution in most solvent conditions, and it is difficult to tell if the compound remains in solution once added to culture media. Despite this, administration of 4’OH-T led to a modest increase in GFP+ cell populations in lineage tracing lines (**Fig 5.1, 5.2**). The second problem with this system was an activation of Cre Recombinase activity and Stoplight excision in the absence of 4’OH-T, indicating leaky activity not controlled temporally by 4’OHT. In this lineage tracing system, ERT2CreERT2 is expressed in the presence of DUX4, and the estrogen receptor ERT2 segments should induce a tighter temporal regulation requiring 4’OH-T. Our data indicate activation of Cre in the absence of 4’OH-T, which was not improved by monoclonal selection in our study. Interestingly, estrogens were shown to antagonize DUX4 activity and reduce transcriptional activity of DUX4 in FSHD cells [226], indicating the administration of 4’OH-T may add confounding variables to tracking endogenous DUX4 expression.

The culmination of concerns with tamoxifen-inducible lineage tracing led to a complete redesign of the lineage tracing system in DUX4-expressing cells. We first adopted a doxycycline-inducible lineage tracing system for expression of Cre only when DUX4 could induce expression of rTA through lentiviral-delivered target gene promoter

constructs and when DOX was added to the cells [223]. This design prevented Cre expression unless DUX4 and DOX were present, leading to regulation at the level of Cre expression, not Cre activity. Moderate success was achieved in polyclonal lines, leading to monoclonal selection. Unfortunately, none of the constitutively controlled *PGK*-lines showed any specific and robust activation of rtTA. Of note, constitutive expression of Cre has been shown to cause reduced proliferation and an increase in cell death [227,228], implying that the clones with properly integrated lineage tracing systems may not have survived to be screened and that a constitutive DOX-inducible temporal control might not be feasible. Nevertheless, all clones from DUX4 target-rtTA G401 lines showed some leaky activation of *Cre*, indicating the system may be flawed due to improper activation of *Cre Recombinase* expression. One reason for this could be due to the properties of lentiviral integration into the host genome, where open, active regions are more accessible and more likely to be lentiviral integration sites [229]. This active chromatin environment may contribute to nonspecific activation of rtTA in our lineage tracing system.

In the final lineage tracing system covered in this chapter, we attempted to combat the leakiness of rtTA expression from randomly integrated lentiviral DUX4 target-rtTA lines by knock in of rtTA into the coding region of DUX4 targets. Adapting the donor plasmid template for AAVS1-mediated KI of silencing reporter constructs shown in **Chapter 2**, donor templates containing homology arms for DUX4 targets *ZSCAN4* and *H3Y1* were created and used to knock rtTA into the coding region for these genes in unaffected and FSHD myoblasts, along with the rest of the lineage tracing system. Though the polyclonal lines showed leaky Stoplight and GFP excision, the H3Y1-rtTA KI lines showed appropriate changes in percentages of GFP⁺ cells in the presence of etoposide and DOX. We hope that screening monoclones will ensure cell lines with proper integration of the whole system and provide a platform for long term tracing of DUX4-positive cells and their progeny.

Materials and Methods

Cell culture

G401 cells were grown in McCoy's 5A Modified Media (Gibco, # 16600082), supplemented with 10% fetal bovine serum (FBS) (Cytiva Hyclone, #SH30396.03) and 1% penicillin/streptomycin (Gibco, #15140122). MB135, MB135iDUX4ca#7, and MB200 cells were grown in Ham's F-10 Media (Gibco, #11550043), supplemented with 10% fetal bovine serum (FBS) (Cytiva Hyclone, #SH30396.03), 1% penicillin/streptomycin (Gibco, #15140122), 10 ng/mL rhFGF, 1 μ M dexamethasone, and 3 μ g/mL puromycin as appropriate to maintain lines carrying the DUX4 transgene. All cells were incubated at 37°C supplemented with 5% CO₂. Cells were passaged by treatment with 0.25% trypsin- Ethylenediaminetetraacetic acid (EDTA) (Gibco, #25200056) for 5 minutes at 37°C followed by growth media addition and subculturing. Cell treatments to induce

DUX4 include treatment with 1uM doxorubicin or 10uM etoposide for 6 hours, followed by washout and addition of lineage tracing chemicals (see below).

Establishment of 4'OH-tamoxifen-inducible lineage tracing lines

pRRLSIN-cPPT-PGK-GFP-WPRE plasmid (Addgene #12252) was linearized to eliminate GFP-WPRE via digestion with KpnI-HF (NEB, #R3142) and AgeI-HF (NEB, #R3552). Primers containing linearized pRRL backbone overhangs were used to PCR amplify ERT2CreERT2-polyA from pCAG-ERT2CreERT2 (Addgene #13777) using Phusion Plus PCR master mix (Thermo Scientific #F631S) according to manufacturer's instructions. Product size was confirmed via 1XTAE agarose gel electrophoresis and gel purified using NucleoSpin Gel and PCR cleanup mini kit (Machery-Nagel, #740609). The ERT2CreERT2 fragment was inserted into the pRRL backbone via HiFi DNA Assembly Master Mix (NEB, #E2621) according to manufacturer's instructions. Plasmids were transformed into Stbl3 competent cells, extracted via PureLink Quick Plasmid Miniprep Kit (Invitrogen, #K210011), and verified via sanger sequencing (FHCC Genomics core). The intermediate pRRL-PGK-ERT2CreERT2 plasmid was digested and linearized with KpnI-HF. PGK-BlastR was amplified with backbone overhangs from pCW57.1-TetON-iDUX4ca-PGK-BlastR (Tapscott Lab) and inserted via HiFi into the backbone. The pRRL-PGK-ERT2CreERT2-PGK-BlastR sequence was confirmed and expanded as previously described. Completed plasmid was further grown in Stbl3 cells and expanded for extraction via Nucleobond Xtra Maxi EF kit (Macherey-Nagel, #740424.50).

DUX4 reporter lines were created by digesting pRRL-PGK-ERT2CreERT2 with EcoRI-HF (NEB, #R3101) and SalI-HF (NEB, #R3138) to excise PGK-ERT2. ERT2 with backbone and target promoter sequence overhangs was PCR amplified. *H3Y1* and *LEUTX* promoter sequences were amplified with backbone and ERT2 overhangs from G401 genomic DNA. The DUX4 target gene amplicon and ERT2 inserts were cloned into the pRRL-CreERT2-PGK-BlastR linearized backbone via HiFi DNA assembly, confirmed, and expanded as stated previously. Verified plasmids pRRL-PGK-ERT2CreERT2-PGK-BlastR, pRRL-*H3Y1* promoter-ERT2CreERT2-PGK-BlastR, pRRL-*LEUTX* promoter-ERT2CreERT2-PGK-BlastR, or pLV-CMV-loxPDsREDloxP-GFP-IRES-PuroR "Stoplight" (Addgene #65726), along with psPAX2 (Addgene, #12260), and pMD2.G (Addgene, #12259) were transfected into 293T cells with Lipofectamine 3000 to produce lentiviral particles. G401 and myoblast cells were then transduced with lentivirus in the presence of 10ug/mL polybrene. Cells were selected for lentiviral integration with growth media supplemented with 3ug/mL puromycin then 3ug/mL blasticidin S.

Establishment of doxycycline-inducible lineage tracing lines

pRRL-PGK-ERT2CreERT2-PGK-BlastR and pRRL-*H3Y1* promoter-ERT2CreERT2-PGK-BlastR plasmids were linearized by digestion with BamHI-HF (NEB, #R3136) and FseI (NEB, #R0588) to excise the ERT2CreERT2 fragment. pRRL-TetON 3G (rtTA) and bGH poly(A) fragments were amplified with backbone overhangs from pCW57.1-TetON-iDUX4ca-PGK-BlastR and AAVS1 HA(L)-CMV-mEmerald-hPGK-PuroR-AAVS1 HA(R) plasmids, respectively, and inserted into the linearized backbones via HiFi DNA assembly. Puromycin resistance was removed from the pLV-CMV-

loxPDsREDloxP-GFP-IRES-PuroR “Stoplight” (Addgene #65726) construct by digestion and linearization with AgeI-HF (NEB, #R3552) and NsiI-HF (NEB, #R3127). The bGH poly(A) fragment was PCR amplified from AAVS1 HA(L)-CMV-mEmerald-hPGK-PuroR-AAVS1 HA(R) with backbone overhangs and was inserted into the linearized backbone via HiFi DNA assembly. For the AAVS1 knock-in Tet-responsive Cre construct, AAVS1 HA(L)-CMV-mEmerald-hPGK-PuroR-AAVS1 HA(R) was linearized by digestion with HindIII-HF (NEB, #R3104) and HpaI (NEB, #R0105). The TRE3GS promoter and Cre Recombinase sequences were amplified from pCW57.1-TetON-iDUX4ca-PGK-BlastR and pCAG-ERT2CreERT2 (Addgene #13777), respectively, with backbone overhangs. These segments were inserted into the linearized backbone via HiFi DNA assembly. All plasmids were confirmed and expanded as previously described. Verified plasmids pRRL-PGK-rtTA-BlastR, pRRL-*H3Y1 promoter*-rtTA-PGK-BlastR, or pLV-CMV-loxPDsREDloxP-GFP “Stoplight”, along with psPAX2 (Addgene, #12260), and pMD2.G (Addgene, #12259) were transfected into 293T cells with Lipofectamine 3000 to produce lentiviral particles. G401 and myoblast cells were then transduced with lentivirus in the presence of 10ug/mL polybrene and selected with growth media supplemented 3ug/mL blasticidin S. For TetOn-Cre knock-in at the AAVS1 locus, HA(L)-AAVS1-TetOn-Cre-PGK-PuroR-HA(R) and pX459-sgAAVS1 (Addgene #62988), which contains a sgRNA for the AAVS1 site and Cas9, were transfected into HeLa cells with the Lipofectamine 3000 kit (Invitrogen, #L3000015) according to manufacturer’s protocol. Lines were selected for proper AAVS1 knock-in using puromycin selection (Sigma-Aldrich, #P8833) at 1.5μg/mL. Cells were finally selected for “stoplight” lentiviral integration by FACS for DsRED+/GFP- cells.

Establishment of endogenous DUX4 target doxycycline-inducible lineage tracing lines

Cell lines contain the same AAVS1-TetOn-Cre-PGK-PuroR-HA(R) and “Stoplight” constructs as described in the previous section. To make plasmid donor templates for knock-in at the endogenous *H3Y1* and *ZSCAN4* promoters, the *ZSCAN4* construct was created first. AAVS1 HA(L)-CMV-mEmerald-hPGK-PuroR-AAVS1 HA(R) was digested with EcoRI-HF (NEB, #R3101), which has two sites outside both homology arms. Five different fragments were amplified via PCR with overhang sequences for HiFi assembly: 841bp upstream of the *ZSCAN4* CDS was amplified from genomic DNA (*ZSCAN4* HA-L), TetON 3G (rtTA)- poly(A) and PGK-BlastR were amplified from pRRL-PGK-rtTA-PGK-BlastR with long primers for PGK-BlastR that incorporate FRT sites, bGH poly(A) was amplified from AAVS1 HA(L)-CMV-mEmerald-hPGK-PuroR-AAVS1 HA(R), and 861 bp downstream from the *ZSCAN4* CDS was amplified from genomic DNA (*ZSCAN4* HA-R). These five fragments were inserted into the backbone via HiFi DNA assembly to form *ZSCAN4* promoter HA-L-rtTA-FRT-PGK-BlastR-Frt-*ZSCAN4* HA-R. The *H3Y1* promoter endogenous KI was created when AAVS1 HA(L)-CMV-mEmerald-hPGK-PuroR-AAVS1 HA(R) was digested with EcoRI-HF (NEB, #R3101), which has two sites outside both homology arms. Three PCRs were amplified with overlapping overhangs for insertion: 816 bp upstream of the *H3Y1* CDS was amplified from genomic DNA (*H3Y1* HA-L), rtTA-FRT-PGK-BlastR-FRT was amplified from *ZSCAN4* promoter HA-L-rtTA-FRT-PGK-BlastR-Frt-*ZSCAN4* HA-R, and 803 bp downstream of the *H3Y1* CDS was amplified from genomic DNA (*H3Y1* HA-L). These three fragments were inserted into the backbone via HiFi DNA assembly to form

H3Y1 promoter HA-L-rtTA-FRT-PGK-BlastR-Frt-H3Y1 HA-R. All plasmids were confirmed and expanded as previously described. In G401 and myoblast lines with “Stoplight” and AAVS1-TetOn-Cre-PGK-PuroR-HA(R) integration, ZSCAN4 promoter HA-L-rtTA-FRT-PGK-BlastR-Frt-ZSCAN4 HA-R or HA-L-rtTA-FRT-PGK-BlastR-Frt-H3Y1 HA-R were transfected with TrueCut Cas9 protein v2, TrueGuide sgRNAs targeting each specific locus, and Lipofectamine CRISPRMAX Transfection Reagent (all Invitrogen, TrueTag Donor DNA kit #A42992) according to manufacturer’s protocols. Cells were selected with growth media supplemented with 3ug/mL blasticidin S. Finally, cells were transfected with pCS-FLPe (Addgene #31130) using the Lipofectamine 3000 kit (Invitrogen, #L3000015) according to manufacturer’s protocol. To excise the BlastR cassette at the endogenous loci.

Flow cytometry and FACS

Cells were trypsinized and resuspended in FACS Buffer (1X PBS with 2%FBS and 5mM EDTA) and kept on ice until use. For flow cytometry analysis, live cell samples were run on one of several analyzers (BD Biosciences: FACSCanto II, LSR Fortessa x50, FACSymphony A3, or FACSymphony A5). For FACS, cells were sorted into tubes or plates with growth media using a BD Biosciences FACSymphony S6 or FACS Aria II instrument. EGFP fluorescence was quantified using Blue 488nm (200mW) laser readouts and DsRED fluorescence was quantified using Green 552nm (200mW) laser readouts. This analysis was supported by the Flow Cytometry shared resource of the Fred Hutch/University of Washington Cancer Consortium (P30 CA015704). FlowJo v10.8 software (BD Biosciences) was used for all flow cytometry analysis. For all GFP/DsRED fluorescence intensity dot plots, cells were first gated for live cells to eliminate debris using FSC and SSC, then gated on single cells to eliminate doublets using FSC-H and FSC-W. Single color and no color (WT) cell lines were used as compensation controls. Compensation matrices were kept consistent within experiments.

Quantitative reverse transcription PCR (RT-qPCR)

Total cellular RNA was isolated using the Nucleospin RNA kit (Machery-Nagel, #740955.250) according to the manufacturer protocol. 0.5-1µg RNA was treated with Amplification grade-DNaseI (Thermo Fisher, #18068015), heat inactivated with EDTA (Thermo Fisher, #AM9260G), and reverse transcribed into cDNA using random hexamers or oligo-dT with the Superscript IV First-Strand Synthesis System (Invitrogen, #18091050) following manufacturer instructions. Quantitative PCR was run on a QuantStudio 7 Flex (Applied Biosystem) using iTaq SYBR Green Supermix (Bio-Rad, #1725124).

ASO-mediated KD of gene expression

Cells were seeded at a density of 1×10^5 cells per well of a 6-well plate 24 hours before transfection. 50pmol 2'MOE ASOs were diluted in OptiMEM and transfected into cells using Lipofectamine RNAiMAX (Invitrogen #13778150) following manufacturer’s instructions. 24 hours after transfection, cells were harvested for flow cytometry of GFP and DsRED fluorescence and RNA isolation for RT-qPCR analysis.

Figures

Figure 5.1: Pilot testing of the 4'OH-Tamoxifen-inducible Cre system shows some promise in temporal control of lineage tracing.

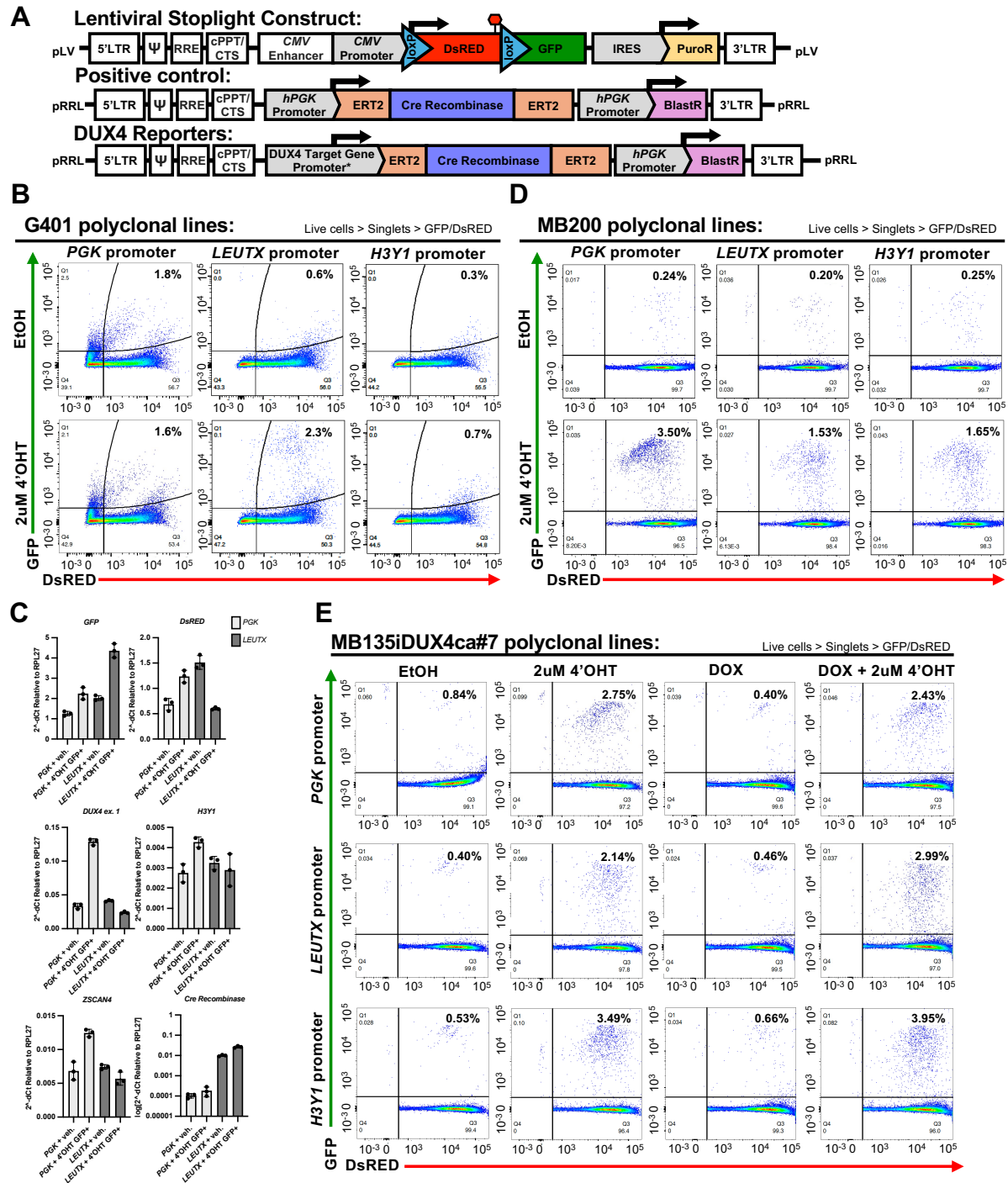


Figure 5.1: (A) Schematic of constructs used in 4'OH-Tamoxifen (4'OH-T)-based lineage tracing system. 4'OH-T treatment of cells with constitutive hPGK-ERT2CreERT2 expression should induce excision of DsRED and expression of *GFP*. Cells with DUX4

target gene promoters should only express ERT2CreERT2 in the presence of DUX4. 4'OH-T addition to *DUX4*-expressing cells with *DUX4 target*-ERT2CreERT2 constructs should induce excision of DsRED and expression of *GFP*. LTR= long terminal repeat, ψ = HIV-1 psi packaging element, RRE = HIV-1 rev response element, cPPT/CTS = central polypurine tract and central termination sequence of HIV-1, loxP = Cre recombinase recognition site, IRES = internal ribosome entry site, PuroR = puromycin resistance, ERT2 = estrogen receptor T2, BlastR = blasticidin resistance. **(B, D)** Flow cytometry dot plots of various G401 **(B)** or MB200 **(D)** polyclonal control or DUX4 lineage tracing lines treated for three days, replenishing every 24 hours, with EtOH vehicle or 2uM 4'OH-T. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated. **(C)** RT-qPCR analysis of expression of various RNAs compared to housekeeping gene *RPL27* expression in G401 *PGK*-ERT2CreERT2 or *LEUTX promoter*-ERT2CreERT2 polyclonal lines treated with EtOH vehicle or 4'OH-T for three days. Data represents mean $\pm$ SD of technical replicates (n=3) of one biological replicate (n=1). **(E)** Flow cytometry dot plots of MB135iDUX4ca#7 polyclonal control or DUX4 lineage tracing lines treated with vehicle or 1ug/mL doxycycline to induce *DUX4* expression for 4 hours before washout and addition of 2uM 4'OH-T or EtOH vehicle for three days, replenishing every 24 hours. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated.

Figure 5.2: Increasing expression of *DUX4* with doxorubicin or etoposide treatment does not improve temporal efficiency of ERT2CreERT2-based lineage tracing system.

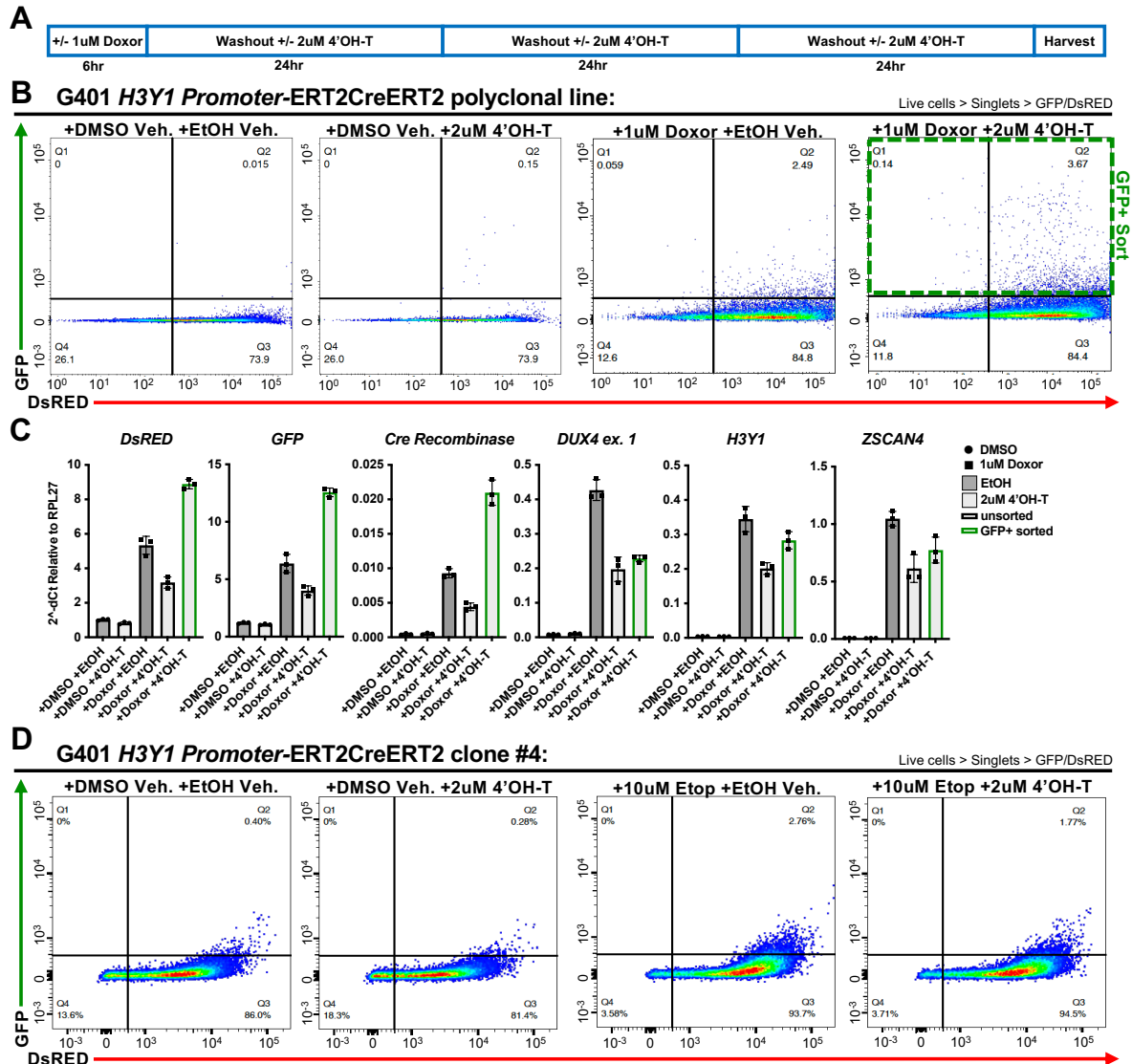


Figure 5.2: (A) Time course of treatment in G401 lineage tracing monoclonal lines. Cells are treated with 1uM doxorubicin to induce *DUX4* expression or equivalent volumes of DMSO vehicle for 6 hours before washout and addition of 2uM 4'OH-T or EtOH vehicle. Cells are treated with 4'OH-T or EtOH for three days, with washout and replenishment every 24 hours, before harvest for flow cytometry and RNA analysis. **(B)** Flow cytometry dot plots of G401 *H3Y1* promoter- ERT2CreERT2 polyclonal lines treated according to the time course in **(A)**. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated. Green box indicates cell population sorted via fluorescence activated cell sorting (FACS) for separate RNA analysis. **(C)** RT-qPCR analysis of expression of various RNAs compared to housekeeping gene *RPL27* expression in G401 *H3Y1* promoter-

ERT2CreERT2 polyclonal cells treated according to **(A)** and/or FACS sorted according to **(B)**. Data represents mean $\pm$ SD of technical replicates (n=3) of one biological replicate (n=1). **(D)** Flow cytometry dot plots of G401 *H3Y1 promoter*- ERT2CreERT2 monoclonal #4 treated according to the time course in **(A)**. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated.

Figure 5.3: Doxycycline-inducible Cre recombinase system increases efficacy of Stoplight construct Cre excision.

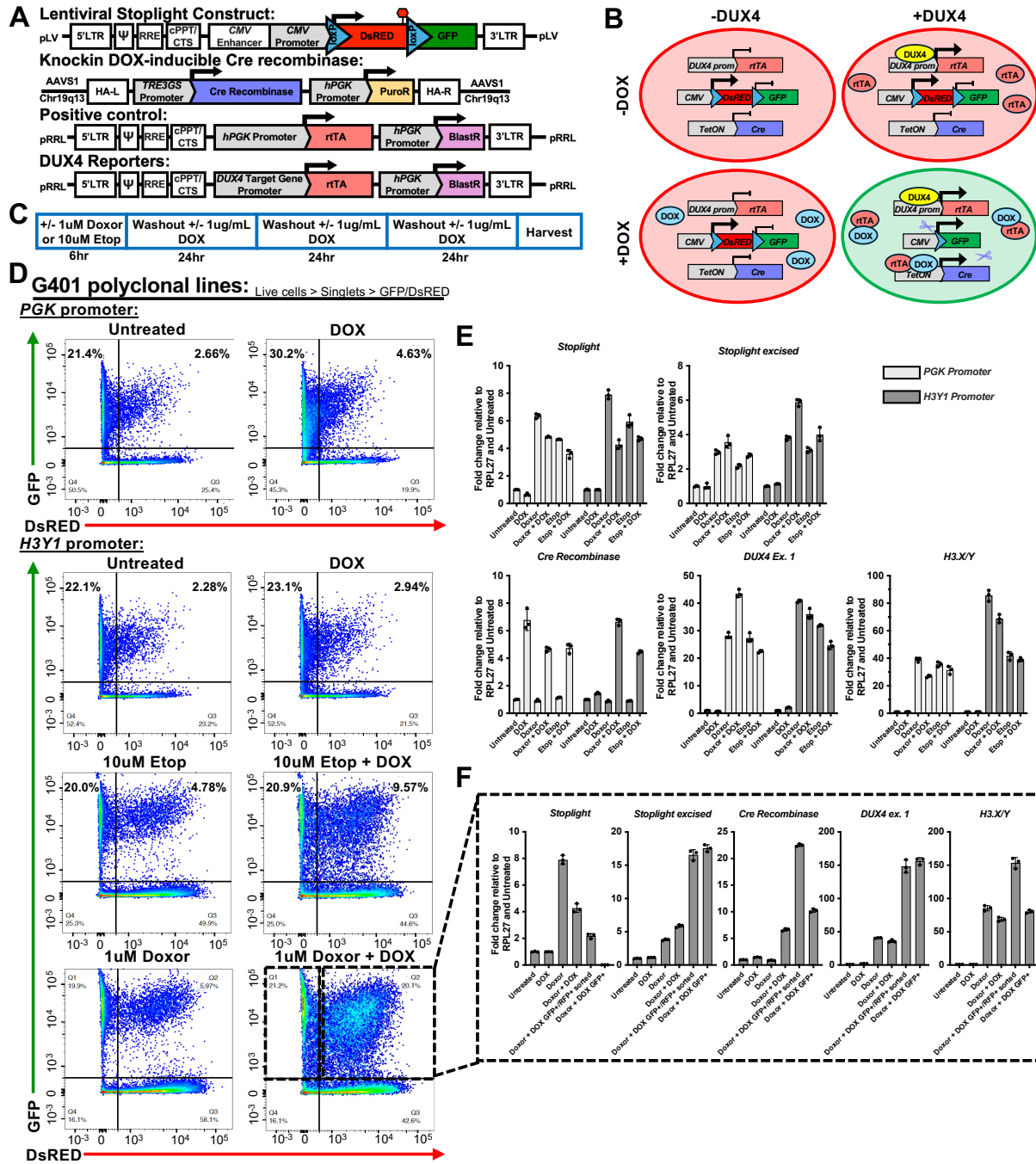


Figure 5.3: (A) Schematic of constructs used in the doxycycline (DOX)-inducible Cre lineage tracing system. LTR= long terminal repeat, ψ = HIV-1 psi packaging element, RRE = HIV-1 rev response element, cPPT/CTS = central polypurine tract and central termination sequence of HIV-1, loxP = Cre recombinase recognition site, HA-L/R = homology arm (left or right), TRE3GS = Tet 3G On promoter, PuroR = puromycin resistance, rTA = TetOn 3G, BlastR = blasticidin resistance. **(B)** Schematic of DOX-inducible lineage tracing system. DOX treatment of cells with constitutive hPGK-rTA

expression should induce excision of DsRED and expression of *GFP*. Cells with DUX4 target gene promoters should only express rtTA in the presence of DUX4. DOX addition to *DUX4*-expressing cells with *DUX4 target*-rtTA constructs should induce TetOn-Cre recombinase expression, leading to excision of DsRED and expression of *GFP*. **(C)** Time course of treatment in G401 lineage tracing monoclonal lines. Cells are treated with 1uM doxorubicin or 10uM etoposide to induce *DUX4* expression or equivalent volumes of DMSO vehicle for 6 hours before washout and addition of 1ug/mL DOX or vehicle. Cells are treated with DOX or vehicle for three days, with washout and replenishment every 24 hours, before harvest for flow cytometry and RNA analysis. **(D)** Top: Flow cytometry dot plots of G401 polyclonal *PGK*-rtTA lines with and without doxycycline. Bottom: Flow cytometry dot plots of G401 *H3Y1 promoter*-rtTA polyclonal lines treated according to the time course in **(C)**. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated. Black boxes indicate populations sorted via FACS for separate RNA analysis. **(E)** RT-qPCR analysis of expression of various RNAs compared to housekeeping gene *RPL27* expression in G401 *PGK* and *H3Y1 promoter*-rtTA polyclonal cells treated according to **(C)** or **(D)**. Data represents mean $\pm$ SD of technical replicates (n=3) of one biological replicate (n=1). **(F)** RT-qPCR analysis of expression of various RNAs compared to housekeeping gene *RPL27* expression in G401 *H3Y1 promoter*-rtTA polyclonal cells treated according to **(C)** with or without sorting via FACS as shown in **(D)**. Data represents mean $\pm$ SD of technical replicates (n=3) of one biological replicate (n=1).

Figure 5.4: G401 PGK-rtTA monoclonal lines are not responsive to DOX induction.

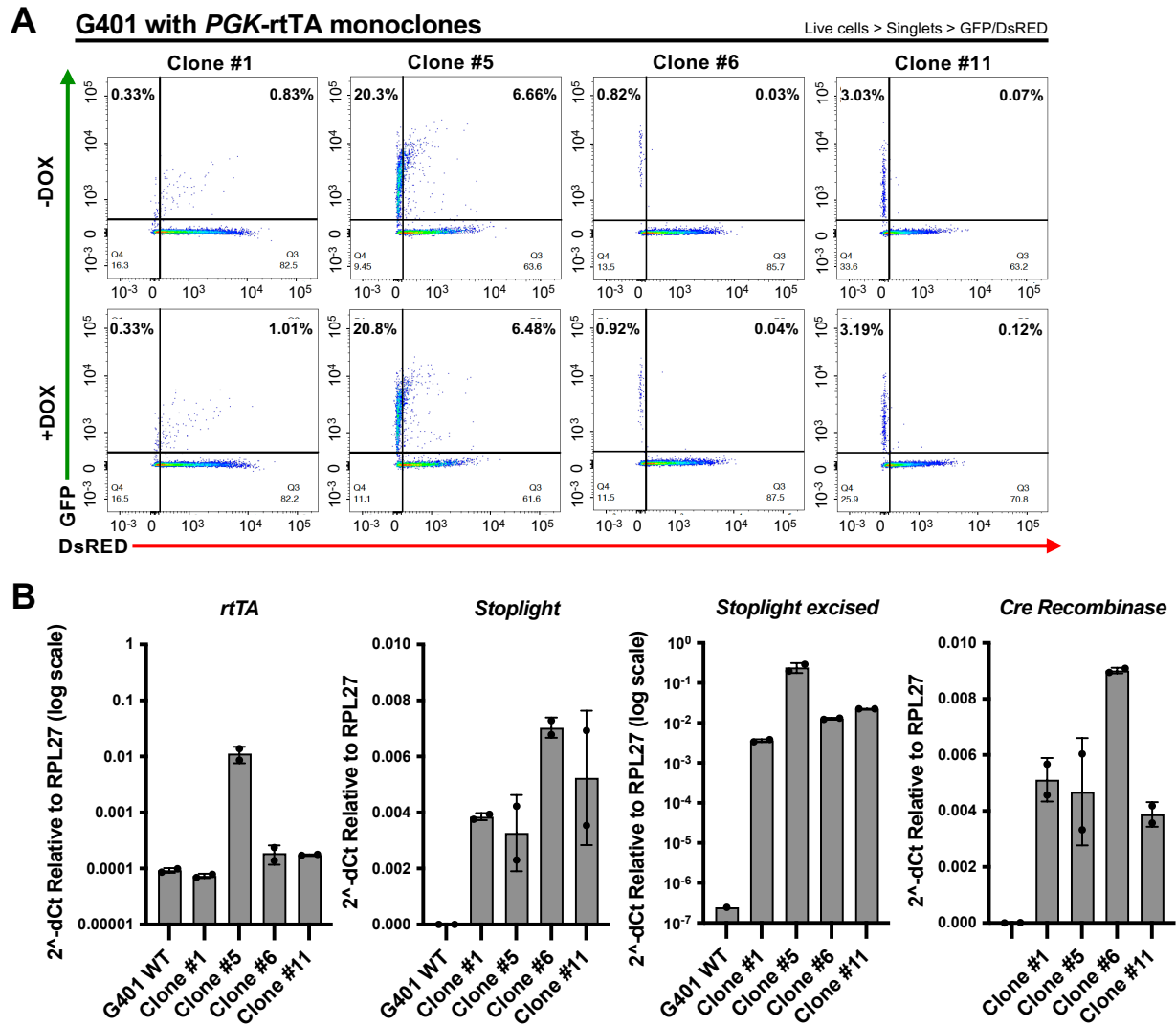


Figure 5.4 (A) Flow cytometry dot plots of G401 PGK-rtTA monoclonal lines with and without three days of doxycycline treatment. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated. **(B)** RT-qPCR analysis of expression of various RNAs compared to housekeeping gene *RPL27* expression in G401 PGK-rtTA monoclonal lines treated according to **(A)**. Data represents mean $\pm$ SD of technical replicates ($n=3$) of one biological replicate ($n=1$).

Figure 5.5: Enhancing DUX4 expression with doxorubicin shows robust response in G401 lineage tracing monoclonal cells treated with doxycycline, but monoclonal expression of GFP is still leaky.

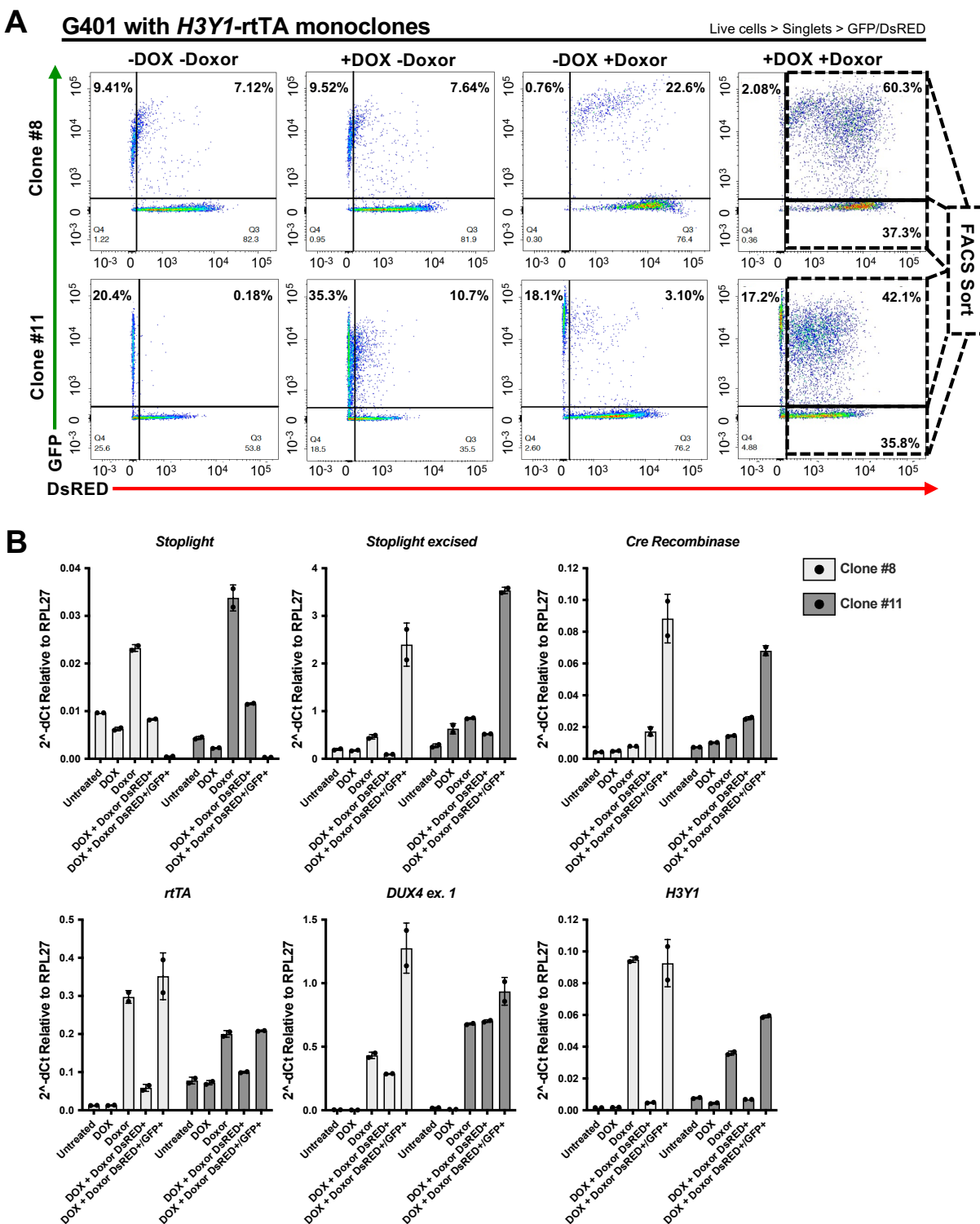


Figure 5.5: (A) Flow cytometry dot plots of G401 *H3Y1 promoter*-rtTA monoclonal lines treated according to the time course in **Fig. 5.2C**. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated. Black boxes indicate populations sorted via FACS for separate RNA analysis. **(B)** RT-qPCR analysis of expression of various RNAs compared to housekeeping gene *RPL27* expression in G401 *H3Y1 promoter*-rtTA monoclonal lines in **(A)**. Data represents mean $\pm$ SD of technical replicates (n=3) of one biological replicate (n=1).

Figure 5.6: Enhancing DUX4 expression with “Catch and Release” shows robust response in G401 lineage tracing monoclonal cells treated with doxycycline, but monoclonal expression of GFP is still leaky.

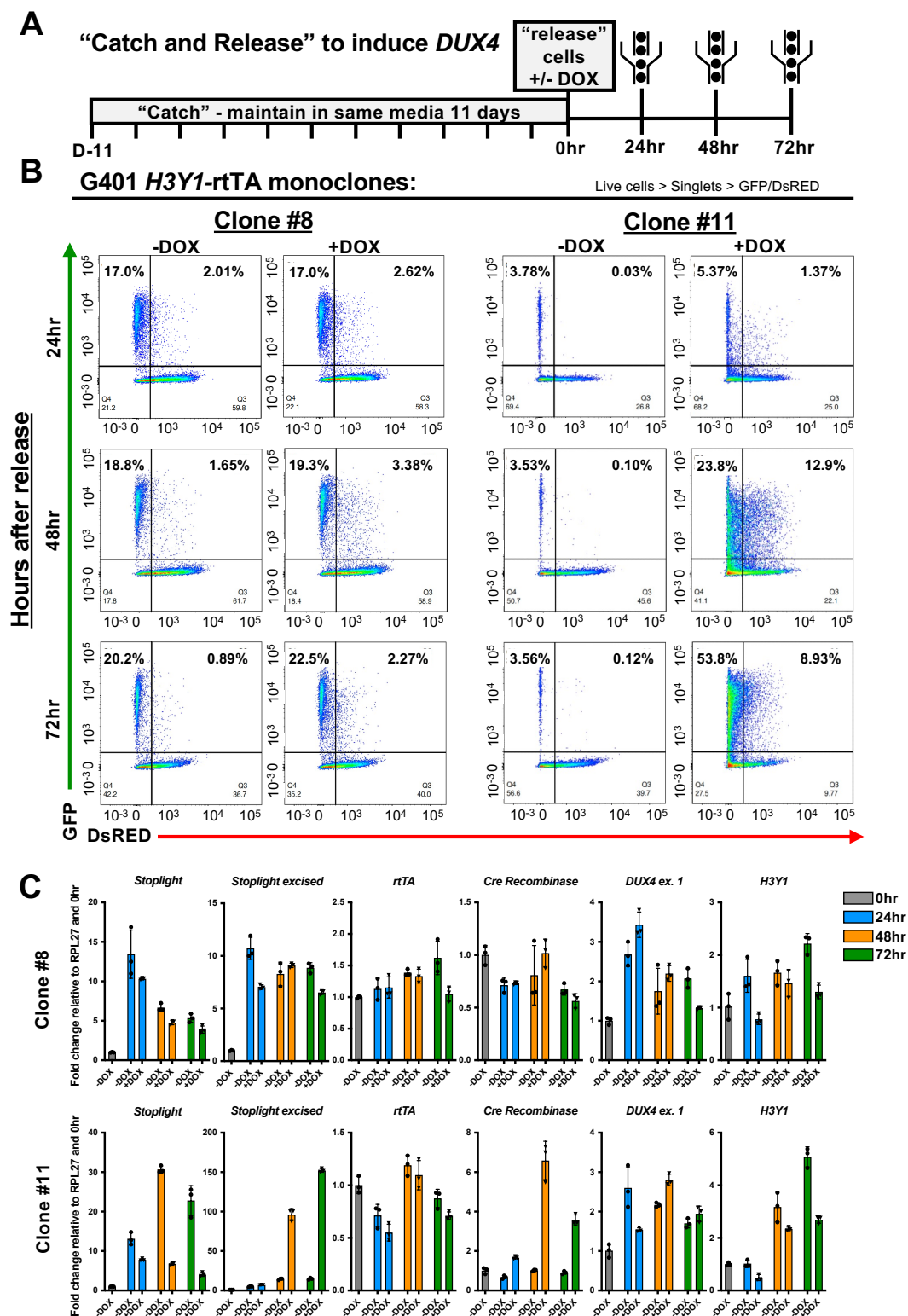


Figure 5.6: (A) Schematic of “catch and release” protocol for induction of *DUX4* in G401 lineage tracing monoclonal lines. Cells are kept and starved in the same growth media for 11 days before trypsinization and “release” into fresh media. Upon release, cells are treated with DOX or vehicle for three days, with harvest for RNA and flow analysis and DOX replenishment every 24 hours. **(B)** Flow cytometry dot plots of G401 *H3Y1 promoter-rtTA* monoclonal lines treated according to the time course in **(A)**. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated. **(B)** RT-qPCR analysis of expression of various RNAs compared to housekeeping gene *RPL27* expression in G401 *H3Y1 promoter-rtTA* monoclonal lines in **(B)**. Data represents mean $\pm$ SD of technical replicates (n=3) of one biological replicate (n=1).

Figure 5.7: Stoplight excision and GFP expression in G401 monoclonal lines is partially DUX4-dependent, though leaky GFP expression is still prevalent.

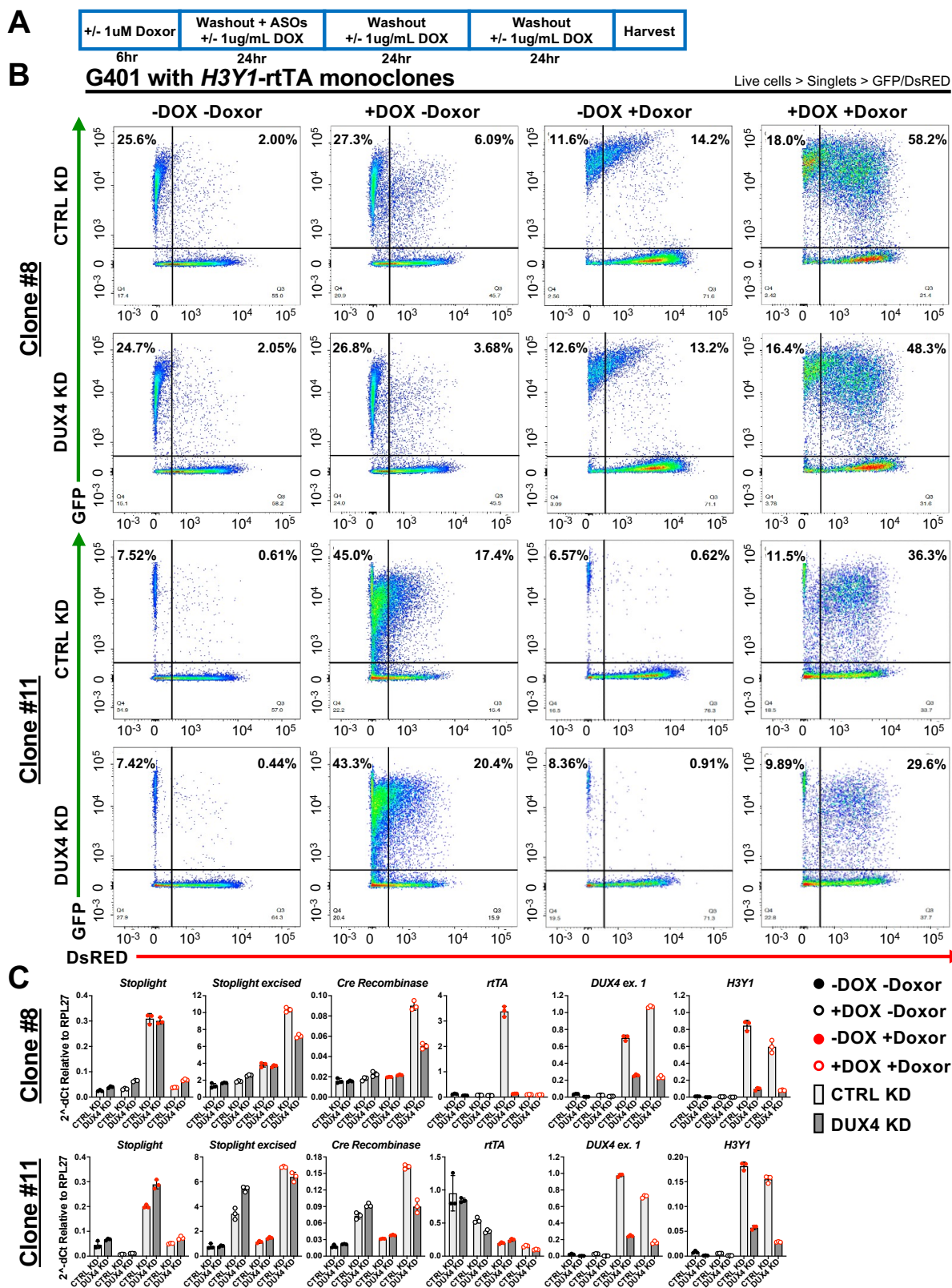


Figure 5.6: **(A)** Schematic of antisense oligonucleotide (ASO)-mediated knockdown (KD) of DUX4 and doxorubicin (Doxor)-mediated induction of *DUX4* in G401 lineage tracing monoclonal lines. Cells are treated with 1 μ M doxor for 6 hours before washout and transfection with CTRL or DUX4 ASOs and with DOX or vehicle treatment for three days, washout DOX replenishment every 24 hours, before harvest for flow cytometry and RNA analysis. **(B)** Flow cytometry dot plots of G401 *H3Y1 promoter*-rtTA monoclonal lines treated according to the time course in **(A)**. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated. **(B)** RT-qPCR analysis of expression of various RNAs compared to housekeeping gene *RPL27* expression in G401 *H3Y1 promoter*-rtTA monoclonal lines in **(B)**. Data represents mean $\pm$ SD of technical replicates (n=3) of one biological replicate (n=1).

Figure 5.8: Myoblast polyclonal lineage tracing lines with *rtTA* knocked into *DUX4* target loci are partially responsive to etoposide treatment.

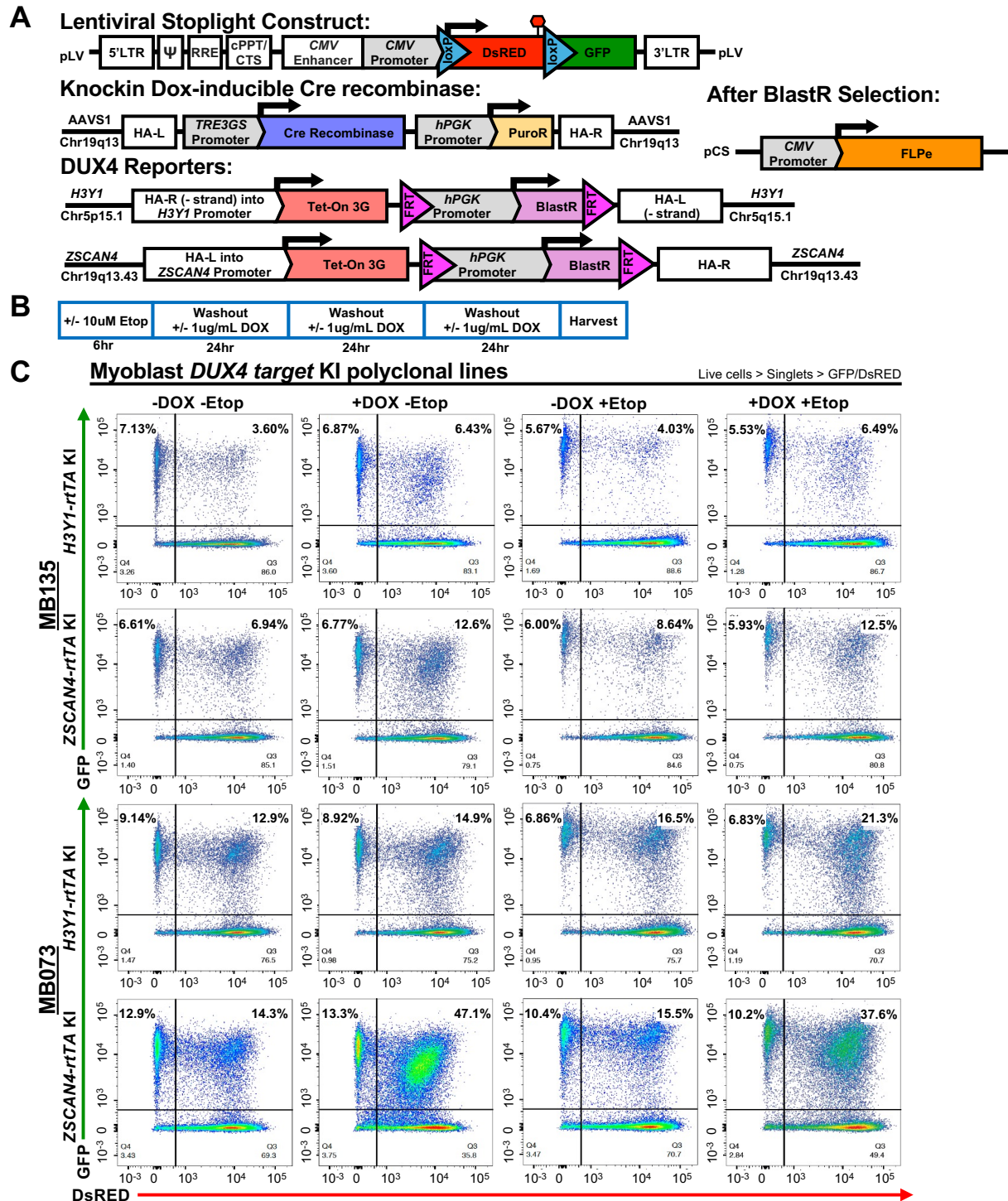


Figure 5.8: (A) Schematic of constructs used in the doxycycline (DOX)-inducible Cre lineage tracing system with *rtTA* knock-ins at endogenous loci. Cells with *rtTA* KI at *DUX4* target gene promoters should only express *rtTA* in the presence of *DUX4*. DOX addition to *DUX4*-expressing cells with *DUX4* target-*rtTA* knock-ins should induce

TetON-Cre recombinase expression, leading to excision of DsRED and expression of *GFP*. To select for rtTA knock-in at endogenous loci, cells are treated with blasticidin. After blasticidin selection, cells can be transfected with pCS-FLPe to excise the blasticidin resistance cassette and prevent spurious expression from the presence of a constitutive promoter at DUX4 target loci. LTR= long terminal repeat, ψ = HIV-1 psi packaging element, RRE = HIV-1 rev response element, cPPT/CTS = central polypurine tract and central termination sequence of HIV-1, loxP = Cre recombinase recognition site, HA-L/R = homology arm (left or right), TRE3GS = Tet 3G On promoter, PuroR = puromycin resistance, rtTA = TetOn 3G, FRT = FLP recombinase target sequence, BlastR = blasticidin resistance, FLPe = enhanced flippase recombinase. **(B)** Time course of treatment in myoblast lineage tracing polyclonal lines. Cells are treated with 10uM etoposide to induce *DUX4* expression or equivalent volumes of DMSO vehicle for 6 hours before washout and addition of 1ug/mL DOX or vehicle. Cells are treated with DOX or vehicle for three days, with washout and replenishment every 24 hours, before harvest for flow cytometry and RNA analysis. **(C)** Flow cytometry dot plots of MB135 (top) or MB073 (bottom) myoblast lines with *H3Y1 promoter-rtTA* or *ZSCAN4 promoter-rtTA* polyclonal knock-in lines treated according to the time course in **(B)**. DsRED fluorescence is on the x axis and GFP fluorescence is on the y axis. Percentages of cells in each quadrant are indicated.

Chapter 6: Discussion

Epigenetic regulation of the D4Z4 and implications for FSHD therapeutic screens

Due to its repetitive nature, the epigenetic landscape of the D4Z4 is presumed to be extremely heterochromatic and silenced during early development. Regulators of the D4Z4 in early development, such as SMCHD1, are known to silence highly repetitive sequences, and are involved in implementation of higher order chromatin structural elements. **In Chapter 2** we have designed a reporter system that can assess the ability of a sequence to recruit factors that drive silencing activity. This system revealed silencing activity in the *LRIF1* promoter and surprisingly, only in one discrete segment of the D4Z4, D4Z4-S5. Though many studies characterizing the D4Z4 and other repetitive genomic loci have revealed large contiguous areas of DNA methylation and repressive histone modifications [37,38,56], they have not focused on detection of specific sequences that may seed and initiate silencing in these regions. Several studies have identified D4Z4 segments that show hypomethylation in FSHD cells when compared to controls. One of these regions, 5P [42–44,121], overlaps with D4Z4-S5.4, the segment in our reporter system that showed the highest silencing activity. Therefore, this region may be an essential regulatory sequence of the D4Z4 that modulates *DUX4* expression, where insufficient activity of epigenetic regulators leads to hypomethylation at this sequence. Further screening using this reporter may provide insight into additional regulators of the D4Z4.

Several KAP1 core repressor proteins were identified in our siRNA panel of potential D4Z4-S5 regulators, such as SETDB1 and its enhancer of nuclear stability cofactor ATF7IP [122]. Although KD of the main scaffold protein in the KAP1 complex TRIM28 did not rescue GFP expression, it has been shown in several studies to induce *DUX4* expression [46,86,88,89,126], indicating its role as a D4Z4 silencer although this activity may not be specific to D4Z4-S5. The silencing activity of SETDB1 is not restricted to its activity through TRIM28, however. Targeting of HP1 isoforms to sequences with incomplete suppressive methylation is sufficient to recruit SETDB1, which deposits H3K9me3 and confers silencing activity in both mouse and human cell systems [230,231]. SETDB1 is also recruited by the Human Silencing Hub (HUSH) complex to H3K9me3-rich sequences to confer both maintenance and spreading of H3K9me3 and heterochromatin [127,128]. SETDB1 interacts directly with DNMT3A to silence promoters [232], increases the activity of Polycomb Repressive Complex 2 (PRC2)-mediated deposition of H3K27me3 [233], and binds directly to H3K14Ac- and H3K9me1/2/3-modified histones to initiate H3K9me3 deposition [234].

The variety of mechanisms by which SETDB1 confers epigenetic silencing is showcased by its diverse roles in the cell. SETDB1 is involved in regulation of cell division and proliferation, maintenance of pluripotency gene expression in early development, X chromosome inactivation, and regulation of retroelement expression (reviewed thoroughly in [235]). Interestingly, SETDB1 and FSHD2 disease modifier SMCHD1 were both identified as *Modifier of murine metastable epiallele (Momme)* genes, shown to silence variegated gene expression from an artificial repeat array in mice [57,236]. SMCHD1 has also been implicated in X chromosome inactivation, early embryonic gene suppression, and silencing of specific autosomal genes [58–60,112]. Despite no known cases of FSHD2 being driven by mutations in SETDB1, the similarity in functions of SETDB1 and SMCHD1 may indicate some overlap in activity and regulation of DUX4 expression. The strong relief in GFP suppression upon KD of SETDB1 or ATF7IP in our system indicates that increasing SETDB1 levels may be another approach to abrogating *DUX4* expression in FSHD, potentially compensating for *SMCHD1* haploinsufficiency. Our ChIP data showed a significant loss of H3K9me3 occupancy at the AAVS1 D4Z4-S5 and LRIF1-FLr loci in cells with SETDB1 or ATF7IP KD, though the endogenous D4Z4-S5.4 and LRIF1-FL loci maintained high levels of H3K9me3 deposition. This may indicate that SETDB1/ATF7IP exhibit silencing initiation at these loci, but are not involved in H3K9me3 maintenance, or that there are compensatory maintenance mechanisms from other D4Z4 regulators that are maintaining D4Z4 heterochromatin at the D4Z4 in these HeLa cells. In parallel with previous work [56], loss of SMCHD1 or LRIF1 did not alter H3K9me3 levels at the D4Z4 in control cells, indicating multiple silencing maintenance mechanisms are functioning in somatic cells. Further characterization of SETDB1/ATF7IP in FSHD is necessary to determine its efficacy as a DUX4 suppressor and its role in D4Z4 silencing initiation and maintenance.

Compared to the striking increase in GFP expression upon SETDB1, ATF7IP, or LRIF1 KD in our reporter system, it was surprising to find that GFP decreased upon loss of SMCHD1. One potential reason for this decrease is that SMCHD1 is a known suppressor of *LRIF1* expression [56], and if LRIF1 is the primary driver of silencing through D4Z4-S5 and LRIF1-FLr, loss SMCHD1 would lead to increased levels of the silencer LRIF1. In our system, however, *LRIF1* expression did not increase upon SMCHD1 KD. Another possible reason for the lack of SMCHD1-mediated silencing activity is the dynamic nature of SMCHD1 and LRIF1 silencing activity in different cell types. In a previous study, KD of SMCHD1 or LRIF1 in control myoblast lines only modestly increased *DUX4* expression, as underlying heterochromatic marks deposited in early development were sufficient to maintain D4Z4 silencing. Indeed, loss of SMCHD1 or LRIF1 in control cells did not decrease D4Z4 heterochromatin marks, as is seen in FSHD2 myoblast lines [56]. The heterochromatin state at the D4Z4 should not impact SMCHD1 modulation of expression from our reporter system, as it is in the

AAVS1 “safe harbor” locus. H3K9me3 has been shown to be crucial for recruitment of SMCHD1 to the D4Z4 in somatic cells [48]; this potentially indicates there is insufficient H3K9me3 at the AAVS1 D4Z4-S5 and LRIF1-FLr loci to robustly recruit SMCHD1 for silencing activity. Our ChIP-qPCR data, however, indicates recruitment of SMCHD1 and LRIF1 to both the endogenous and exogenous AAVS1 sequences above levels of a control region and robust H3K9me3 enrichment at the AAVS1 insert region. These results suggest that while H3K9me3 deposition at the AAVS1 insert region may recruit SMCHD1, it is not conveying silencing activity, indicating more surrounding sequence is necessary for recruitment of cofactors that mediate SMCHD1 activity, or that other factors may be more important drivers of silencing at this locus in somatic cells. Repeating the reporter construct experiment in FSHD cell lines may provide insight into the dynamics of SMCHD1 activity at this sequence and its relevance as a therapeutic intervention for FSHD.

Several screens have identified p38 inhibitors as repressors of *DUX4* and *DUX4* target gene expression [64,65]. Furthermore, p38 inhibition in our reporter system further decreases GFP expression, indicating that p38 kinases or their downstream effectors may be modulating *DUX4* expression through interactions with the D4Z4-S5 sequence. It is currently unknown why p38 inhibition decreases *DUX4* expression in FSHD drug screens. The p38 pathway is a known regulator of muscle differentiation; p38 α kinase activity increases throughout myogenic differentiation and is essential for proper myoblast fusion [237–239]. A recent study has found that the effects of p38 inhibition in FSHD myoblast differentiation are temporally regulated; in early myogenesis, the p38 pathway drives *DUX4* and *DUX4* target gene expression, while in late differentiation, the p38 pathway is dispensable for expression of *DUX4* and its target genes [66]. This may explain why the clinical trial for losmapimod failed to meet its primary endpoint of decreasing target expression, as *DUX4* levels are not changed upon p38 inhibition in late-differentiating myoblasts. Of interest, our study revealed that prolonged p38-inhibitor treatment of HeLa reporter D4Z4-S5 cells under live cell imaging resulted in a short-term decrease in GFP confluency that recovered at later timepoints, while in the LRIF1-FLr cells, GFP confluency remained suppressed. These cells were not differentiating, so there may be some growth or confluency-dependent mechanism to p38 inhibitor-mediated transcriptional suppression. If these trends are consistent at the endogenous D4Z4-S5 locus, then this would indicate that long-term treatment of FSHD with p38 inhibitors may be less effective at silencing *DUX4* expression over time.

Ultimately, the discovery of D4Z4-S5-mediated silencing activity in this isolated reporter system provides a robust platform for identification of new D4Z4-regulating factors, pathways, and therapeutics. We used HeLa cells to establish initial proof-of-concept use of this system and believe it will be easily adaptable for other cell types. Flow cytometry of GFP expression is a simple and sensitive system for future CRISPR-

or RNAi-mediated screens of epigenetic regulators that may interact with this important D4Z4-S5/5P regulatory sequence. This study also shows that cells with D4Z4-S5 can be used for temporal studies in multi-well live-cell imaging screens to detect robust changes in GFP+ cell confluency over time in response to treatment. Our hope is that this system may be used as a platform for future discoveries in FSHD research.

Long term effects of DUX4 expression in development and disease

In addition to previous studies of DUX4-induced post-transcriptional regulation and immune modulation discovered by our lab and others, the work outlined here describes a mechanism in which transcriptionally active DUX4 expression leads to an inhibition in translation and reshaping of the translome, allowing *DUX4*-induced RNAs to be translated while other cellular RNAs are suppressed [99]. Combining this translational reprogramming with the induction of a totipotent, 2C-like transcriptional program [105] implicates cancer cell fate may be altered by expression of DUX4.

Although suppression of translation and subsequent cell growth seems counterintuitive to cancer pathogenesis, a growing body of work suggests translational suppression can confer increased adaptive plasticity, or ability to alter cell fate through translational or epigenetic alterations. Cancer plasticity leads to tumor heterogeneity and the ability to adapt to cellular stressors, such as harsh tumor microenvironments (TME) and therapeutic interventions [107,108]. The rapid expansion of tumor cells leads to the development of harsh TMEs with competition for access to nutrients and oxygen [240]. The early embryonic microenvironment is also hypoxic [241], so DUX4-induction of a 2C-like “totipotent” program may confer better fitness in hypoxic TMEs. I hypothesize that DUX4 expression in cancer cells will confer higher adaptive plasticity and fitness in response to extracellular stressors.

Tantamount to determination of the long-term effects of DUX4 expression in a subset of cancer cells is development of a lineage tracing system. Though I was unsuccessful in creating a lineage tracing system that was responsive, specific, and temporally regulated, I believe I have laid the groundwork for future development. Using the classical 4’OH-T (tamoxifen)-inducible system for lineage tracing proved both nonspecific and insufficiently-responsive in the control line, possibly indicating issues from complications with DUX4 and estrogen signaling or negative effects of constitutive *Cre* expression [226–228]. The introduction of a doxycycline-inducible lineage tracing system in **Chapter 5** was an effort to decrease unnecessary *Cre* activity by temporal control at the level of expression instead of activity. Further optimization led to knock-in of the rtTA at the endogenous DUX4-target promoters, eliminating bias in transgene activity due to location of genomic integration. We hope that identification of monoclonal lines with proper integration can be used to robustly track DUX4+ cells and their progeny.

Chapter 7: References

1. Vastenhouw NL, Cao WX, Lipshitz HD. The maternal-to-zygotic transition revisited. *Development*. 2019;146: dev161471. doi:10.1242/dev.161471
2. Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim J-W, Wike CL, et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet*. 2017;49: 925–934. doi:10.1038/ng.3844
3. De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D. A family of double-homeodomain transcription factors regulates zygotic genome activation in placental mammals. *Nat Genet*. 2017;49: 941–945. doi:10.1038/ng.3858
4. Jukam D, Shariati SAM, Skotheim JM. Zygotic genome activation in vertebrates. *Dev Cell*. 2017;42: 316–332. doi:10.1016/j.devcel.2017.07.026
5. Whiddon JL, Langford AT, Wong C-J, Zhong JW, Tapscott SJ. Conservation and innovation in the DUX4-family gene network. *Nat Genet*. 2017;49: 935–940. doi:10.1038/ng.3846
6. Taubenschmid-Stowers J, Rostovskaya M, Santos F, Ljung S, Argelaguet R, Krueger F. 8C-like cells capture the human zygotic genome activation program in vitro. *Cell Stem Cell*. 2022;29: 449–459.
7. Chen Z, Zhang Y. Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. *Nat Genet*. 2019;51: 947–951. doi:10.1038/s41588-019-0418-7
8. Guo M, Zhang Y, Zhou J, Bi Y, Xu J, Xu C, et al. Precise temporal regulation of Dux is important for embryo development. *Cell Res*. 2019;29: 956–959. doi:10.1038/s41422-019-0238-4
9. Snider L, Geng LN, Lemmers RJLF, Kyba M, Ware CB, Nelson AM, et al. Facioscapulohumeral Dystrophy: Incomplete Suppression of a Retrotransposed Gene. *PLOS Genet*. 2010;6: e1001181. doi:10.1371/journal.pgen.1001181
10. Lemmers RJ, van der Vliet PJ, Blatnik A, Balog J, Zidar J, Henderson D, et al. Chromosome 10q-linked FSHD identifies DUX4 as principal disease gene. *J Med Genet*. 2022;59: 180–188. doi:10.1136/jmedgenet-2020-107041
11. Lemmers RJLF, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, Dauwerse JG, et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science*. 2010;329: 1650–1653. doi:10.1126/science.1189044
12. Dixit M, Anseau E, Tassin A, Winokur S, Shi R, Qian H, et al. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional

- activator of PITX1. *Proc Natl Acad Sci U S A*. 2007;104: 18157–18162. doi:10.1073/pnas.0708659104
13. Deidda G, Cacurri S, Grisanti P, Vigneti E, Piazzo N, Felicetti L. Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur J Hum Genet EJHG*. 1995;3: 155–167. doi:10.1159/000472291
 14. Bakker E, Wijmenga C, Vossen RH, Padberg GW, Hewitt J, van der Wielen M, et al. The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. *Muscle Nerve Suppl*. 1995;2: S39-44.
 15. Guraju N, Jump V, Nallamilli B, Liu R, Luo Z, Hegde M. P476: Identification and accurate sizing of D4Z4 repeat units in patients suspected of facioscapulohumeral muscular dystrophy (FSHD) using optical genome mapping*. *Genet Med Open*. 2023;1: 100523. doi:10.1016/j.gimo.2023.100523
 16. Deenen JCW, Arnts H, van der Maarel SM, Padberg GW, Verschuuren JJGM, Bakker E, et al. Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology*. 2014;83: 1056–1059. doi:10.1212/WNL.0000000000000797
 17. Hamel J, Tawil R. Facioscapulohumeral Muscular Dystrophy: Update on Pathogenesis and Future Treatments. *Neurotherapeutics*. 2018;15: 863–871. doi:10.1007/s13311-018-00675-3
 18. Henke C, Spiesshoefer J, Kabitz H-J, Herkenrath S, Randerath W, Brix T, et al. Respiratory muscle weakness in facioscapulohumeral muscular dystrophy. *Muscle Nerve*. 2019;60: 679–686. doi:10.1002/mus.26717
 19. Tupler R, Berardinelli A, Barbierato L, Frants R, Hewitt JE, Lanzi G, et al. Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. *J Med Genet*. 1996;33: 366–370. doi:10.1136/jmg.33.5.366
 20. Wijmenga C, Frants RR, Brouwer OF, Moerer P, Weber JL, Padberg GW. Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet Lond Engl*. 1990;336: 651–653. doi:10.1016/0140-6736(90)92148-b
 21. Wijmenga C, Padberg GW, Moerer P, Wiegant J, Liem L, Brouwer OF, et al. Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. *Genomics*. 1991;9: 570–575. doi:10.1016/0888-7543(91)90348-i
 22. Upadhyaya M, Lunt P, Sarfarazi M, Broadhead W, Farnham J, Harper PS. The mapping of chromosome 4q markers in relation to facioscapulohumeral muscular dystrophy (FSHD). *Am J Hum Genet*. 1992;51: 404–410.

23. Wijmenga C, Hewitt JE, Sandkuijl LA, Clark LN, Wright TJ, Dauwerse HG, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet.* 1992;2: 26–30. doi:10.1038/ng0992-26
24. Deutekom JCTV, Wijmenga C, Tlenhoven EAEV, Gruter A-M, Hewitt JE, Padberg GW, et al. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum Mol Genet.* 1993;2: 2037–2042. doi:10.1093/hmg/2.12.2037
25. Gabriëls J, Beckers MC, Ding H, De Vriese A, Plaisance S, van der Maarel SM, et al. Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene.* 1999;236: 25–32. doi:10.1016/s0378-1119(99)00267-x
26. Lemmers RJLF, de Kievit P, Sandkuijl L, Padberg GW, van Ommen G-JB, Frants RR, et al. Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat Genet.* 2002;32: 235–236. doi:10.1038/ng999
27. Lemmers RJFL, Wohlgemuth M, Frants RR, Padberg GW, Morava E, van der Maarel SM. Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am J Hum Genet.* 2004;75: 1124–1130. doi:10.1086/426035
28. Kowalijow V, Marcowycz A, Ansseau E, Conde CB, Sauvage S, Mattéotti C, et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord NMD.* 2007;17: 611–623. doi:10.1016/j.nmd.2007.04.002
29. Lemmers RJLF, van der Vliet PJ, Klooster R, Sacconi S, Camaño P, Dauwerse JG, et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science.* 2010;329: 1650–1653. doi:10.1126/science.1189044
30. Tawil R, Kissel JT, Heatwole C, Pandya S, Gronseth G, Benatar M. Evidence-based guideline summary: Evaluation, diagnosis, and management of facioscapulohumeral muscular dystrophy. *Neurology.* 2015;85: 357–364. doi:10.1212/WNL.0000000000001783
31. Goselink RJM, Voermans NC, Okkersen K, Brouwer OF, Padberg GW, Nikolic A, et al. Early onset facioscapulohumeral dystrophy – a systematic review using individual patient data. *Neuromuscul Disord.* 2017;27: 1077–1083. doi:10.1016/j.nmd.2017.09.007
32. Chen T-H, Wu Y-Z, Tseng Y-H. Early-Onset Infantile Facioscapulohumeral Muscular Dystrophy: A Timely Review. *Int J Mol Sci.* 2020;21: 7783. doi:10.3390/ijms21207783

33. Tawil R, Maarel SM, Tapscott SJ. Facioscapulohumeral dystrophy: the path to consensus on pathophysiology. *Skelet Muscle*. 2014;4. doi:10.1186/2044-5040-4-12
34. Lemmers RJLF, Tawil R, Petek LM, Balog J, Block GJ, Santen GWE, et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat Genet*. 2012;44: 1370–1374. doi:10.1038/ng.2454
35. Hamanaka K, Šikrová D, Mitsuhashi S, Masuda H, Sekiguchi Y, Sugiyama A, et al. Homozygous nonsense variant in LRIF1 associated with facioscapulohumeral muscular dystrophy. *Neurology*. 2020;94: e2441–e2447. doi:10.1212/WNL.00000000000009617
36. van den Boogaard ML, Lemmers RJLF, Balog J, Wohlgemuth M, Auranen M, Mitsuhashi S, et al. Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. *Am J Hum Genet*. 2016;98: 1020–1029. doi:10.1016/j.ajhg.2016.03.013
37. Huichalaf C, Micheloni S, Ferri G, Caccia R, Gabellini D. DNA Methylation Analysis of the Macrosatellite Repeat Associated with FSHD Muscular Dystrophy at Single Nucleotide Level. *PLOS ONE*. 2014;9: e115278. doi:10.1371/journal.pone.0115278
38. van Overveld PGM, Lemmers RJFL, Sandkuijl LA, Enthoven L, Winokur ST, Bakels F, et al. Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet*. 2003;35: 315–317. doi:10.1038/ng1262
39. Strafella C, Caputo V, Bortolani S, Torchia E, Megalizzi D, Trastulli G, et al. Whole exome sequencing highlights rare variants in CTCF, DNMT1, DNMT3A, EZH2 and SUV39H1 as associated with FSHD. *Front Genet*. 2023;14. doi:10.3389/fgene.2023.1235589
40. Das S, Chadwick BP. Influence of Repressive Histone and DNA Methylation upon D4Z4 Transcription in Non-Myogenic Cells. *PLoS ONE*. 2016;11: e0160022. doi:10.1371/journal.pone.0160022
41. Strafella C, Megalizzi D, Trastulli G, Proietti Piorgo E, Colantoni L, Tasca G, et al. Integrating D4Z4 methylation analysis into clinical practice: improvement of FSHD molecular diagnosis through distinct thresholds for 4qA/4qA and 4qA/4qB patients. *Clin Epigenetics*. 2024;16: 148. doi:10.1186/s13148-024-01747-2
42. Hartweck LM, Anderson LJ, Lemmers RJ, Dandapat A, Toso EA, Dalton JC, et al. A focal domain of extreme demethylation within D4Z4 in FSHD2. *Neurology*. 2013;80: 392–399. doi:10.1212/WNL.0b013e31827f075c

43. Roche S, Dion C, Broucqsault N, Laberthonnière C, Gaillard M-C, Robin JD, et al. Methylation hotspots evidenced by deep sequencing in patients with facioscapulohumeral dystrophy and mosaicism. *Neurol Genet.* 2019;5: e372. doi:10.1212/NXG.0000000000000372
44. Butterfield RJ, Dunn DM, Duval B, Moldt S, Weiss RB. Deciphering D4Z4 CpG methylation gradients in facioscapulohumeral muscular dystrophy using nanopore sequencing. *Genome Res.* 2023;33: 1439–1454. doi:10.1101/gr.277871.123
45. Jones TI, Yan C, Sapp PC, McKenna-Yasek D, Kang PB, Quinn C, et al. Identifying diagnostic DNA methylation profiles for facioscapulohumeral muscular dystrophy in blood and saliva using bisulfite sequencing. *Clin Epigenetics.* 2014;6: 23. doi:10.1186/1868-7083-6-23
46. Campbell AE, Shadle SC, Jagannathan S, Lim J-W, Resnick R, Tawil R, et al. NuRD and CAF-1-mediated silencing of the D4Z4 array is modulated by DUX4-induced MBD3L proteins. *eLife.* 2018;7: e31023. doi:10.7554/eLife.31023
47. Zeng W, Greef JC de, Chen Y-Y, Chien R, Kong X, Gregson HC, et al. Specific Loss of Histone H3 Lysine 9 Trimethylation and HP1 γ /Cohesin Binding at D4Z4 Repeats Is Associated with Facioscapulohumeral Dystrophy (FSHD). *PLOS Genet.* 2009;5: e1000559. doi:10.1371/journal.pgen.1000559
48. Zeng W, Chen Y-Y, Newkirk DA, Wu B, Balog J, Kong X, et al. Genetic and epigenetic characteristics of FSHD-associated 4q and 10q D4Z4 that are distinct from non-4q/10q D4Z4 homologs. *Hum Mutat.* 2014;35: 998–1010. doi:10.1002/humu.22593
49. Gabellini D, Green MR, Tupler R. Inappropriate Gene Activation in FSHD: A Repressor Complex Binds a Chromosomal Repeat Deleted in Dystrophic Muscle. *Cell.* 2002;110: 339–348. doi:10.1016/S0092-8674(02)00826-7
50. Ottaviani A, Rival-Gervier S, Boussouar A, Foerster AM, Rondier D, Sacconi S, et al. The D4Z4 Macrosatellite Repeat Acts as a CTCF and A-Type Lamins-Dependent Insulator in Facio-Scapulo-Humeral Dystrophy. *PLOS Genet.* 2009;5: e1000394. doi:10.1371/journal.pgen.1000394
51. Lim J-W, Snider L, Yao Z, Tawil R, Van Der Maarel SM, Rigo F, et al. DICER/AGO-dependent epigenetic silencing of D4Z4 repeats enhanced by exogenous siRNA suggests mechanisms and therapies for FSHD. *Hum Mol Genet.* 2015;24: 4817–4828. doi:10.1093/hmg/ddv206
52. Daxinger L, Tapscott SJ, Van Der Maarel SM. Genetic and epigenetic contributors to FSHD. *Curr Opin Genet Dev.* 2015;33: 56–61. doi:10.1016/j.gde.2015.08.007
53. Goossens R, Tihaya MS, Heuvel A van den, Tabot-Ndip K, Willemsen IM, Tapscott SJ, et al. A proteomics study identifying interactors of the FSHD2 gene product

- SMCHD1 reveals RUVBL1-dependent DUX4 repression. *Sci Rep.* 2021;11: 23642. doi:10.1038/s41598-021-03030-3
54. Larsen M, Rost S, El Hajj N, Ferbert A, Deschauer M, Walter MC, et al. Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. *Eur J Hum Genet.* 2015;23: 808–816. doi:10.1038/ejhg.2014.191
 55. Goossens R, Boogaard ML van den, Lemmers RJLF, Balog J, Vliet PJ van der, Willemsen IM, et al. Intronic SMCHD1 variants in FSHD: testing the potential for CRISPR-Cas9 genome editing. *J Med Genet.* 2019;56: 828–837. doi:10.1136/jmedgenet-2019-106402
 56. Šikrová D, Testa AM, Willemsen I, van den Heuvel A, Tapscott SJ, Daxinger L, et al. SMCHD1 and LRIF1 converge at the FSHD-associated D4Z4 repeat and LRIF1 promoter yet display different modes of action. *Commun Biol.* 2023;6: 677. doi:10.1038/s42003-023-05053-0
 57. Blewitt ME, Vickaryous NK, Hemley SJ, Ashe A, Bruxner TJ, Preis JL, et al. An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc Natl Acad Sci U S A.* 2005;102: 7629–7634. doi:10.1073/pnas.0409375102
 58. Blewitt ME, Gendrel A-V, Pang Z, Sparrow DB, Whitelaw N, Craig JM, et al. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet.* 2008;40: 663–669. doi:10.1038/ng.142
 59. Nozawa R-S, Nagao K, Igami K-T, Shibata S, Shirai N, Nozaki N, et al. Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat Struct Mol Biol.* 2013;20: 566–573. doi:10.1038/nsmb.2532
 60. Mason AG, Sliker RC, Balog J, Lemmers RJLF, Wong C-J, Yao Z, et al. SMCHD1 regulates a limited set of gene clusters on autosomal chromosomes. *Skelet Muscle.* 2017;7: 12. doi:10.1186/s13395-017-0129-7
 61. Cohen J, DeSimone A, Lek M, Lek A. Therapeutic Approaches in Facioscapulohumeral Muscular Dystrophy. *Trends Mol Med.* 2021;27: 123–137. doi:10.1016/j.molmed.2020.09.008
 62. Rickard AM, Petek LM, Miller DG. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum Mol Genet.* 2015;24: 5901–5914. doi:10.1093/hmg/ddv315
 63. Himeda CL, Jones PL. FSHD Therapeutic Strategies: What Will It Take to Get to Clinic? *J Pers Med.* 2022;12: 865. doi:10.3390/jpm12060865

64. Rojas LA, Valentine E, Accorsi A, Maglio J, Shen N, Robertson A, et al. p38 α Regulates Expression of DUX4 in a Model of Facioscapulohumeral Muscular Dystrophy. *J Pharmacol Exp Ther*. 2020;374: 489–498. doi:10.1124/jpet.119.264689
65. Oliva J, Galasinski S, Richey A, Campbell AE, Meyers MJ, Modi N, et al. Clinically Advanced p38 Inhibitors Suppress DUX4 Expression in Cellular and Animal Models of Facioscapulohumeral Muscular Dystrophy. *J Pharmacol Exp Ther*. 2019;370: 219–230. doi:10.1124/jpet.119.259663
66. Vangipurapu R, Oliva J, Fox A, Sverdrup FM. Temporal variation in p38-mediated regulation of DUX4 in facioscapulohumeral muscular dystrophy. *Sci Rep*. 2024;14: 26437. doi:10.1038/s41598-024-77911-8
67. Campbell AE, Oliva J, Yates MP, Zhong JW, Shadle SC, Snider L, et al. BET bromodomain inhibitors and agonists of the beta-2 adrenergic receptor identified in screens for compounds that inhibit DUX4 expression in FSHD muscle cells. *Skelet Muscle*. 2017;7: 16. doi:10.1186/s13395-017-0134-x
68. Cruz JM, Hupper N, Wilson LS, Concannon JB, Wang Y, Oberhauser B, et al. Protein kinase A activation inhibits DUX4 gene expression in myotubes from patients with facioscapulohumeral muscular dystrophy. *J Biol Chem*. 2018;293: 11837–11849. doi:10.1074/jbc.RA118.002633
69. Ciszewski L, Lu-Nguyen N, Slater A, Brennan A, Williams HEL, Dickson G, et al. G-quadruplex ligands mediate downregulation of DUX4 expression. *Nucleic Acids Res*. 2020;48: 4179–4194. doi:10.1093/nar/gkaa146
70. Block GJ, Narayanan D, Amell AM, Petek LM, Davidson KC, Bird TD, et al. Wnt/ β -catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. *Hum Mol Genet*. 2013;22: 4661–4672. doi:10.1093/hmg/ddt314
71. Vanderplanck C, Anseau E, Charron S, Stricwant N, Tassin A, Laoudj-Chenivresse D, et al. The FSHD atrophic myotube phenotype is caused by DUX4 expression. *PloS One*. 2011;6: e26820. doi:10.1371/journal.pone.0026820
72. Anseau E, Vanderplanck C, Wauters A, Harper SQ, Coppée F, Belayew A. Antisense Oligonucleotides Used to Target the DUX4 mRNA as Therapeutic Approaches in FaciosScapuloHumeral Muscular Dystrophy (FSHD). *Genes*. 2017;8: 93. doi:10.3390/genes8030093
73. Marsollier A-C, Ciszewski L, Mariot V, Popplewell L, Voit T, Dickson G, et al. Antisense targeting of 3' end elements involved in DUX4 mRNA processing is an efficient therapeutic strategy for facioscapulohumeral dystrophy: a new gene-silencing approach. *Hum Mol Genet*. 2016;25: 1468–1478. doi:10.1093/hmg/ddw015

74. Chen JC, King OD, Zhang Y, Clayton NP, Spencer C, Wentworth BM, et al. Morpholino-mediated Knockdown of DUX4 Toward Facioscapulohumeral Muscular Dystrophy Therapeutics. *Mol Ther*. 2016;24: 1405–1411. doi:10.1038/mt.2016.111
75. Lim KRQ, Maruyama R, Echigoya Y, Nguyen Q, Zhang A, Khawaja H, et al. Inhibition of DUX4 expression with antisense LNA gapmers as a therapy for facioscapulohumeral muscular dystrophy. *Proc Natl Acad Sci U S A*. 2020;117: 16509–16515. doi:10.1073/pnas.1909649117
76. Rashnonejad A, Amini-Chermahini G, Taylor NK, Wein N, Harper SQ. Designed U7 snRNAs inhibit *DUX4* expression and improve FSHD-associated outcomes in *DUX4* overexpressing cells and FSHD patient myotubes. *Mol Ther - Nucleic Acids*. 2021;23: 476–486. doi:10.1016/j.omtn.2020.12.004
77. Wallace LM, Saad NY, Pyne NK, Fowler AM, Eidahl JO, Domire JS, et al. Pre-clinical Safety and Off-Target Studies to Support Translation of AAV-Mediated RNAi Therapy for FSHD. *Mol Ther Methods Clin Dev*. 2018;8: 121–130. doi:10.1016/j.omtm.2017.12.005
78. Wallace LM, Liu J, Domire JS, Garwick-Coppens SE, Guckes SM, Mendell JR, et al. RNA Interference Inhibits DUX4-induced Muscle Toxicity In Vivo: Implications for a Targeted FSHD Therapy. *Mol Ther*. 2012;20: 1417–1423. doi:10.1038/mt.2012.68
79. Lim KRQ, Bittel A, Maruyama R, Echigoya Y, Nguyen Q, Huang Y, et al. DUX4 Transcript Knockdown with Antisense 2'-O-Methoxyethyl Gapmers for the Treatment of Facioscapulohumeral Muscular Dystrophy. *Mol Ther*. 2021;29: 848–858. doi:10.1016/j.ymthe.2020.10.010
80. Bouwman LF, den Hamer B, van den Heuvel A, Franken M, Jackson M, Dwyer CA, et al. Systemic delivery of a DUX4-targeting antisense oligonucleotide to treat facioscapulohumeral muscular dystrophy. *Mol Ther - Nucleic Acids*. 2021;26: 813–827. doi:10.1016/j.omtn.2021.09.010
81. Lu-Nguyen N, Malerba A, Antoni Pineda M, Dickson G, Popplewell L. Improving Molecular and Histopathology in Diaphragm Muscle of the Double Transgenic ACTA1-MCM/FLEXDUX4 Mouse Model of FSHD with Systemic Antisense Therapy. *Hum Gene Ther*. 2022;33: 923–935. doi:10.1089/hum.2021.251
82. Lu-Nguyen N, Malerba A, Herath S, Dickson G, Popplewell L. Systemic antisense therapeutics inhibiting DUX4 expression ameliorates FSHD-like pathology in an FSHD mouse model. *Hum Mol Genet*. 2021;30: 1398–1412. doi:10.1093/hmg/ddab136
83. Klingler C, Ashley J, Shi K, Stiefvater A, Kyba M, Sinnreich M, et al. DNA aptamers against the DUX4 protein reveal novel therapeutic implications for FSHD. *FASEB J*. 2020;34: 4573–4590. doi:10.1096/fj.201902696

84. Mariot V, Joubert R, Marsollier A-C, Hourdé C, Voit T, Dumonceaux J. A Deoxyribonucleic Acid Decoy Trapping DUX4 for the Treatment of Facioscapulohumeral Muscular Dystrophy. *Mol Ther Nucleic Acids*. 2020;22: 1191–1199. doi:10.1016/j.omtn.2020.10.028
85. Joubert R, Mariot V, Charpentier M, Concordet JP, Dumonceaux J. Gene Editing Targeting the DUX4 Polyadenylation Signal: A Therapy for FSHD? *J Pers Med*. 2020;11: 7. doi:10.3390/jpm11010007
86. Das S, Chadwick BP. CRISPR mediated targeting of DUX4 distal regulatory element represses DUX4 target genes dysregulated in Facioscapulohumeral muscular dystrophy. *Sci Rep*. 2021;11: 12598. doi:10.1038/s41598-021-92096-0
87. Šikrová D, Cadar VA, Ariyurek Y, Laros JFJ, Balog J, van der Maarel SM. Adenine base editing of the DUX4 polyadenylation signal for targeted genetic therapy in facioscapulohumeral muscular dystrophy. *Mol Ther Nucleic Acids*. 2021;25: 342–354. doi:10.1016/j.omtn.2021.05.020
88. Himeda CL, Jones TI, Jones PL. Targeted epigenetic repression by CRISPR/dSaCas9 suppresses pathogenic DUX4-fl expression in FSHD. *Mol Ther Methods Clin Dev*. 2020;20: 298–311. doi:10.1016/j.omtm.2020.12.001
89. Himeda CL, Jones TI, Jones PL. CRISPR/dCas9-mediated Transcriptional Inhibition Ameliorates the Epigenetic Dysregulation at D4Z4 and Represses DUX4-fl in FSH Muscular Dystrophy. *Mol Ther*. 2016;24: 527–535. doi:10.1038/mt.2015.200
90. Sacconi S, Briand-Suleau A, Gros M, Baudoin C, Lemmers RJLF, Rondeau S, et al. FSHD1 and FSHD2 form a disease continuum. *Neurology*. 2019;92: e2273–e2285. doi:10.1212/WNL.00000000000007456
91. Hareendran S, Balakrishnan B, Sen D, Kumar S, Srivastava A, Jayandharan GR. Adeno-associated virus (AAV) vectors in gene therapy: immune challenges and strategies to circumvent them. *Rev Med Virol*. 2013;23: 399–413. doi:10.1002/rmv.1762
92. Dyle MC, Kolakada D, Cortazar MA, Jagannathan S. How to get away with nonsense: Mechanisms and consequences of escape from nonsense-mediated RNA decay. *Wiley Interdiscip Rev RNA*. 2020;11: e1560. doi:10.1002/wrna.1560
93. Homma S, Beermann ML, Boyce FM, Miller JB. Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. *Ann Clin Transl Neurol*. 2015;2: 151–166. doi:10.1002/acn3.158
94. Campbell AE, Dyle MC, Albanese R, Matheny T, Sudheendran K, Cortázar MA, et al. Compromised nonsense-mediated RNA decay results in truncated RNA-binding protein production upon DUX4 expression. *Cell Rep*. 2023;42: 112642. doi:10.1016/j.celrep.2023.112642

95. Jagannathan S, Ogata Y, Gafken PR, Tapscott SJ, Bradley RK. Quantitative proteomics reveals key roles for post-transcriptional gene regulation in the molecular pathology of facioscapulohumeral muscular dystrophy. *eLife*. 2019;8: e41740. doi:10.7554/eLife.41740
96. Feng Q, Snider L, Jagannathan S, Tawil R, van der Maarel SM, Tapscott SJ, et al. A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *eLife*. 2015;4: e04996. doi:10.7554/eLife.04996
97. Jagannathan S, Shadle SC, Resnick R, Snider L, Tawil RN, van der Maarel SM, et al. Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum Mol Genet*. 2016;25: 4419–4431. doi:10.1093/hmg/ddw271
98. Campbell AE, Dyle MC, Calviello L, Matheny T, Cortazar MA, Forman T, et al. The myopathic transcription factor DUX4 induces the production of truncated RNA-binding proteins in human muscle cells. *Genomics*; 2021 Jun. doi:10.1101/2021.06.28.450189
99. Hamm DC, Paatela EM, Bennett SR, Wong C-J, Campbell AE, Wladyka CL, et al. The transcription factor DUX4 orchestrates translational reprogramming by broadly suppressing translation efficiency and promoting expression of DUX4-induced mRNAs. *PLoS Biol*. 2023;21: e3002317. doi:10.1371/journal.pbio.3002317
100. Chew G-L, Campbell AE, De Neef E, Sutliff NA, Shadle SC, Tapscott SJ, et al. DUX4 Suppresses MHC Class I to Promote Cancer Immune Evasion and Resistance to Checkpoint Blockade. *Dev Cell*. 2019;50: 658-671.e7. doi:10.1016/j.devcel.2019.06.011
101. Yoshimoto T, Tanaka M, Homme M, Yamazaki Y, Takazawa Y, Antonescu CR, et al. CIC-DUX4 Induces Small Round Cell Sarcomas Distinct from Ewing Sarcoma. *Cancer Res*. 2017;77: 2927–2937. doi:10.1158/0008-5472.CAN-16-3351
102. Tian L, Shao Y, Nance S, Dang J, Xu B, Ma X, et al. Long-read sequencing unveils IGH-DUX4 translocation into the silenced IGH allele in B-cell acute lymphoblastic leukemia. *Nat Commun*. 2019;10: 2789. doi:10.1038/s41467-019-10637-8
103. Pineda JMB, Bradley RK. DUX4 is a common driver of immune evasion and immunotherapy failure in metastatic cancers. *eLife*. 2024;12: RP89017. doi:10.7554/eLife.89017
104. Spens AE, Sutliff NA, Bennett C SR, AE T, S.J. Human DUX4 and mouse Dux interact with STAT1 and broadly inhibit interferon-stimulated gene induction. *Elife*. 2023;12. doi:10.7554/eLife.82057

105. Smith AA, Nip Y, Bennett SR, Hamm DC, Lemmers RJLF, van der Vliet PJ, et al. DUX4 expression in cancer induces a metastable early embryonic totipotent program. *Cell Rep.* 2023;42: 113114. doi:10.1016/j.celrep.2023.113114
106. Grow EJ, Weaver BD, Smith CM, Guo J, Stein P, Shadle SC, et al. p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. *Nat Genet.* 2021;53: 1207–1220. doi:10.1038/s41588-021-00893-0
107. Qin S, Jiang J, Lu Y, Nice EC, Huang C, Zhang J, et al. Emerging role of tumor cell plasticity in modifying therapeutic response. *Signal Transduct Target Ther.* 2020;5: 1–36. doi:10.1038/s41392-020-00313-5
108. Lee LJ, Papadopoli D, Jewer M, Del Rincon S, Topisirovic I, Lawrence MG, et al. Cancer Plasticity: The Role of mRNA Translation. *Trends Cancer.* 2021;7: 134–145. doi:10.1016/j.trecan.2020.09.005
109. Schätzl T, Kaiser L, Deigner H-P. Facioscapulohumeral muscular dystrophy: genetics, gene activation and downstream signalling with regard to recent therapeutic approaches: an update. *Orphanet J Rare Dis.* 2021;16: 129. doi:10.1186/s13023-021-01760-1
110. Campbell AE, Belleville AE, Resnick R, Shadle SC, Tapscott SJ. Facioscapulohumeral dystrophy: activating an early embryonic transcriptional program in human skeletal muscle. *Hum Mol Genet.* 2018;27: R153–R162. doi:10.1093/hmg/ddy162
111. Geng LN, Yao Z, Snider L, Fong AP, Cech JN, Young JM, et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev Cell.* 2012;22: 38–51. doi:10.1016/j.devcel.2011.11.013
112. Ruebel ML, Vincent KA, Schall PZ, Wang K, Latham KE. 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.* 2019;317: C655–C664. doi:10.1152/ajpcell.00116.2019
113. Huang Z, Yu J, Cui W, Johnson BK, Kim K, Pfeifer GP. The chromosomal protein SMCHD1 regulates DNA methylation and the 2c-like state of embryonic stem cells by antagonizing TET proteins. *Sci Adv.* 2021;7. doi:10.1126/sciadv.abb9149
114. Himeda CL, Jones TI, Virbasius C-M, Zhu LJ, Green MR, Jones PL. Identification of Epigenetic Regulators of DUX4-fl for Targeted Therapy of Facioscapulohumeral Muscular Dystrophy. *Mol Ther.* 2018;26: 1797–1807. doi:10.1016/j.ymthe.2018.04.019

115. Greco A, Goossens R, van Engelen B, van der Maarel SM. Consequences of epigenetic derepression in facioscapulohumeral muscular dystrophy. *Clin Genet*. 2020;97: 799–814. doi:10.1111/cge.13726
116. Sadelain M, Papapetrou EP, Bushman FD. Safe harbours for the integration of new DNA in the human genome. *Nat Rev Cancer*. 2012;12: 51–58. doi:10.1038/nrc3179
117. Himeda CL, Debarnot C, Homma S, Beermann ML, Miller JB, Jones PL, et al. Myogenic Enhancers Regulate Expression of the Facioscapulohumeral Muscular Dystrophy-Associated DUX4 Gene. *Mol Cell Biol*. 2014 [cited 26 Jan 2025]. doi:10.1128/MCB.00149-14
118. Balog J, Thijssen PE, Shadle S, Straasheijm KR, van der Vliet PJ, Krom YD, et al. Increased DUX4 expression during muscle differentiation correlates with decreased SMCHD1 protein levels at D4Z4. *Epigenetics*. 2015;10: 1133–1142. doi:10.1080/15592294.2015.1113798
119. Doni Jayavelu N, Jajodia A, Mishra A, Hawkins RD. Candidate silencer elements for the human and mouse genomes. *Nat Commun*. 2020;11: 1061. doi:10.1038/s41467-020-14853-5
120. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*. 2010;20: 110–121. doi:10.1101/gr.097857.109
121. Gaillard M-C, Roche S, Dion C, Tasmadjian A, Bouget G, Salort-Campana E, et al. Differential DNA methylation of the D4Z4 repeat in patients with FSHD and asymptomatic carriers. *Neurology*. 2014;83: 733–742. doi:10.1212/WNL.0000000000000708
122. Timms RT, Tchasovnikarova IA, Antrobus R, Dougan G, Lehner PJ. ATF7IP-Mediated Stabilization of the Histone Methyltransferase SETDB1 Is Essential for Heterochromatin Formation by the HUSH Complex. *Cell Rep*. 2016;17: 653–659. doi:10.1016/j.celrep.2016.09.050
123. Wong MM-K, Hachmer S, Gardner E, Runfola V, Arezza E, Megeney LA, et al. SMCHD1 activates the expression of genes required for the expansion of human myoblasts. *Nucleic Acids Res*. 2024;52: 9450–9462. doi:10.1093/nar/gkae600
124. Le Gall L, Sidlauskaitė E, Mariot V, Dumonceaux J. Therapeutic Strategies Targeting DUX4 in FSHD. *J Clin Med*. 2020;9: 2886. doi:10.3390/jcm9092886
125. Tawil R, Wagner KR, Hamel JI, Leung DG, Statland JM, Wang LH, et al. Safety and efficacy of losmapimod in facioscapulohumeral muscular dystrophy (ReDUX4): a randomised, double-blind, placebo-controlled phase 2b trial. *Lancet Neurol*. 2024;23: 477–486. doi:10.1016/S1474-4422(24)00073-5

126. Dion C, Roche S, Laberthonnière C, Broucqsault N, Mariot V, Xue S, et al. SMCHD1 is involved in de novo methylation of the DUX4-encoding D4Z4 macrosatellite. *Nucleic Acids Res.* 2019;47: 2822–2839. doi:10.1093/nar/gkz005
127. Tchasovnikarova IA, Timms RT, Matheson NJ, Wals K, Antrobus R, Göttgens B, et al. GENE SILENCING. Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science.* 2015;348: 1481–1485. doi:10.1126/science.aaa7227
128. Fukuda K, Shinkai Y. SETDB1-Mediated Silencing of Retroelements. *Viruses.* 2020;12: 596. doi:10.3390/v12060596
129. Müller I, Helin K. Keep quiet: the HUSH complex in transcriptional silencing and disease. *Nat Struct Mol Biol.* 2024;31: 11–22. doi:10.1038/s41594-023-01173-7
130. Pedersen BS, Eyring K, De S, Yang IV, Schwartz DA. Fast and accurate alignment of long bisulfite-seq reads. *arXiv;* 2014. doi:10.48550/arXiv.1401.1129
131. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience.* 2021;10: giab008. doi:10.1093/gigascience/giab008
132. R Core Team R. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2024. Available: <https://www.r-project.org/>
133. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for Computing and Annotating Genomic Ranges. *PLOS Comput Biol.* 2013;9: e1003118. doi:10.1371/journal.pcbi.1003118
134. Pagès H, Aboyoun P, Gentleman R, DebRoy S. Biostrings: Efficient manipulation of biological strings. 2024. Available: <https://bioconductor.org/packages/Biostrings>
135. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. *J Open Source Softw.* 2019;4: 1686. doi:10.21105/joss.01686
136. Nelson JD, Denisenko O, Bomsztyk K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat Protoc.* 2006;1: 179–185. doi:10.1038/nprot.2006.27
137. Conerly ML, Yao Z, Zhong JW, Groudine M, Tapscott SJ. Distinct Activities of Myf5 and MyoD Indicate Separate Roles in Skeletal Muscle Lineage Specification and Differentiation. *Dev Cell.* 2016;36: 375. doi:10.1016/j.devcel.2016.01.021
138. Rodriguez-Terrones D, Gaume X, Ishiuchi T, Weiss A, Kopp A, Kruse K, et al. A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat Genet.* 2018;50: 106–119. doi:10.1038/s41588-017-0016-5

139. Thijssen PE, Balog J, Yao Z, Pham TP, Tawil R, Tapscott SJ, et al. DUX4 promotes transcription of FRG2 by directly activating its promoter in facioscapulohumeral muscular dystrophy. *Skelet Muscle*. 2014;4: 19. doi:10.1186/2044-5040-4-19
140. Bello T, Paindelli C, Diaz-Gomez LA, Melchiorri A, Mikos AG, Nelson PS, et al. Computational modeling identifies multitargeted kinase inhibitors as effective therapies for metastatic, castration-resistant prostate cancer. *Proc Natl Acad Sci U S A*. 2021;118: e2103623118. doi:10.1073/pnas.2103623118
141. Gujral TS, Peshkin L, Kirschner MW. Exploiting polypharmacology for drug target deconvolution. *Proc Natl Acad Sci*. 2014;111: 5048–5053. doi:10.1073/pnas.1403080111
142. Anastassiadis T, Deacon SW, Devarajan K, Ma H, Peterson JR. Comprehensive assay of kinase catalytic activity reveals features of kinase inhibitor selectivity. *Nat Biotechnol*. 2011;29: 1039–1045. doi:10.1038/nbt.2017
143. Duc TN, Hassanzadeh-Ghassabeh G, Saerens D, Peeters E, Charlier D, Muyldermans S. Nanobody-based chromatin immunoprecipitation. *Methods Mol Biol Clifton NJ*. 2012;911: 491–505. doi:10.1007/978-1-61779-968-6_31
144. Fujita T, Fujii H. Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochem Biophys Res Commun*. 2013;439: 132–136. doi:10.1016/j.bbrc.2013.08.013
145. Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics*. 2020;36: 2628–2629. doi:10.1093/bioinformatics/btz931
146. Vastenhouw NL, Cao WX, Lipshitz HD. The maternal-to-zygotic transition revisited. *Development*. 2019;146. doi:10.1242/dev.161471
147. Sha Q-Q, Zheng W, Wu Y-W, Li S, Guo L, Zhang S. Dynamics and clinical relevance of maternal mRNA clearance during the oocyte-to-embryo transition in humans. *Nat Commun*. 2020;11. doi:10.1038/s41467-020-18680-6
148. Teixeira FK, Lehmann R. Translational Control during Developmental Transitions. *Cold Spring Harb Perspect Biol*. 2019;11: a032987. doi:10.1101/cshperspect.a032987
149. Sampath P, Pritchard DK, Pabon L, Reinecke H, Schwartz SM, Morris DR. A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. *Cell Stem Cell*. 2008;2: 448–460. doi:10.1016/j.stem.2008.03.013

150. Ingolia NT, Ghaemmamghami S, Newman JRS, Weissman JS. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*. 2009;324: 218–223. doi:10.1126/science.1168978
151. Signer RA, Qi L, Zhao Z, Thompson D, Sigova AA, Fan ZP. The rate of protein synthesis in hematopoietic stem cells is limited partly by 4E-BPs. *Genes Dev*. 2016;30: 1698–1703. doi:10.1101/gad.282756.116
152. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*. 2012;487: 57–63. doi:10.1038/nature11244
153. Hung SSC, Wong RCB, Sharov AA, Nakatake Y, Yu H, Ko MSH. Repression of Global Protein Synthesis by Eif1a-Like Genes That Are Expressed Specifically in the Two-Cell Embryos and the Transient Zscan4-Positive State of Embryonic Stem Cells. *DNA Res*. 2013;20: 391–402. doi:10.1093/dnares/dst018
154. Eckersley-Maslin MA, Svensson V, Krueger C, Stubbs TM, Giehr P, Krueger F. MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. *CellReports*. 2016;17: 179–192. doi:10.1016/j.celrep.2016.08.087
155. Himeda CL, Jones PL. The Genetics and Epigenetics of Facioscapulohumeral Muscular Dystrophy. *Annu Rev Genomics Hum Genet*. 2019;20: 265–291. doi:10.1146/annurev-genom-083118-014933
156. Young JM, Whiddon JL, Yao Z, Kasinathan B, Snider L, Geng LN, et al. DUX4 Binding to Retroelements Creates Promoters That Are Active in FSHD Muscle and Testis. Feschotte C, editor. *PLoS Genet*. 2013;9: e1003947. doi:10.1371/journal.pgen.1003947
157. Bosnakovski D, Gearhart MD, Toso EA, Ener ET, Choi SH, Kyba M. Low level DUX4 expression disrupts myogenesis through deregulation of myogenic gene expression. *Sci Rep*. 2018;8. doi:10.1038/s41598-018-35150-8
158. Dmitriev P, Bou Saada Y, Dib C, Anseau E, Barat A, Hamade A. DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. *Free Radic Biol Med*. 2016;99: 244–258. doi:10.1016/j.freeradbiomed.2016.08.007
159. Turki A, Hayot M, Carnac G, Pillard F, Passerieux E, Bommart S. Functional muscle impairment in facioscapulohumeral muscular dystrophy is correlated with oxidative stress and mitochondrial dysfunction. *Free Radic Biol Med*. 2012;53: 1068–1079. doi:10.1016/j.freeradbiomed.2012.06.041
160. Banerji CRS, Zammit PS. Pathomechanisms and biomarkers in facioscapulohumeral muscular dystrophy: roles of DUX4 and PAX7. *EMBO Mol Med*. 2021;13. doi:10.15252/emmm.202013695

161. Wong CJ, Wang LH, Friedman SD, Shaw D, Campbell AE, Budech CB. Longitudinal measures of RNA expression and disease activity in FSHD muscle biopsies. *Hum Mol Genet.* 2020;29: 1030–1043. doi:10.1093/hmg/ddaa031
162. Resnick R, Wong C-J, Hamm DC, Bennett SR, Skene PJ, Hake SB, et al. DUX4-Induced Histone Variants H3.X and H3.Y Mark DUX4 Target Genes for Expression. *Cell Rep.* 2019;29: 1812–1820.e5. doi:10.1016/j.celrep.2019.10.025
163. Shadle SC, Zhong JW, Campbell AE, Conerly ML, Jagannathan S, Wong C-J, et al. DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. Cox GA, editor. *PLOS Genet.* 2017;13: e1006658. doi:10.1371/journal.pgen.1006658
164. Sudhakar A, Ramachandran A, Ghosh S, Hasnain SE, Kaufman RJ, Ramaiah KV. Phosphorylation of serine 51 in initiation factor 2 alpha (eIF2 alpha) promotes complex formation between eIF2 alpha(P) and eIF2B and causes inhibition in the guanine nucleotide exchange activity of eIF2B. *Biochemistry.* 2000;39: 12929–12938. doi:10.1021/bi0008682
165. Pyronnet S, Imataka H, Gingras AC, Fukunaga R, Hunter T, Sonenberg N. Human eukaryotic translation initiation factor 4G (eIF4G) recruits mnk1 to phosphorylate eIF4E. *EMBO J.* 1999;18: 270–279. doi:10.1093/emboj/18.1.270
166. Gingras AC, Raught B, Sonenberg N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem.* 1999;68: 913–963. doi:10.1146/annurev.biochem.68.1.913
167. Hsieh AC, Costa M, Zollo O, Davis C, Feldman ME, Testa JR, et al. Genetic dissection of the oncogenic mTOR pathway reveals druggable addiction to translational control via 4EBP-eIF4E. *Cancer Cell.* 2010;17: 249–261. doi:10.1016/j.ccr.2010.01.021
168. Moerke NJ, Aktas H, Chen H, Cantel S, Reibarkh MY, Fahmy A. Small-Molecule Inhibition of the Interaction between the Translation Initiation Factors eIF4E and eIF4G. *Cell.* 2007;128: 257–267. doi:10.1016/j.cell.2006.11.046
169. Houghton FD. HYPOXIA AND REPRODUCTIVE HEALTH: Hypoxic regulation of preimplantation embryos: lessons from human embryonic stem cells. *Reproduction.* 2021;161. doi:10.1530/REP-20-0322
170. Baumann CG, Morris DG, Sreenan JM, Leese HJ. The quiet embryo hypothesis: Molecular characteristics favoring viability. *Mol Reprod Dev.* 2007;74: 1345–1353. doi:10.1002/mrd.20604
171. Leese HJ. Quiet please, do not disturb: a hypothesis of embryo metabolism and viability. *Bioessays.* 2002;24: 845–849. doi:10.1002/bies.10137

172. Rodriguez-Terrones D, Hartleben G, Gaume X, Eid A, Guthmann M, Iturbide A. A distinct metabolic state arises during the emergence of 2-cell-like cells. *EMBO Rep.* 2020;21. doi:10.15252/embr.201948354
173. Atashpaz S, Samadi Shams S, Gonzalez JM, Sebestyén E, Arghavanifard N, Gnocchi A. ATR expands embryonic stem cell fate potential in response to replication stress. *Elife.* 2020;9. doi:10.7554/eLife.54756
174. Srinivasan R, Nady N, Arora N, Hsieh LJ, Swigut T, Narlikar GJ. Zscan4 binds nucleosomal microsatellite DNA and protects mouse two-cell embryos from DNA damage. *Sci Adv.* 2020;6. doi:10.1126/sciadv.aaz9115
175. Lek A, Zhang Y, Woodman KG, Huang S, DeSimone AM, Cohen J, et al. Applying genome-wide CRISPR-Cas9 screens for therapeutic discovery in facioscapulohumeral muscular dystrophy. *Sci Transl Med.* 2020;12: eaay0271. doi:10.1126/scitranslmed.aay0271
176. Heher P, Ganassi M, Weidinger A, Engquist EN, Pruller J, Nguyen TH. Interplay between mitochondrial reactive oxygen species, oxidative stress and hypoxic adaptation in facioscapulohumeral muscular dystrophy: Metabolic stress as potential therapeutic target. *Redox Biol.* 2022;51. doi:10.1016/j.redox.2022.102251
177. Cockman E, Anderson P, Ivanov P. TOP mRNPs: Molecular Mechanisms and Principles of Regulation. *Biomolecules.* 2020;10: 969. doi:10.3390/biom10070969
178. Hsieh AC, Liu Y, Edlind MP, Ingolia NT, Janes MR, Sher A, et al. The translational landscape of mTOR signalling steers cancer initiation and metastasis. *Nature.* 2012;485: 55–61. doi:10.1038/nature10912
179. Thoreen CC, Chantranupong L, Keys HR, Wang T, Gray NS, Sabatini DM. A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature.* 2012;485: 109–113. doi:10.1038/nature11083
180. Gismondi A, Caldarola S, Lisi G, Juli G, Chellini L, Iadevaia V, et al. Ribosomal stress activates eEF2K–eEF2 pathway causing translation elongation inhibition and recruitment of Terminal Oligopyrimidine (TOP) mRNAs on polysomes. *Nucleic Acids Res.* 2014;42: 12668–12680. doi:10.1093/nar/gku996
181. Knopp P, Krom YD, Banerji CRS, Panamarova M, Moyle LA, den Hamer B, et al. DUX4 induces a transcriptome more characteristic of a less-differentiated cell state and inhibits myogenesis. *J Cell Sci.* 2016;129: 3816–3831. doi:10.1242/jcs.180372
182. Bosnakovski D, Toso EA, Hartweck LM, Magli A, Lee HA, Thompson ER. The DUX4 homeodomains mediate inhibition of myogenesis and are functionally exchangeable with the Pax7 homeodomain. *J Cell Sci.* 2017;130: 3685–3697. doi:10.1242/jcs.205427

183. Yao Z, Snider L, Balog J, Lemmers RJLF, Van Der Maarel SM, Tawil R, et al. DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum Mol Genet.* 2014;23: 5342–5352. doi:10.1093/hmg/ddu251
184. Kozak M. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol.* 1989;9: 5134–5142. doi:10.1128/mcb.9.11.5134-5142.1989
185. Svitkin YV, Pause A, Haghighat A, Pyronnet S, Witherell G, Belsham GJ. The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *RNA.* 2001;7: 382–394. doi:10.1017/s135583820100108x
186. Dieudonne FX, O'Connor PB, Gubler-Jaquier P, Yasrebi H, Conne B, Nikolaev S. The effect of heterogeneous Transcription Start Sites (TSS) on the translome: implications for the mammalian cellular phenotype. *BMC Genomics.* 2015;16. doi:10.1186/s12864-015-2179-8
187. Liu B, Qian SB. Translational reprogramming in cellular stress response. *Wiley Interdiscip Rev RNA.* 2014;5: 301–315. doi:10.1002/wrna.1212
188. Wang R, Amoyel M. mRNA Translation Is Dynamically Regulated to Instruct Stem Cell Fate. *Front Mol Biosci.* 2022;9. doi:10.3389/fmolb.2022.863885
189. Kovalski K-O JR, D R, D. Protein synthesis control in cancer: selectivity and therapeutic targeting. *EMBO J.* 2022;41. doi:10.15252/embj.2021109823
190. Leprivier G, Rotblat B, Khan D, Jan E, Sorensen PH. Stress-mediated translational control in cancer cells. *Biochim Biophys Acta.* 2015;1849: 845–860. doi:10.1016/j.bbarm.2014.11.002
191. Brennan CM, Hill AS, Andre M, Li X, Madeti V, Breitkopf S. DUX4 expression activates JNK and p38 MAP kinases in myoblasts. *Model Mech.* 2022;15. doi:10.1242/dmm.049516
192. Campbell AE, Dyle MC, Albanese R, Matheny T, Sudheendran K, Cortazar MA. Compromised nonsense-mediated RNA decay results in truncated RNA-binding protein production upon DUX4 expression. *Cell Rep.* 2023;42. doi:10.1016/j.celrep.2023.112642
193. Zhang C, Wang M, Li Y, Zhang Y. Profiling and functional characterization of maternal mRNA translation during mouse maternal-to-zygotic transition. *Sci Adv.* 2022;8. doi:10.1126/sciadv.abj3967
194. Zhu L, Zhou T, Iyyappan R, Ming H, Dvoran M, Wang Y, et al. High-resolution ribosome profiling reveals translational selectivity for transcripts in bovine preimplantation embryo development. *Dev Camb Engl.* 2022;149: dev200819. doi:10.1242/dev.200819

195. Zou Z, Zhang C, Wang Q, Hou Z, Xiong Z, Kong F. Translatome and transcriptome co-profiling reveals a role of TPRXs in human zygotic genome activation. *Science*. 2022;378. doi:10.1126/science.abo7923
196. Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee S-L. Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature*. 2010;464: 858–863. doi:10.1038/nature08882
197. Ishiuchi T, Enriquez-Gasca R, Mizutani E, Bošković A, Ziegler-Birling C, Rodriguez-Terrones D. Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat Struct Mol Biol*. 2015;22: 662–671. doi:10.1038/nsmb.3066
198. Sun Z, Yu H, Zhao J, Tan T, Pan H, Zhu Y. LIN28 coordinately promotes nucleolar/ribosomal functions and represses the 2C-like transcriptional program in pluripotent stem cells. *Protein Cell*. 2022;13: 490–512. doi:10.1007/s13238-021-00864-5
199. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature*. 2001;414: 105–111. doi:10.1038/35102167
200. Bhat M, Robichaud N, Hulea L, Sonenberg N, Pelletier J, Topisirovic I. Targeting the translation machinery in cancer. *Nat Rev Drug Discov*. 2015;14: 261–278. doi:10.1038/nrd4505
201. C P, BA W, P G, RJ S. Translation initiation factors and their relevance in cancer. *Curr Opin Genet Dev*. 2018;48: 82–88. doi:10.1016/j.gde.2017.11.001
202. Jana S, Deo R, Hough RP, Liu Y, Horn JL, Wright JL, et al. mRNA translation is a therapeutic vulnerability necessary for bladder epithelial transformation. *JCI Insight*. 2021;6: 144920. doi:10.1172/jci.insight.144920
203. Liu Y, Horn JL, Banda K, Goodman AZ, Lim Y, Jana S. The androgen receptor regulates a druggable translational regulon in advanced prostate cancer. *Sci Transl Med*. 2019;11. doi:10.1126/scitranslmed.aaw4993
204. Xu Y, Ruggero D. The Role of Translation Control in Tumorigenesis and Its Therapeutic Implications. *Annu Rev Cancer Biol*. 2020;4: 437–457.
205. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature*. 2013;501: 328–337. doi:10.1038/nature12624
206. Holcik M. Could the eIF2 α -Independent Translation Be the Achilles Heel of Cancer? *Front Oncol*. 2015;5.
207. Xu Y, Poggio M, Jin HY, Shi Z, Forester CM, Wang Y. Translation control of the immune checkpoint in cancer and its therapeutic targeting. *Nat Med*. 2019;25: 301–311. doi:10.1038/s41591-018-0321-2

208. Siddiqui N, Sonenberg N. Signalling to eIF4E in cancer. *Biochem Soc Trans.* 2015;43: 763–772. doi:10.1042/BST20150126
209. Karakas D, Ozpolat B. Eukaryotic elongation factor-2 kinase (eEF2K) signaling in tumor and microenvironment as a novel molecular target. *J Mol Med Berl.* 2020;98: 775–787. doi:10.1007/s00109-020-01917-8
210. Schulz KN, Harrison MM. Mechanisms regulating zygotic genome activation. *Nat Rev Genet.* 2019;20: 221–234. doi:10.1038/s41576-018-0087-x
211. Li Y, Banerjee S, Goldstein SA, Dong B, Gaughan C, Rath S. Ribonuclease I mediates the cell-lethal phenotype of double-stranded RNA editing enzyme ADAR1 deficiency in a human cell line. *Elife.* 2017;6.
212. Calviello L, Mukherjee N, Wyler E, Zauber H, Hirsekorn A, Selbach M. Detecting actively translated open reading frames in ribosome profiling data. *Nat Methods.* 2016;13: 165–170. doi:10.1038/nmeth.3688
213. Lorenz R, Bernhart SH, Siederdisen C, Tafer H, Flamm C, Stadler PF. ViennaRNA Package 2.0. *Algorithms Mol Biol.* 2011;6. doi:10.1186/1748-7188-6-26
214. Karpukhina A, Vassetzky Y. DUX4, a Zygotic Genome Activator, Is Involved in Oncogenesis and Genetic Diseases. *Ontog Russ J Dev Biol.* 2020;51: 176–182. doi:10.1134/S1062360420030078
215. Grow EJ, Weaver BD, Smith CM, Guo J, Stein P, Shadle SC, et al. p53 convergently activates Dux/DUX4 in embryonic stem cells and in facioscapulohumeral muscular dystrophy cell models. *Nat Genet.* 2021;53: 1207–1220. doi:10.1038/s41588-021-00893-0
216. Bosnakovski D, Xu Z, Gang EJ, Galindo CL, Liu M, Simsek T, et al. An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J.* 2008;27: 2766–2779. doi:10.1038/emboj.2008.201
217. Wallace LM, Garwick SE, Mei W, Belayew A, Coppee F, Ladner KJ, et al. DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann Neurol.* 2011;69: 540–552. doi:10.1002/ana.22275
218. Kretzschmar K, Watt FM. Lineage Tracing. *Cell.* 2012;148: 33–45. doi:10.1016/j.cell.2012.01.002
219. Van Duyne GD. Cre Recombinase. *Microbiol Spectr.* 2015;3: MDNA3-0014–2014. doi:10.1128/microbiolspec.MDNA3-0014-2014

220. Yang YS, Hughes TE. Cre Stoplight: A Red/Green Fluorescent Reporter of Cre Recombinase Expression in Living Cells. *BioTechniques*. 2001;31: 1036–1041. doi:10.2144/01315st03
221. Casanova E, Fehsenfeld S, Lemberger T, Shimshek DR, Sprengel R, Mantamadiotis T. ER-based double iCre fusion protein allows partial recombination in forebrain. *Genes N Y N* 2000. 2002;34: 208–214. doi:10.1002/gene.10153
222. Matsuda T, Cepko CL. Controlled expression of transgenes introduced by in vivo electroporation. *Proc Natl Acad Sci U S A*. 2007;104: 1027–1032. doi:10.1073/pnas.0610155104
223. Le Y-Z, Zheng W, Rao P-C, Zheng L, Anderson RE, Esumi N, et al. Inducible expression of cre recombinase in the retinal pigmented epithelium. *Invest Ophthalmol Vis Sci*. 2008;49: 1248–1253. doi:10.1167/iovs.07-1105
224. Verkhusha VV, Kuznetsova IM, Stepanenko OV, Zaraisky AG, Shavlovsky MM, Turoverov KK, et al. High stability of Discosoma DsRed as compared to Aequorea EGFP. *Biochemistry*. 2003;42: 7879–7884. doi:10.1021/bi034555t
225. Feil S, Valtcheva N, Feil R. Inducible Cre mice. *Methods Mol Biol Clifton NJ*. 2009;530: 343–363. doi:10.1007/978-1-59745-471-1_18
226. Teveroni E, Pellegrino M, Sacconi S, Calandra P, Cascino I, Farioli-Vecchioli S, et al. Estrogens enhance myoblast differentiation in facioscapulohumeral muscular dystrophy by antagonizing DUX4 activity. *J Clin Invest*. 2017;127: 1531–1545. doi:10.1172/JCI89401
227. Higashi AY, Ikawa T, Muramatsu M, Economides AN, Niwa A, Okuda T, et al. Direct Hematological Toxicity and Illegitimate Chromosomal Recombination Caused by the Systemic Activation of CreERT21. *J Immunol*. 2009;182: 5633–5640. doi:10.4049/jimmunol.0802413
228. Kurachi M, Ngiow SF, Kurachi J, Chen Z, Wherry EJ. Hidden Caveat of Inducible Cre Recombinase. *Immunity*. 2019;51: 591–592. doi:10.1016/j.immuni.2019.09.010
229. Yang S-H, Cheng P-H, Sullivan RT, Thomas JW, Chan AWS. Lentiviral Integration Preferences in Transgenic Mice. *Genes N Y N* 2000. 2008;46: 711–718. doi:10.1002/dvg.20435
230. Maksakova IA, Goyal P, Bullwinkel J, Brown JP, Bilenky M, Mager DL, et al. H3K9me3-binding proteins are dispensable for SETDB1/H3K9me3-dependent retroviral silencing. *Epigenetics Chromatin*. 2011;4: 12. doi:10.1186/1756-8935-4-12

231. Kourmouli N, Sun Y, van der Sar S, Singh PB, Brown JP. Epigenetic regulation of mammalian pericentric heterochromatin in vivo by HP1. *Biochem Biophys Res Commun.* 2005;337: 901–907. doi:10.1016/j.bbrc.2005.09.132
232. Li H, Rauch T, Chen Z-X, Szabó PE, Riggs AD, Pfeifer GP. The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *J Biol Chem.* 2006;281: 19489–19500. doi:10.1074/jbc.M513249200
233. Fei Q, Yang X, Jiang H, Wang Q, Yu Y, Yu Y, et al. SETDB1 modulates PRC2 activity at developmental genes independently of H3K9 trimethylation in mouse ES cells. *Genome Res.* 2015;25: 1325–1335. doi:10.1101/gr.177576.114
234. Chandrasekaran TT, Choudalakis M, Bröhm A, Weirich S, Kouroukli AG, Ammerpohl O, et al. SETDB1 activity is globally directed by H3K14 acetylation via its Triple Tudor Domain. *Nucleic Acids Res.* 2024;52: 13690–13705. doi:10.1093/nar/gkae1053
235. Markouli M, Strepkos D, Piperi C. Structure, Activity and Function of the SETDB1 Protein Methyltransferase. *Life.* 2021;11: 817. doi:10.3390/life11080817
236. Daxinger L, Harten SK, Oey H, Epp T, Isbel L, Huang E, et al. An ENU mutagenesis screen identifies novel and known genes involved in epigenetic processes in the mouse. *Genome Biol.* 2013;14: R96. doi:10.1186/gb-2013-14-9-r96
237. Perdiguero E, Ruiz-Bonilla V, Gresh L, Hui L, Ballestar E, Sousa-Victor P, et al. Genetic analysis of p38 MAP kinases in myogenesis: fundamental role of p38 α in abrogating myoblast proliferation. *EMBO J.* 2007;26: 1245–1256. doi:10.1038/sj.emboj.7601587
238. Keren A, Tamir Y, Bengal E. The p38 MAPK signaling pathway: a major regulator of skeletal muscle development. *Mol Cell Endocrinol.* 2006;252: 224–230. doi:10.1016/j.mce.2006.03.017
239. Lluís F, Perdiguero E, Nebreda AR, Muñoz-Cánoves P. Regulation of skeletal muscle gene expression by p38 MAP kinases. *Trends Cell Biol.* 2006;16: 36–44. doi:10.1016/j.tcb.2005.11.002
240. Petrova V, Annicchiarico-Petruzzelli M, Melino G, Amelio I. The hypoxic tumour microenvironment. *Oncogenesis.* 2018;7: 1–13. doi:10.1038/s41389-017-0011-9
241. Simon MC, Keith B. The role of oxygen availability in embryonic development and stem cell function. *Nat Rev Mol Cell Biol.* 2008;9: 285–296. doi:10.1038/nrm2354