

The DUX4
cytotoxic cascade,
and CRISPR
mitigation
methods

Ator Ashoti

The work described in this thesis was performed at the Hubrecht institute for Developmental Biology and Stem Cell Research (The Royal Netherlands Academy of Arts and Sciences, KNAW) within the framework of the PhD programme Regenerative Medicine (RM), which is part of the Utrecht Graduate school of Life Sciences (Utrecht University).

Cover: "Lamassu" by Batul Ashoti

Layout and design by Ator Ashoti and Batul Ashoti

Printed by Ridderprint: www.ridderprint.nl

ISBN: 978-94-6416-638-5

Copyright © 2021 by Ator Ashoti. All right reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means, without priot permission of the author.

The DUX4 cytotoxic cascade, and CRISPR mitigation methods

De DUX4-cytotoxische cascade, en de CRISPR-mitigatie methoden
(Met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

donderdag 17 juni 2021
des middags te 12.15 uur

door

Ator Rafael Odisho Ashoti

geboren op 25 november 1989
te Dihok, Irak

Promotoren:

Prof. dr. W.L. de Laat

Prof. dr. N. Geijsen

Table of Content

- Chapter 1 7
The Mystery that is FSHD
- Chapter 2 27
Generation of a cellular model to dissect early molecular events leading to DUX4-induced toxicity
- Chapter 3 85
DUX4 induces a homogeneous sequence of molecular changes, culminating in cellular apoptosis
- Chapter 4 143
A genome-wide CRISPR/Cas phenotypic screen for modulators of DUX4 cytotoxicity reveals screen complications
- Chapter 5 179
CRISPR-mediated functional silencing of DUX4
- Chapter 6 195
General discussion
- Addendum 203
 - Nederlandse samenvatting 204
 - Acknowledgements 206
 - Curriculum Vitae 210
 - List of publications 211



Illustration based on a stone carving on display at the Louvre

Chapter 1

The Mystery that is FSHD

Ator Ashoti & Niels Geijsen

FSHD: A brief history

1

Facioscapulohumeral muscular dystrophy (FSHD) is one of the most prevalent genetic muscular disorders¹. The disorder was documented for the first time over a 136 years ago in 1884 by Landouzy and Dejerine² and was further investigated in the 1900s, with a large familial study published in 1950. This large study of 1249 descendants included physical examination of 240 individuals and data gathering of deceased family members to create a pedigree. Of the 1249 descendants, 159 had the familial anomaly. Due to the large size of the individual family groups, with some of them being of a polygamous nature, the frequency of the disease was much higher within this family tree than in the general population. This gave the authors a unique chance for a large case study, where the pattern of inheritance and clinical features were documented and described³. FSHD is presently described as a hereditary autosomal dominant trait, however, 10-30% of the cases originate sporadically by de novo mutations⁴⁻⁸. Many of these sporadic cases are somatic mosaic, which likely originate from mitotic repeat rearrangement⁹⁻¹⁵.

FSHD patients experience muscle atrophy in a asymmetric fashion, starting in the facial muscles and muscles of the shoulder blades, and slowly progresses to the muscles in the upper arms, areas of the truck and in some cases the muscles in the lower extremities^{3,16}. The prognosis for FSHD is compared to other muscular dystrophies one of the best, as it generally progresses at a slow pace and rarely effects cardiac output, with the majority of the cases having a normal life expectancy^{16,17}. However, the psychological and psychosocial impact can be severe, as the facial muscles that show human emotion are the first to be affected.

Through microsatellite linkage analysis it was determined in 1990 that the origin of FSHD lies on chromosome 4^{18,19}. This was quickly narrowed down to the subtelomeric region on the q-arm (4q35), often referred to as the 4q35-ter or 4qter^{20,21}. In 1992 it was established that FSHD was linked to a 3.2kb repeated structure in the 4qter which was named a D4Z4 macrosatellite repeat sequence²². A contraction of these repeated sequences was linked to the development of FSHD. The number of repeats can range between 1-100, arranged in a head to tail orientation. Healthy individuals generally possess between 11-100 repeats, whereas 95% of the FSHD patients display a contraction to 1-10 D4Z4 repeats^{4,22,23}. A homeobox-containing gene was a likely suspect driving FSHD pathogenesis, since two homeobox sequences were identified within the 3.2kb repeat^{4,22}.

While these repeated structures were studied more in depth, assays for detecting D4Z4 sequences became more specific, using southern blot analysis and in-situ hybridization. It became apparent that the D4Z4 repeats were not restricted to the 4qter, but were also found at multiple other loci in the human genome. D4Z4 copies were found on chromosomes 1, 3, 9, 13, 14, 15, 20, 21, 22 and Y, and as a similar tandem repeat structure on the q-arm of chromosome 10 (10qter, 10q26)²³⁻²⁵. This repeat array on chromosome 10q26 shares 98% in sequence homology to chromosome 4q35²⁶, yet the 10qter shows no association with FSHD²⁶⁻²⁸. Fortunately for diagnostic purposes, these two highly similar tandem repeat arrays could still be discriminated from one another due to a specific BlnI (also known as AvrII) restriction enzyme recognition site on each D4Z4 repeat located on chromosome 10q26²⁹.

At the same time, it was also discovered that the double homeodomains found in the D4Z4 repeat units were contained in an open reading frame (ORF). The authors assumed it to be unlikely that the ORF would code for a functional protein, since no transcript of this ORF or any other sequences found at or around the D4Z4 repeated array had been identified^{23,24}.

Additionally, hybrid repeat arrays (chromosome 10 repeats found on chromosome 4 and vice versa) found in healthy individuals contributed to the idea that these repeat sequences do not encode for a functional protein, as they believed that this rearrangement would disrupt the FSHD-related gene²⁸. A different scenario was therefore proposed, in which FSHD was caused by a position effect due to the large deletions on the 4qter, and deletion of a critical number of D4Z4 repeat units could affect the expression of genes located in close proximity of this truncated repeated array^{23,24,28}.

Hewitt et al. hypothesized that if the ORF would produce a functional protein, it would either be a large polymorphic gene encoding multiple homeodomains, or only one copy would be responsible for the production of a functional protein²³. This last assumption, as we now know, proved to be true. A minimum of one D4Z4 repeat, containing an ORF that was later identified as double homeobox 4 (DUX4)³⁰, is necessary for the development of FSHD³¹. This gene is a pioneer transcription factor^{32,33}, that is normally expressed during early embryonic development (4-cell stage)^{34,35} and in the thymus³⁶ and testis³⁷. The first evidence supporting the involvement of DUX4 in FSHD was published in 2007, by Kowaljow et al.³⁸ and Dixit et al.³⁹. The authors showed an upregulation of DUX4 in FSHD muscle biopsies compared to biopsies of healthy controls^{38,39}, and the pro-apoptotic feature of DUX4³⁸. This result was quickly corroborated by the group of Stephan Tapscott, indeed showing an upregulation of DUX4 expression in FSHD-derived muscle cells, together with many other sense and anti-sense RNA transcripts, novel mRNAs and other RNA fragments that are encoded within the D4Z4 repeat array. They furthermore confirmed the hypothesis of Hewitt et al. that one copy of the DUX4 ORF is involved with pathophysiology of FSHD, as they showed that a polyadenylated DUX4 transcript comes from the most distal (most telomeric) D4Z4 repeat⁴⁰.

Genetic background and DUX4 expression

To uncover why a contracted repeat array on chromosome 10 is not associated with FSHD, differences in telomeric structures between the 10qter and the disease-linked 4qter were studied. Both chromosomes contain a sequence directly adjacent to the most distal D4Z4 unit, called the pLAM sequence. This sequence was previously used for the characterization of rearranged D4Z4 fragments, through the use of the pLAM probe²². Both chromosomes also possess an inverted D4Z4 repeat 42kb upstream of the main repeat array. However, the inverted repeat on chromosome 10 misses a portion, which also happens to be the breakpoint in the 4q and 10q proximal homology. Downstream of this breakpoint the two chromosomes share a high degree of sequence homology (Fig. 1). In the process of uncovering the differences between the 4qter and 10qter, the authors also found two variants of the 4qter: one containing a pLAM sequence (4qA), and the other not (4qB)⁴¹ (Fig. 1). These two variants can be found almost equally frequent in the population, yet only 4qA is associated with FSHD⁴²⁻⁴⁴. When the role of DUX4 in the development of FSHD was established, with the discovery of a stable polyadenylated DUX4 transcript in FSHD-affected muscle cells^{39,40}, it did not take long to connect the missing piece as to why only 4qA is linked to the development of FSHD. About a year after, a genetic model for FSHD was published, where the authors identified an ATAAA polyadenylation signal (PolyA) in the pLAM region on the 4qA allele. Since the 4qB allele lacks the pLAM sequence, DUX4 transcripts from this allele are not polyadenylated, which is necessary to stabilize the DUX4 transcript. While a pLAM sequence is present in the most distal 10qter repeat, a single nucleotide polymorphism (ATCAA)⁴⁵ at this locus disrupts the poly-adenylation sequence (Fig. 1). Without this essential polyA sequence, the DUX4 protein cannot be stably expressed.

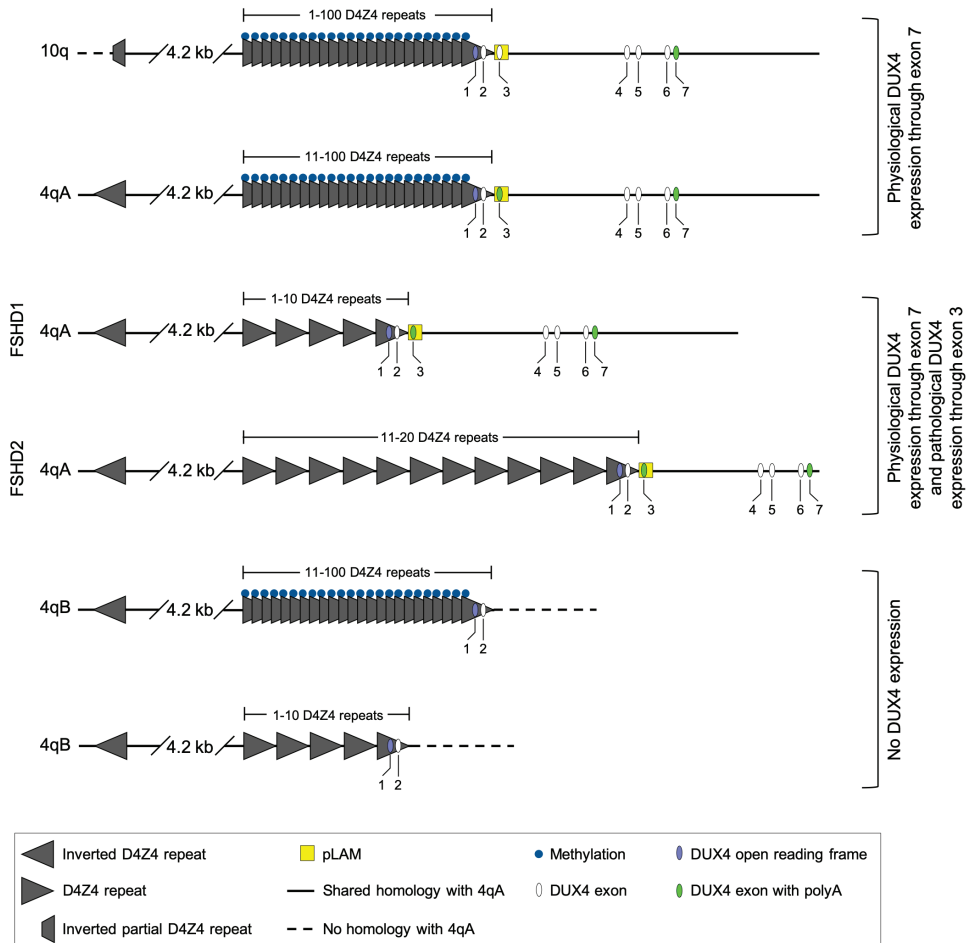


Figure 1. Overview of the D4Z4 tandem repeat arrays on chromosome 4 and 10. Schematic representation of the organization of the D4Z4 tandem repeat arrays and the DUX4 gene in healthy and FSHD affected individuals, and the shared sequence homology between the 4qter and 10qter, deduced from van Geel et al.⁴¹, Lemmers et al.⁴², Lemmers et al.⁴⁵, and Snider et al.³⁷. Physiological DUX4 expression occurs through the use of the polyadenylation signal in exon 7, which can be found on chromosome 10 and 4qA. Pathological DUX4 expression occurs on a 4qA allele through the use of the polyadenylation signal on exon 3, and after loss of methylation at the 4q35 loci. DUX4 is not expressed on a 4qB allele, due to a lack of a polyadenylation signal.

Specific haplotypes of the 4qter were identified, based on subtle and consistent sequence variations in the D4Z4 repeated array, and its flanking regions⁴⁵⁻⁴⁷. Among the most common haplotypes (4A161, 4B163 and 4A166), only contractions on 4A161 are pathogenic, due to the fact that this haplotype contains a poly-adenylation signal in exon 3, which stabilizes the DUX4 transcript⁴⁵⁻⁴⁷.

The DUX4 transcription factor is physiologically expressed during early embryonic development, as well as in the adult testis³⁷ and thymus³⁶. Stabilization of the physiological transcript is regulated through a different polyA sequence than the one found in exon 3 in the 4qA genetic background (Fig. 1). Downstream of the most distal D4Z4 unit lie 4

more additional exons, exon 4 to 7. Exon 7 contains a polyA sequence which appears to be more tightly regulated, and is the one used for the physiological expression of DUX4 during development and in mature tissues such as the testis and thymus. It therefore appears that the polyA sequence in exon 3 is pathological, as this transcript is only found in FSHD affected muscle cells³⁷.

FSHD2

The contracted D4Z4 repeat array on chromosome 4q35 was only found in 90-95% of all FSHD cases (FSHD1)^{10,48}. This suggested the existence of a second locus or event linked to the disorder. These FSHD2 patients with normal a D4Z4 repeat length are clinically indistinguishable from FSHD1 patients that carry a contracted D4Z4 repeat array⁴⁹⁻⁵². A common feature between FSHD1 and FSHD2 is the presence of a permissive, hypomethylated 4qA allele^{50,51}. In FSHD1 patients, the contraction of D4Z4 repeat array itself causes the loss of methylation and repressive chromatin, which leads to the permissive state of the D4Z4 array.^{22,23,49} While the D4Z4 region in FSHD2 patients is not contracted, all D4Z4 repeat arrays in FSHD2 subjects are hypomethylated, which includes both the 4q and 10q alleles. This is in contrast to FSHD1 patients, in which only the contracted repeat array is hypomethylated⁴⁹⁻⁵². The fact that all D4Z4 repeat arrays are hypomethylated in FSHD2 subjects implies the loss of a gene responsible for the methylation of these loci. Indeed, many FSHD2 patients possessed a heterozygous mutation in Structural Maintenance of Chromosome Flexible Hinge Domain Containing gene 1 (SMCHD1)⁵³, a gene known for its role in X inactivation through the hypermethylation of CpG islets⁵⁴⁻⁵⁶. The loss of SMCHD1 co-segregates with the hypomethylated status of the D4Z4 repeat array, and even heterozygous loss of SMCHD1 can thus cause the hypomethylated state on the D4Z4 arrays in patients diagnosed with FSHD2^{53, 57}.

In recent years, DNA Methyltransferase 2 Beta (DNMT3B) has also been identified in rare cases of FSHD2⁵⁸. This gene is involved in de novo methylation during early embryonic development and likely plays a role in the hypermethylation and inactivation of the D4Z4 array as well^{56,59-61}.

Notably, even though the D4Z4 repeat array in FSHD2 patients is not considered contracted (<10), the number of D4Z4 repeat units in most FSHD2 patients is lower (11-16) than most healthy individuals (11-100)^{52,62-65}. This suggests that haploinsufficiency of SMCHD1 is on its own not sufficient to fully derepress the permissive D4Z4 array, unless the number of D4Z4 repeats drops below a certain threshold (Fig. 1).

Thus, both FSHD1 and FSHD2 are caused by the inheritance of at least two dominant traits, a FSHD-permissive 4qA allele, and hypomethylated D4Z4 repeat array caused through either a contraction event or a mutated modifier gene. These events lead to the misexpression of DUX4 and subsequently the development of FSHD^{45,53}.

D4Z4 contractions in FSHD

As described above, hypomethylation of the D4Z4 array can be caused in two ways: the contraction of the D4Z4 repeated array, or a mutation in a chromatin modifier gene (e.g. SMCHD1) necessary to establish and/or maintain the hypermethylated status of this locus. There are several hypotheses about how a contraction of the D4Z4 array leads to a more relaxed chromatin, which subsequently initiates the transcription of the DUX4 gene. The D4Z4 array is often described as heterochromatin as it has some similar features: its proximity to telomeres, an unusually high GC content, the presence of hsp3 and LSau repeats that are

predominantly found in heterochromatin regions within the human genome^{23,24}, and the abundance of H3K9me3 and H3K27me3 marks^{36,66}. The loss of some of the heterochromatin signature may lead to local chromatin relaxation, allowing the transcription of genes within the area. However, it is argued that the D4Z4 array is missing an important feature of heterochromatin, as the H4 acetylation levels at the D4Z4 array are not low enough to be classified as heterochromatin, and corresponds more to that of unexpressed euchromatin⁶⁷. One study has shown that a contracted D4Z4 array enables the binding of CCCTC-binding factor (CTCF) and A-type Lamins to the contracted array, which could change the spatial positioning of the 4qter in the nuclear envelope. They hypothesize that a normal length D4Z4 array keeps the 4qter in a repressive compartment, and that binding of CTCF and A-type lamins to a contracted D4Z4 array positions the 4qter in a more permissive compartment at the nuclear envelope⁶⁸. Other chromatin-binding proteins that bind to the D4Z4 repeat array and influence the expression of nearby genes have also been identified. These proteins: YY1, HMGB2, nucleolin and EZH2, are either part of, or are associated with the polycomb group (PcG)^{69,70}. PcG complexes are known for their repressive effects on gene expression by adding repressive histone modification marks to nucleosomal histones⁷¹. The lowered occupancy of these proteins at the D4Z4 repeated array of FSHD-affected muscle cells leads to a reduction of these repressive marks, like the repressive histone mark H3K27me3⁷⁰. It appears that the loss of a piece of chromatin at the 4qter carrying essential repressive features, including DNA methylation and binding motifs for repressive proteins, causes major epigenetic dysregulation upon their loss, which leads to an open locus that is permissive for transcription.

The mechanism of the D4Z4 contraction is a topic of discussion as well. Due to the telomeric location of the D4Z4 repeats, rearrangements of this region during either meiosis or mitosis are likely to occur. With FSHD1, the contraction of the D4Z4 repeated array primarily occurs during mitotic cell division in early embryonic development^{14,45,72}. D4Z4 rearrangement can occur through either intrachromosomal or interchromosomal rearrangements, with interchromosomal rearrangements appearing to be the more common event^{14,72}. Partners for interchromosomal rearrangement in FSHD1 can be sister chromatids, or chromosome 10, as this D4Z4 repeat array shares high sequence homology with the repeat array on the 4qter^{26,41}. Interchromosomal rearrangement with the sister chromatid as a partner seems to be a logical option, as this plays a major role in double-stranded break repair in mammalian cells⁷³, however, as of yet no FSHD1 cases caused by this type of rearrangement have been identified. The more likely course of events is therefore mitotic interchromosomal rearrangements between the 4qter and 10qter⁷², as several of these types rearrangements have been identified^{14,48,72,74}. See figure two for a schematic overview.

FSHD: a muscle-specific disorder

FSHD is described as a muscle disorder, because it mainly effects muscle tissue. Other tissues are either less severely affected, have a lower impact on the patient's quality of life, or are rare occurrences only effecting a small percentage of patients. These symptoms include mild to moderate retinal pathologies, high-tone hearing loss, and in rare cases, that are predominantly early onset, patients can suffer from intellectual disabilities and epilepsy^{5,75-78}. Skeletal muscle is the most affected tissue, likely due to the cellular structure and other muscle-specific characteristics. Muscle fibers are long multinucleated structures, some reaching ~20 cm⁷⁹, containing dozens of myonuclei per mm of fiber.

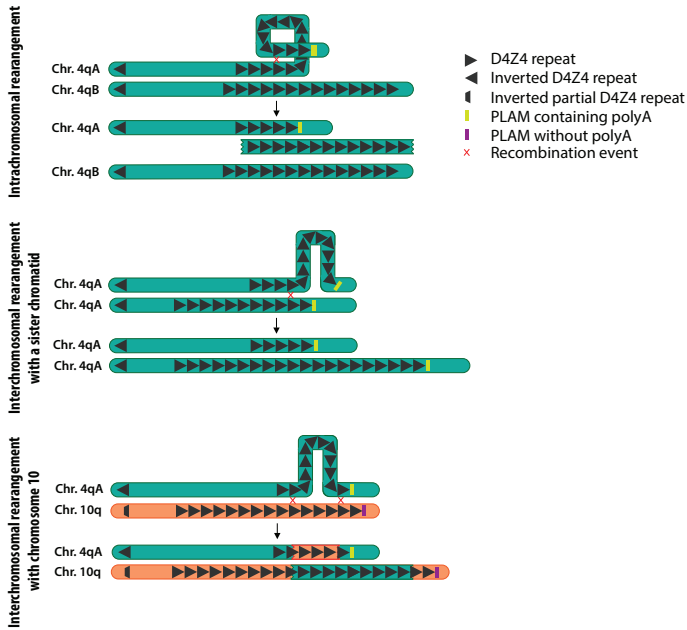


Figure 2. Schematic representation of chromosomal rearrangement events between D4Z4 arrays on chromosome 4 and 10. Top panel: Intrachromosomal rearrangement between D4Z4 repeats on chromosome 4. Middle panel: Interchromosomal rearrangement between D4Z4 repeat arrays on chromosome 4 sister chromatids. Bottom panel: Interchromosomal rearrangement between D4Z4 repeat arrays on chromosome 4 and chromosome 10.

Thus, depending on the size of the fiber, many will contain hundreds or thousands of nuclei^{80–82}. Considering that FSHD is caused by a burst of DUX4 expression in approximately 1/200 to 1/1000 nuclei^{37,83,84}, many affected muscle fibers of FSHD patients will likely contain one or more of these nuclei. These bursts of DUX4 expressions are rare, with the translation of the DUX4 transcript occurring in the cytoplasm of the myofiber. Both DUX4 transcript and protein can diffuse to other parts of the myofiber, forming a gradient^{84–86}.

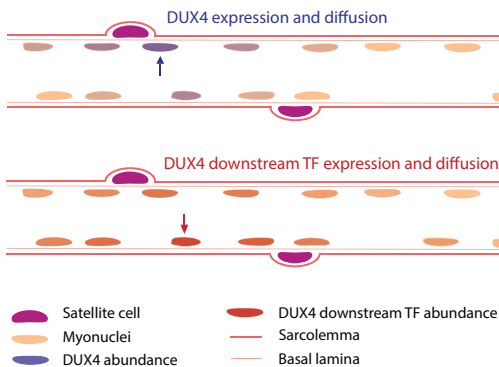


Figure 3. Depiction of the diffusion of DUX4 protein and DUX4 regulated transcription factors within a myofiber. Panels demonstrate a section of a myofiber containing multiple nuclei. Top panel: A DUX4 burst expression occurs in one nucleus (blue arrow), which diffuses and is taken up in the surrounding nuclei. As DUX4 diffuses away from the DUX4 expressing nuclei, it forms a gradient in the surrounding area, which is also reflected in the surrounding nuclei. Bottom panel: an example of surrounding nuclei taking in small amounts of DUX4 protein, which induced the expression of another transcription factor that forms its own gradient (red arrow).

As DUX4 is a transcription factor, it possesses NLS signals⁸⁷, enabling DUX4 protein to enter surrounding nuclei, activating and continuing gene expression changes that are ultimately cytotoxic. DUX4 diffusion into surrounding nuclei is also made evident due to the presence of DUX4 protein in more myonuclei than the DUX4 transcript. Detection of the DUX4 protein ranged between 0.5 to 16.5% of counted myonuclei in primary FSHD cells⁸⁸, whereas the transcript is found between 0.1 to 0.5%^{37,83,84}. Other transcription factors whose expression is regulated by DUX4 can in turn also diffuse along the length of the myofiber's, entering surrounding nuclei and continuing the cytotoxicity (Fig. 3). Therefore, due to the multinucleated nature of muscle tissue, the toxic effects caused by aberrant DUX4 expression are amplified. This, in combination with the low turnover of skeletal muscle cells⁸⁹, makes the muscle tissue more prone to manifest visible symptoms.

The cellular structure of muscle is not the only factor that makes muscle more prone for the development of FSHD symptoms. Two muscle-specific enhancers have been identified, located upstream of the D4Z4 repeats, that are able to control expression of genes in their surroundings, including DUX4⁹⁰. These enhancers possess binding motifs for (myogenic) transcription factors, but also binding motifs for CTCF proteins. CTCF can also bind to the contracted D4Z4 array⁶⁸, which could subsequently facilitate the looping of the enhancers to the DUX4 promoter and lead to gene activation. This looping is less likely to occur with a normal-sized D4Z4 array (11-100), as the chromatin is more compacted, containing more repressive motifs, thereby preventing binding of CTCF to the D4Z4 array. In contrast, a contracted D4Z4 array lowers the competition between the DUX4 promoters contained in the D4Z4 units, to bind to the enhancers. This increases the odds of the enhancers associating with the most proximal D4Z4 repeat unit, which in a 4qA genetic background is connected to a polyA sequence (Fig. 4).

Development and Severity

The severity of FSHD1 is inversely correlated to the length of the D4Z4 repeat array on a permissive haplotype^{7,91,92}. In mosaic FSHD1 patients the severity also depends on the timing of de novo rearrangement during embryonic development, which determines the number and types of tissues that contain affected cells, as well as the proportion of affected cells. As de novo mitotic D4Z4 rearrangement is a common reoccurrence and leads to mosaicism, gametes of a mosaic FSHD1 carrier/patient can be made up of cells containing a contracted 4qter, and cells with normal-sized 4qters. This frequently leads to offspring with FSHD1, that are more severely affected than the parent, since they carry the mutation in all their cells⁹⁻¹⁵.

In FSHD2 patients, the type of mutation in the disease-causing modifier genes can influence the disease severity. FSHD2 patients can be affected more or less, due to the impact of the mutations on the activity of the modifiers, like SMCHD1 and DNMT3B^{57,62,65,93}.

It should be noted that FSHD1 and FSHD2 are not mutually exclusive. There are patients possessing both defects (FSHD1+2), which often exacerbates the disease development, progression and overall severity. FSHD is therefore considered a disease continuum rather than a disease with specific subclasses, as many factors (known and unknown) influence the development and progression of FSHD when occurring in a permissive genetic background^{57,58,62,65,93,94}.

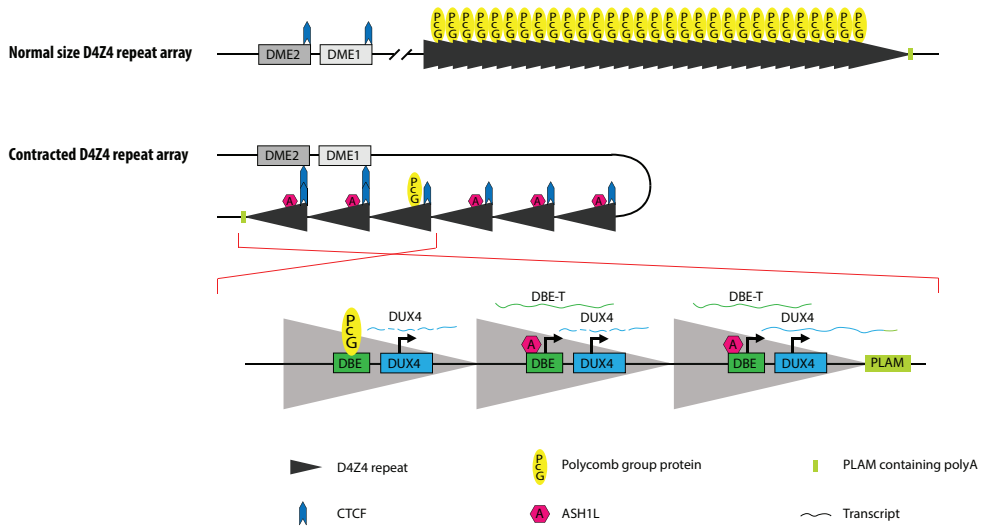


Figure 4. Model of derepression of the D4Z4 array and subsequent expression of DUX4 and DBE-T in muscle cells. Top panel: a normal sized D4Z4 array, carrying repressive Polycomb group (PcG) marks, in a condensed state. Depicted upstream of the D4Z4 array lie DUX4 myogenic enhancer 1 and 2 (DME1 and DME2), bound by CTCF. Lower panel: a contracted D4Z4 array, in a more relaxed state. Relaxation of this area creates openings for CTCF to bind to D4Z4 units. D4Z4-CTCF bind to the CTCFs bound to DME1 and DME2 through looping. Relaxation of the D4Z4 array leads to the expression of DBE-T. DBE-T recruits ASH1L. ASH1L counteracts PcG repression, and promotes further relaxation and expression of surrounding genes. This allows transcription of DUX4, with stable DUX4 transcript being expressed from the most distal D4Z4 unit, attached to the polyA containing PLAM.

In addition to the genetic factors described above (D4Z4 array length, presence of a poly-A signal in exon 3, and the status of SMCHD1 and DNMT3B), several other factors influence the onset and severity of the FSHD phenotype:

- *Number of permissive alleles*

The number of permissive alleles also influences certain aspects of FSHD, including the age of onset, disease progression and severity. In FSHD1, if both 4q35 alleles are contracted and permissive (4qA), DUX4 expression can occur on both alleles, increasing both the likelihood of DUX4 burst expression and potentially the level of DUX4 expression in myotubes. With FSHD2 or FSHD1+2, all D4Z4 repeat arrays will be hypomethylated, thus enabling transcription on both permissive alleles^{93,95}.

- *Telomere length*

Another factor that can play an important role in the development of FSHD is the telomeric length on chromosome 4. Telomere length can influence the expression of genes relatively close to the telomere (up to ~10mb upstream of the telomere) through telomere looping^{96,97}. As the telomeres shorten, telomere looping is diminished, and previous areas of the chromosome that had been in close proximity to the heterochromatin signature of the telomere ends, have lost that repressive connection and are therefore more prone for transcription⁹⁷. Shorter telomeres have shown to be inversely proportional with DUX4 expression, as myoblasts with shorter telomeres have higher DUX4 expression. The

1 involvement of the telomeres in the development of FSHD can explain the late age of onset seen in many patients, as significant shortening of the telomeres would be required to contribute to the derepression of the D4Z4 array⁹⁸.

- *Non-coding RNAs*

The D4Z4 repeated array contains not only the DUX4 gene, but many other transcriptional start sites, in both the sense and anti-sense direction, suggesting it can give rise to other transcripts such as long non-coding RNAs (lncRNA), or small non-coding RNAs such as small interfering RNA (siRNA), micro RNA (miRNA), or piwi-interacting RNA (piRNA)^{40,99,100}. Indeed, many other transcripts have been identified in FSHD-affected cells, that map back to D4Z4 region^{40,99,100}. These ncRNAs are hypothesized to influence DUX4 expression. Some anti-sense RNA fragments are thought to silence the D4Z4 array⁹⁹, whereas a specific lncRNA in the sense orientation has been shown to further induce DUX4 expression. This lncRNA lies upstream of the DUX4 ORF within each D4Z4 unit, and is known to be a D4Z4-binding element (DBE) for the Polycomb group (PcG) proteins^{69,70}. A transcript of DBE (DBE-T) has been discovered in FSHD-affected muscle, which aids the de-repression of the D4Z4 array further¹⁰⁰. As DBE is normally bound by PcG proteins, a contraction of the D4Z4 array will lead to the loss of repressive binding motifs, diminishing PcG occupancy, resulting in chromatin relaxation^{69,70}. This relaxation can be enough for the expression of the DBE-T, which in turn can recruit the Trithorax group protein ASH1L to the D4Z4 array (Fig. 4). ASH1L is a transcriptional activator that can counteracts the PcG repression¹⁰¹⁻¹⁰⁴, therefore further derepressing the D4Z4 array. Furthermore, as ASH1L de-represses the locus further, it promotes the expression of DBE-T, continuing the de-repression of the D4Z4 array as positive feedback loop¹⁰⁰.

With FSHD2, the genetic defect in a chromatin-modifier gene leads to the loss of CpG methylation at the D4Z4 array, resulting in a more permissive chromatin. It would therefore stand to reason that this too could be enough to facilitate the transcription of DBE-T, which again would cause further depression of the D4Z4 region, subsequently leading to the expression of DUX4.

Animal models

Animal models are widely used to study human diseases in a more physiological context. Since DUX4 is more primate-specific, finding a suitable model is challenging. In addition, modeling FSHD in other species is challenged by the wide clinical variability, the high potency of DUX4 cytotoxicity when overexpressed, and its stochastic expression in FSHD-affected tissue. Although DUX4 is not conserved in most of the conventional animal models, many of the downstream genes and pathways are. Several animal models have thus been generated to study the effect of FSHD candidate genes, primarily based on (induced) ectopic DUX4 expression. FSHD models have been created in *Xenopus*¹⁰⁵, zebrafish¹⁰⁶⁻¹⁰⁸, and even *Drosophila*¹⁰⁹. Some of these models recapitulate aspects of FSHD^{107,108}, however, the general effect of expressing DUX4 in these models is embryonic lethality, caused by major cellular loss that is not muscle-specific, and much more severe than what is seen in humans. As there is no DUX4 ortholog found in any of these species, it is not surprising these models respond differently to ectopic DUX4 expression. However, zebrafish models did point out an interesting possibility of a potential developmental role of DUX4 in causing FSHD in later life^{104,105}. Conditional expression of human DUX4 in developing zebrafish embryos resulted

in an asymmetric degenerative effect in the adult zebrafish, after they initially appeared to function normally. If there is indeed a developmental origin of FSHD, this too can be taken into account when considering treatment options or developing therapeutic interventions. Mice do have a Dux gene that shares some sequence and functional homology with its human DUX4 ortholog, allowing them to bind each other's binding motifs and activate a number of the same set of genes¹¹⁰. The Dux gene in mice likely evolved independently from the same ancestor retrogene as DUX4, known as DUXC, a gene that is now lost in both species^{111–114}. Although DUX4 and Dux are similar, a Dux mouse model will likely have problems determining the underlying mechanism, as not all genes and pathways are conserved between species. Additionally, due to the differences between human DUX4 and mouse Dux, the model is a less ideal candidate for preclinical testing of therapeutic applications that target DUX4 transcript or protein. Yet multiple different mouse models have been generated. Mouse models which ectopically express DUX4 show a high variability of disease manifestation^{106,115–118}. Some mouse models were not, or very mildly affected, displaying only an eye phenotype¹¹⁵. Others display a more evident phenotype, that is not limited to skeletal muscle and is rather severe^{106,116}, much more so than in human FSHD patients. Recently two DUX4-inducible mouse models have been generated that recapitulate mild, moderate and severe forms of FSHD through the titration of an inducing agent (doxycycline or tamoxifen). These models show more similarities in disease manifestation and even recapitulate some of the underlying molecular mechanisms, such as DUX4 expression in sporadic nuclei similar to the burst expression seen in FSHD patients, and in the differentially expressed downstream genes and pathways^{117–120}. The DUX4 animal models mentioned here can be helpful for research and drug-screening purposes. However, ectopic expression of the human-specific DUX4 gene in animal model systems can lead to unspecific effects. Moreover, as with the Dux mouse model, these model systems are not suitable to study mechanisms activated by DUX4 expression, as they lack the appropriate human genetic context. While many genes and pathways that are linked to DUX4 expression are conserved in these model systems, this does not necessarily apply for their regulatory regions. The DUX4 binding motif found in the regulatory regions of human genes could be missing in animal models. Indeed, FSHD-related gene 1 (FRG1) is a downstream target gene of DUX4 in human cells¹²¹ and the gene itself is conserved in mice, but is not activated after ectopic human DUX4 expression in murine cells¹¹⁵. In humans, FRG1 contains a functional intronic DUX4 binding site, but this DUX4 binding motif is absent in the Frg1 murine counterpart¹²¹. It is therefore difficult to predict if a drug that is successful in an animal model, would act the same in humans. Moreover, we do not possess a complete understanding of the underlying mechanisms of FSHD, and could therefore be missing important players, that might again not be recapitulated in other species. These models could however be useful when developing treatments that target DUX4 transcript or protein directly.

A relatively novel approach to allow the study of human-specific disorders, is the use of xenograft models. Immunodeficient mice are engrafted with human primary tissue or human cells, to produce human tissue in a physiological animal system. These models have already been successful in cancer research, where antitumor treatments were identified, that then moved into clinical trials where similar effects were observed^{122–124}. These types of models have also been generated for the purpose of studying FSHD and its underlying molecular mechanisms, by engrafting or growing muscle tissue derived from FSHD patients in the animal model. The human muscle tissue in these mice was shown to be vascularized, innervated and functional^{125–127}. Naturally, these models too have their limitations: such as

the variability of engraftment in mice, the fact that a whole-body assessment is not possible, and the need for the use of immunodeficient animals to avoid human tissue rejection, which can potentially effect disease manifestation. These models will however enable researchers to test a broad range of therapeutics that could affect pathways acting upstream or downstream of DUX4 activation, and therefore hold great promise and value in finding a working therapeutic treatment.

Final remarks

Since both the DUX4 transcript and protein are notoriously difficult to detect, and DUX4 derepression in FSHD patients is caused by many underlying factors, the main cause of FSHD remained elusive for more than 100 years. This chapter has given an overview of the work and discoveries that have led to the unmasking of the main, but not sole, culprit of FSHD. It demonstrates that this muscular dystrophy is not as simple as one mutation in one gene, but requires a combination of genetic and epigenetic factors or events for the disease to manifest. Factors such as:

- The length of the D4Z4 repeat array
- A genetic defect in a chromatin-modifier gene or other FSHD-related genes (both known and unknown)
- The type of mutation in FSHD-related genes involved in FSHD pathogenesis
- Telomere length
- Heterozygosity or homozygosity for the 4qA-permissive alleles
- The ‘degree’ of mosaicism if FSHD is not familial and originated *de novo*
- Genetic variations in FSHD-linked enhancers
- The expression and abundance of ncRNAs such as small ncRNA and long ncRNA

All these variations in the population can influence FSHD penetrance, age of onset, disease progression and overall severity. It is therefore not surprising that there is such a large variability between FSHD patients, or even between closely related family members. Patients can range from asymptomatic carriers to being wheelchair-dependent and even requiring ventilation^{13,128}.

There are still many unknowns regarding the molecular mechanisms of the disease, including which transcription factors, co-factors and or kinases are involved in the expression and activation of DUX4. Some transcription factors are suspected to be involved, due to the presence of binding motifs in the myogenic enhancers identified by Himeda et al⁹⁰. These enhancers contain E-box motifs that can be bound by basic helix-loop-helix (bHLH) factors such as MyoD and Myogenin, and homeodomain motifs that can bind homeodomain-containing genes such as the PAX family of transcription factors⁹⁰. The presence of these binding motifs in enhancer and promotor regions is not sufficient evidence of their involvement in the expression of DUX4, but does make them justified suspects. One study identified Bromodomain-containing 4 (BRD4), a member of the bromodomain and extra terminal domain (BET) family of proteins, to be involved in the activation of DUX4, and demonstrated that BET inhibitors decreased DUX4 expression in FSHD patient-derived myoblasts¹²⁹. BET inhibitors are therefore interesting candidates for future clinical trials.

Scope and outline of this thesis

Many of the underlying mechanism of FSHD remain unclear, which hampers the development of effective methods for therapeutic intervention. Work will therefore continue to either clarify these unknown areas of the molecular mechanism, or to modulate DUX4 directly. The work described in this thesis was done with this goal in mind. We generated a versatile human *in vitro* model and applied this cell model to analyze the sequential occurrence of events following expression of DUX4 through RNA sequencing (**Chapter 2 and 3**). In attempts to find novel key players that mediate the cytotoxic effects downstream of DUX4; we used our *in vitro* model to perform a (genome-wide) CRISPR/Cas9 knockout screen (**Chapter 2 and 4**); and as the cell model also contains the genomic sequence of the first three exons of the DUX4 gene (including the pathological polyA sequence in exon 3), it was used to directly target the DUX4 transgene with genome-editing tools (**Chapter 5**) in order to find new and safe ways of knocking-out DUX4. The last chapter integrates the findings of this thesis with current and potential future perspectives of the field (**Chapter 6**).

References

1. Deenen, J. C. W. *et al.* Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* **83**, 1056–1059 (2014).
2. Landouzy, L. & Dejerine, J. De la myopathie atrophique progressive (Myopathie héréditaire débutant, dans l'enfance, par la face, sans altération du système nerveux). *C R Acad Sci* **98**, 53–55 (1884).
3. Tyler, F. H. & Stephens, F. E. Studies in disorders of muscle. II Clinical manifestations and inheritance of facioscapulohumeral dystrophy in a large family. *Ann. Intern. Med.* **32**, 640–660 (1950).
4. Wijmenga, C. *et al.* Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* **2**, 26–30 (1992).
5. Padberg, G. W. *et al.* Facioscapulohumeral muscular dystrophy in the dutch population. *Muscle Nerve* **18**, S81–S84 (1995).
6. Zatz, M. *et al.* High proportion of new mutations and possible anticipation in Brazilian facioscapulohumeral muscular dystrophy families. *Am. J. Hum. Genet.* **56**, 99–105 (1995).
7. Tawil, R. *et al.* Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. *Ann. Neurol.* **39**, 744–748 (1996).
8. Lunt, P. W. 44th ENMC International Workshop: Facioscapulohumeral muscular dystrophy: Molecular studies, 19–21 July 1996, Naarden, The Netherlands. *Neuromuscul. Disord.* **8**, 126–130 (1998).
9. Upadhyaya, M. *et al.* Germinal mosaicism in facioscapulohumeral muscular dystrophy (FSHD). *Muscle Nerve* **18**, S45–S49 (1995).
10. Bakker, E. *et al.* Diagnostic, predictive, and prenatal testing for facioscapulohumeral muscular dystrophy: Diagnostic approach for sporadic and familial cases. *J. Med. Genet.* **33**, 29–35 (1996).
11. Köhler, J., Rupilius, B., Otto, M., Bathke, K. & Koch, M. C. Germline mosaicism in 4q35 facioscapulohumeral muscular dystrophy (FSHD1A) occurring predominantly in oogenesis. *Hum. Genet.* **98**, 485–490 (1996).
12. Zatz, M. *et al.* The facioscapulohumeral muscular dystrophy (FSHD1) gene affects males more severely and more frequently than females. *Am. J. Med. Genet.* **77**, 155–161 (1998).
13. Galluzzi, G. *et al.* Molecular analysis of 4q35 rearrangements in facioscapulohumeral muscular dystrophy (FSHD): Application to family studies for a correct genetic advice and a reliable prenatal diagnosis of the disease. *Neuromuscul. Disord.* **9**, 190–198 (1999).
14. Van Der Maarel, S. M. *et al.* De novo facioscapulohumeral muscular dystrophy: Frequent somatic mosaicism, sex-dependent phenotype, and the role of mitotic transchromosomal repeat interaction between chromosomes 4 and 10. *Am. J. Hum. Genet.* **66**, 26–35 (2000).
15. Lemmers, R. J. L. F. *et al.* Somatic mosaicism in FSHD often goes undetected. *Ann. Neurol.* **55**, 845–850 (2004).
16. Tawil, R. & Van Der Maarel, S. M. Facioscapulohumeral muscular dystrophy. *Muscle and Nerve* **34**, 1–15 (2006).
17. van Dijk, G. P. *et al.* High prevalence of incomplete right bundle branch block in facioscapulohumeral muscular dystrophy without cardiac symptoms. *Funct. Neurol.* **29**, 159–165 (2014).
18. Wijmenga, C. *et al.* Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet* **336**, 651–653 (1990).
19. Upadhyaya, M. *et al.* DNA marker applicable to presymptomatic and prenatal diagnosis of facioscapulohumeral disease. *Lancet* **336**, 1320–1321 (1990).
20. Wijmenga, C. *et al.* Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. *Genomics* **9**, 570–575 (1991).
21. Sarfarazi, M. *et al.* Regional mapping of facioscapulohumeral muscular dystrophy gene on 4q35: Combined analysis of an international consortium. *Am. J. Hum. Genet.* **51**, 396–403 (1992).
22. Deutekom, J. C. T. va. *et al.* FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* **2**, 2037–2042 (1993).
23. Hewitt, J. E. *et al.* Analysis of the tandem repeat locus D4Z4 associated with

- facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* **3**, 1287–1295 (1994).
24. Winokur, S. T. *et al.* The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: Implications for a role of chromatin structure in the pathogenesis of the disease. *Chromosom. Res.* **2**, 225–234 (1994).
 25. Deidda, G. *et al.* Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *European Journal of Human Genetics* vol. 3 155–167 (1995).
 26. Cacurri, S. *et al.* Sequence homology between 4qter and 10qter loci facilitates the instability of subtelomeric KpnI repeat units implicated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **63**, 181–190 (1998).
 27. Bakker, E. *et al.* The FSHD-linked locus D4F104S1 (p13E-11) ON 4q35 has a homologue on 10qter. *Muscle Nerve* **18**, S39–S44 (1995).
 28. Lemmers, R. *et al.* Inter- and intrachromosomal sub-telomeric rearrangements on 4q35: implications for facioscapulohumeral muscular dystrophy (FSHD) aetiology and diagnosis. *Hum. Mol. Genet.* **7**, 1207–1214 (1998).
 29. Deidda, G., Cacurri, S., Piazza, N. & Felicetti, L. Direct detection of 4q35 rearrangements implicated in facioscapulohumeral muscular dystrophy (FSHD). *J. Med. Genet.* **33**, 361–365 (1996).
 30. Gabriëls, J. *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* **236**, 25–32 (1999).
 31. Tupler, R. *et al.* Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. *J. Med. Genet.* **33**, 366–370 (1996).
 32. Choi, S. H. *et al.* DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **44**, 5161–5173 (2016).
 33. Vuoristo, S. *et al.* DUX4 regulates oocyte to embryo transition in human. *Biorxiv* (2019) doi:<http://dx.doi.org/10.1101/732289>.
 34. Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **49**, 935–940 (2017).
 35. Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
 36. Das, S. & Chadwick, B. P. Influence of repressive histone and DNA methylation upon D4Z4 transcription in non-myogenic cells. *PLoS One* **11**, 1–26 (2016).
 37. Snider, L. *et al.* Facioscapulohumeral dystrophy: Incomplete suppression of a retrotransposed gene. *PLoS Genet.* **6**, 1–14 (2010).
 38. Kowaljow, V. *et al.* The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul. Disord.* **17**, 611–623 (2007).
 39. Dixit, M. *et al.* DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18157–18162 (2007).
 40. Snider, L. *et al.* RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: New candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum. Mol. Genet.* **18**, 2414–2430 (2009).
 41. Van Geel, M. *et al.* Genomic analysis of human chromosome 10q and 4q telomeres suggests a common origin. *Genomics* **79**, 210–217 (2002).
 42. Lemmers, R. J. L. F. *et al.* Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat. Genet.* **32**, 235–236 (2002).
 43. Lemmers, R. J. F. L. *et al.* Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **75**, 1124–1130 (2004).
 44. Thomas, N. S. T. *et al.* A large patient study confirming that facioscapulohumeral muscular dystrophy (FSHD) disease expression is almost exclusively associated with an FSHD locus located on a 4qA-defined 4qter subtelomere. *J. Med. Genet.* **44**, 215–218 (2007).
 45. Lemmers, R. J. L. F. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science (80-.)*. **329**, 1650–1653 (2010).
 46. Lemmers, R. J. L. F. *et al.* Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **81**, 884–894 (2007).
 47. Lemmers, R. J. L. F. *et al.* Worldwide Population Analysis of the 4q and 10q Subtelomeres Identifies Only Four Discrete Interchromosomal Sequence Transfers in Human Evolution.

- 1
- Am. J. Hum. Genet.* **86**, 364–377 (2010).
48. van Deutekom, J. Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1. *Hum. Mol. Genet.* **5**, 1997–2003 (1996).
 49. Van Overveld, P. G. M. *et al.* Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* **35**, 315–317 (2003).
 50. De Greef, J. C. *et al.* Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. *Neurology* **69**, 1018–1026 (2007).
 51. de Greef, J. C. *et al.* Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. *Hum. Mutat.* **30**, 1449–1459 (2009).
 52. de Greef, J. *et al.* Clinical features of facioscapulohumeral muscular dystrophy 2. *Neurology* **75**, 1548–1554 (2010).
 53. Lemmers, R. J. L. F. *et al.* Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44**, 1370–1374 (2012).
 54. Blewitt, M. E. *et al.* An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7629–7634 (2005).
 55. Blewitt, M. E. *et al.* SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* **40**, 663–669 (2008).
 56. Gendrel, A. V. *et al.* Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome. *Dev. Cell* **23**, 265–279 (2012).
 57. Larsen, M. *et al.* Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. *Eur. J. Hum. Genet.* **23**, 808–816 (2015).
 58. Van Den Boogaard, M. L. *et al.* Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. *Am. J. Hum. Genet.* **98**, 1020–1029 (2016).
 59. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases Non-invasive sexing of preimplantation stage mammalian embryos. *Nat. Am. Inc.* **19**, 219–220 (1998).
 60. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
 61. Auclair, G., Guibert, S., Bender, A. & Weber, M. Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. *Genome Biol.* **15**, 545 (2014).
 62. Lemmers, R. J. L. F. *et al.* Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2. *Hum. Mol. Genet.* **24**, 659–669 (2015).
 63. Lemmers, R. J. L. F. *et al.* Hemizyosity for SMCHD1 in Facioscapulohumeral Muscular Dystrophy Type 2: Consequences for 18p Deletion Syndrome. *Hum. Mutat.* **36**, 679–683 (2015).
 64. Balog, J. *et al.* Monosomy 18p is a risk factor for facioscapulohumeral dystrophy. *J. Med. Genet.* **55**, 469–478 (2018).
 65. Sacconi, S. *et al.* FSHD1 and FSHD2 form a disease continuum. *Neurology* **92**, E2273–E2285 (2019).
 66. Zeng, W. *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.* **5**, (2009).
 67. Jiang, G. *et al.* Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum. Mol. Genet.* **12**, 2909–2921 (2003).
 68. Ottaviani, A. *et al.* The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in Facio-Scapulo-Humeral dystrophy. *PLoS Genet.* **5**, (2009).
 69. Gabellini, D., Green, M. R. & Tupler, R. Inappropriate gene activation in FSHD: A repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* **110**, 339–348 (2002).
 70. Bodega, B. *et al.* Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1)

- expression during human myogenic differentiation. *BMC Biol.* **7**, 41 (2009).
71. Golbabapour, S. *et al.* Gene silencing and polycomb group proteins: An overview of their structure, mechanisms and phylogenetics. *Omi. A J. Integr. Biol.* **17**, 283–296 (2013).
 72. Lemmers, R. J. L. F. *et al.* Mechanism and timing of mitotic rearrangements in the subtelomeric D4Z4 repeat involved in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **75**, 44–53 (2004).
 73. Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J.* **19**, 3398–3407 (2000).
 74. Overveld, P. G. M. v. Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity. *Hum. Mol. Genet.* **9**, 2879–2884 (2000).
 75. Padberg, G. W. *et al.* On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. *Muscle Nerve* **18**, S73–S80 (1995).
 76. Funakoshi, M., Goto, K. & Arahata, K. Epilepsy and mental retardation in a subset of early onset 4q35- facioscapulohumeral muscular dystrophy. *Neurology* **50**, 1791–1794 (1998).
 77. Trevisan, C. P. *et al.* Facioscapulohumeral muscular dystrophy: Hearing loss and other atypical features of patients with large 4q35 deletions. *Eur. J. Neurol.* **15**, 1353–1358 (2008).
 78. Goselink, R. J. M. *et al.* Ophthalmological findings in facioscapulohumeral dystrophy. *Brain Commun.* **1**, 1–9 (2019).
 79. Heron, M. I. & Richmond, F. J. R. In-series fiber architecture in long human muscles. *J. Morphol.* **216**, 35–45 (1993).
 80. Kawano, F. *et al.* Essential role of satellite cells in the growth of rat soleus muscle fibers. *Am. J. Physiol. - Cell Physiol.* **295**, 458–467 (2008).
 81. van der Meer, S. F. T., Jaspers, R. T. & Degens, H. Is the myonuclear domain size fixed? *J. Musculoskelet. Neuronal Interact.* **11**, 286–297 (2011).
 82. Kramer, I. F. *et al.* Extensive Type II Muscle Fiber Atrophy in Elderly Female Hip Fracture Patients. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **72**, 1369–1375 (2017).
 83. Jones, T. I. *et al.* Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: Evidence for disease modifiers and a quantitative model of pathogenesis. *Hum. Mol. Genet.* **21**, 4419–4430 (2012).
 84. Tassin, A. *et al.* DUX4 expression in FSHD muscle cells: How could such a rare protein cause a myopathy? *J. Cell. Mol. Med.* **17**, 76–89 (2013).
 85. Ferreboeuf, M. *et al.* Nuclear protein spreading: Implication for pathophysiology of neuromuscular diseases. *Hum. Mol. Genet.* **23**, 4125–4133 (2014).
 86. Rickard, A. M., Petek, L. M. & Miller, D. G. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum. Mol. Genet.* **24**, 5901–5914 (2015).
 87. Corona, E. D., Jacquelin, D., Gatica, L. & Rosa, A. L. Multiple Protein Domains Contribute to Nuclear Import and Cell Toxicity of DUX4, a Candidate Pathogenic Protein for Facioscapulohumeral Muscular Dystrophy. *PLoS One* **8**, 1–11 (2013).
 88. Block, G. J. *et al.* Wnt/ β -catenin signaling suppresses DUX4 expression and prevents apoptosis of FSHD muscle cells. *Hum. Mol. Genet.* **22**, 4661–4672 (2013).
 89. Claydon, A. J., Thom, M. D., Hurst, J. L. & Beynon, R. J. Protein turnover: Measurement of proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids. *Proteomics* **12**, 1194–1206 (2012).
 90. Himeda, C. L. *et al.* Myogenic Enhancers Regulate Expression of the Facioscapulohumeral Muscular Dystrophy-Associated DUX4 Gene. *Mol. Cell. Biol.* **34**, 1942–1955 (2014).
 91. Lunt, P. W. *et al.* Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). *Hum. Mol. Genet.* **4**, 951–958 (1995).
 92. Lunt, P. W. *et al.* Phenotypic–genotypic correlation will assist genetic counseling in 4q35-facioscapulohumeral muscular dystrophy. *Muscle Nerve* **18**, S103–S109 (1995).
 93. Lemmers, R. J. *et al.* Deep characterization of a common D4Z4 variant identifies biallelic DUX4 expression as a modifier for disease penetrance in FSHD2. *Eur. J. Hum. Genet.* **26**, 94–106 (2018).
 94. Sacconi, S. *et al.* The FSHD2 gene SMCHD1 Is a modifier of disease severity in families affected by FSHD1. *Am. J. Hum. Genet.* **93**, 744–751 (2013).

- 1
95. Renard, D. *et al.* Inflammatory facioscapulohumeral muscular dystrophy type 2 in 18p deletion syndrome. *Am. J. Med. Genet. Part A* **176**, 1760–1763 (2018).
 96. Wright, W. E. & Shay, J. W. Telomere positional effects and the regulation of cellular senescence. *Trends Genet.* **8**, 193–197 (1992).
 97. Robin, J. D. *et al.* Telomere position effect: Regulation of gene expression with progressive telomere shortening over long distances. *Genes Dev.* **28**, 2464–2476 (2014).
 98. Stadler, G. *et al.* Telomere position effect regulates DUX4 in human facioscapulohumeral muscular dystrophy. *Nat. Struct. Mol. Biol.* **20**, 671–678 (2013).
 99. Block, G. J. *et al.* Asymmetric bidirectional transcription from the FSHD-causing D4Z4 array modulates DUX4 production. *PLoS One* **7**, e35532 (2012).
 100. Cabianca, D. S. *et al.* A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in fshd muscular dystrophy. *Cell* **149**, 819–831 (2012).
 101. Klymenko, T. & Jürg, M. The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep.* **5**, 373–377 (2004).
 102. Schmitges, F. W. *et al.* Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol. Cell* **42**, 330–341 (2011).
 103. Yuan, W. *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286**, 7983–7989 (2011).
 104. Miyazaki, H. *et al.* Ash1l Methylates Lys36 of Histone H3 Independently of Transcriptional Elongation to Counteract Polycomb Silencing. *PLoS Genet.* **9**, (2013).
 105. Wuebbles, R. D., Long, S. W., Hanel, M. L. & Jones, P. L. Testing the effects of FSHD candidate gene expression in vertebrate muscle development. *Int. J. Clin. Exp. Pathol.* **3**, 386–400 (2010).
 106. Wallace, L. M. *et al.* DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann. Neurol.* **69**, 540–552 (2011).
 107. Mitsuhashi, H., Mitsuhashi, S., Lynn-jones, T., Kawahara, G. & Kunkel, L. M. Expression of DUX4 in zebrafish development recapitulates facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.* **22**, 568–577 (2013).
 108. Pakula, A. *et al.* Transgenic zebrafish model of DUX4 misexpression reveals a developmental role in FSHD pathogenesis. *Hum. Mol. Genet.* **28**, 320–331 (2019).
 109. Jones, T. I., Parrilla, M. & Jones, P. L. Transgenic drosophila for investigating dux4 and frg1, two genes associated with facioscapulohumeral muscular dystrophy (fshd). *PLoS One* **11**, (2016).
 110. Eidahl, J. O. *et al.* Mouse Dux is myotoxic and shares partial functional homology with its human paralog DUX4. *Hum. Mol. Genet.* **25**, ddw287 (2016).
 111. Winokur, S. T., Ulla, B., Vargas, J. C., Wasmuth, J. J. & Altherr, M. R. The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum. Mol. Genet.* **6**, 502–502 (1997).
 112. Clark, L. N., Koehler, U., Ward, D. C., Wienberg, J. & Hewitt, J. E. Analysis of the organisation and localisation of the FSHD-associated tandem array in primates: Implications for the origin and evolution of the 3.3 kb repeat family. *Chromosoma* **105**, 180–189 (1996).
 113. Clapp, J. *et al.* Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **81**, 264–279 (2007).
 114. Leidenroth, A. *et al.* Evolution of DUX gene macrosatellites in placental mammals. *Chromosoma* **121**, 489–497 (2012).
 115. Krom, Y. D. *et al.* Intrinsic Epigenetic Regulation of the D4Z4 Macrosatellite Repeat in a Transgenic Mouse Model for FSHD. *PLoS Genet.* **9**, (2013).
 116. Dandapat, A. *et al.* Dominant Lethal Pathologies in Male Mice Engineered to Contain an X-Linked DUX4 Transgene. *Cell Rep.* **8**, 1484–1496 (2014).
 117. Bosnakovski, D. *et al.* Muscle pathology from stochastic low level DUX4 expression in an FSHD mouse model. *Nat. Commun.* **8**, 1–9 (2017).
 118. Jones, T. & Jones, P. L. A cre-inducible DUX4 transgenic mouse model for investigating facioscapulohumeral muscular dystrophy. *PLoS ONE* vol. 13 (2018).
 119. Bosnakovski, D. *et al.* Transcriptional and cytopathological hallmarks of FSHD in chronic DUX4-expressing mice. *J. Clin. Invest.* **130**, 2465–2477 (2020).

120. Jones, T. I. *et al.* Transgenic mice expressing tunable levels of DUX4 develop characteristic facioscapulohumeral muscular dystrophy-like pathophysiology ranging in severity. *Skelet. Muscle* **10**, 1–28 (2020).
121. Ferri, G., Huichalaf, C. H., Caccia, R. & Gabellini, D. Direct interplay between two candidate genes in FSHD muscular dystrophy. *Hum. Mol. Genet.* **24**, 1256–1266 (2015).
122. Adams, J. Development of the Proteasome Inhibitor PS-341. *Oncologist* **7**, 9–16 (2002).
123. Baselga, J. Phase I and II clinical trials of trastuzumab.pdf. *Ann. Oncol.* **12**, S49–S55 (2001).
124. Pegram, M. & Ngo, D. Application and potential limitations of animal models utilized in the development of trastuzumab (Herceptin®): A case study. *Adv. Drug Deliv. Rev.* **58**, 723–734 (2006).
125. Zhang, Y. *et al.* Humanskeletal musclexenograft as anewpreclinical model for muscle disorders. *Hum. Mol. Genet.* **23**, 3180–3188 (2014).
126. Mueller, A. L. *et al.* Muscle xenografts reproduce key molecular features of facioscapulohumeral muscular dystrophy. *Exp. Neurol.* **320**, 113011 (2019).
127. Sakellariou, P. *et al.* Neuromuscular electrical stimulation promotes development in mice of mature human muscle from immortalized human myoblasts. *Skelet. Muscle* **6**, 1–14 (2016).
128. Tawil, R., van der Maarel, S. M. & Tapscott, S. J. Facioscapulohumeral dystrophy: The path to consensus on pathophysiology. *Skelet. Muscle* **4**, 1–15 (2014).
129. Campbell, A. E. *et al.* BET bromodomain inhibitors and agonists of the beta-2 adrenergic receptor identified in screens for compounds that inhibit DUX4 expression in FSHD muscle cells. *Skelet. Muscle* **7**, 1–18 (2017).

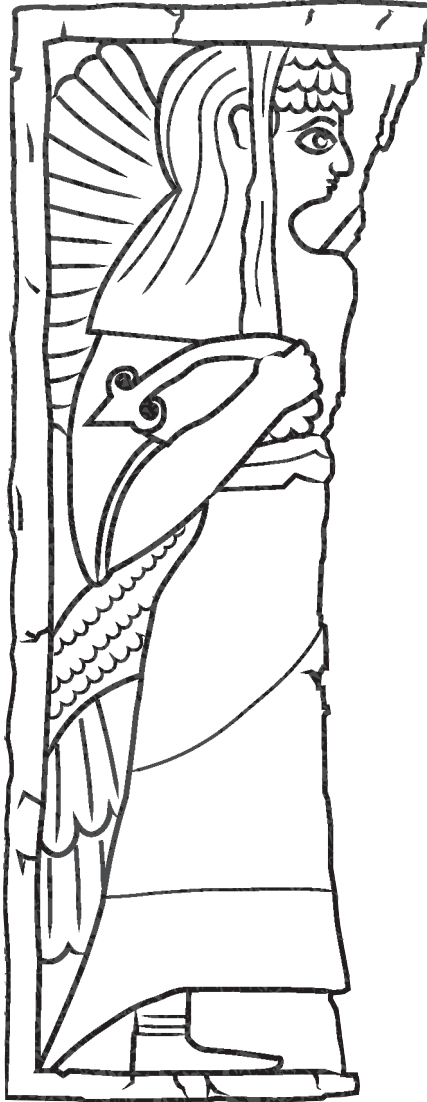


Illustration based on a stone carving on display at the MET

Chapter 2

Generation of a cellular model to dissect early molecular events leading to DUX4-induced toxicity

Ator Ashoti, Mirna Baak, Menno Creyghton, Niels Geijsen

Parts of this chapter are available in an adapted form on bioRxiv

Ashoti, A. et al. A genome-wide CRISPR/Cas phenotypic screen for modulators of DUX4 cytotoxicity reveals screen complications. bioRxiv (2020).

doi:10.1101/2020.07.27.223420

Abstract

Facioscapulohumeral muscular dystrophy (FSHD) is a complex disease that can be caused by several genetic and epigenetic factors. One such factor is the failure to epigenetically silence the sub-telomeric region of chromosome 4, causing misexpression of the Double homeobox 4 (DUX4) gene. Expression of DUX4 is normally tightly regulated and restricted to the thymus, testis and early cleavage stage embryos. Aberrant expression of DUX4 in skeletal muscle underlies the pathogenesis of FSHD. To gain insight into the pathophysiology of FSHD, we aimed to identify the downstream targets leading to DUX4-induced cytotoxicity and assess if manipulation of downstream targets could potentially mitigate DUX4 cytotoxicity. We developed a cell line that, upon inducible DUX4 expression, recapitulates the FSHD transcriptional signature and ultimately undergoes apoptotic cell death. We also developed a small-scale screening assay to knockout DUX4 target genes that were expressed early after DUX4 induction, in order to test their ability to modulate DUX4-induced cytotoxicity. Thus, we have developed a robust system to investigate the molecular and cellular events that follow DUX4 expression and are causal to the emergence of FSHD.

Introduction

Muscular dystrophies are a group of genetic disorders characterized by progressive loss of muscle strength and muscle degeneration. These diseases often have few treatment options, if any, and current therapies primarily focus on symptom relief, not resolving the underlying cause. Treatments consist of a combination of physical therapy and anti-inflammatory agents aimed at preserving muscle function as much as possible. Facioscapulohumeral muscular dystrophy (FHD) is one of the most prevalent forms of muscular dystrophy worldwide¹. FHD first manifests in muscle groups of the face, effecting speech and facial expression. Patients lose the ability to express emotion which hampers their ability to engage in social interactions. FHD progresses sequentially and in an asymmetric fashion from the face, the shoulders, upper arms, trunk and the lower extremities². Due to the consequences of muscle weakening in the face, the disorder can affect the patient's societal interactions and can cause significant emotional stress.

A unified model for the underlying genetics of FHD, published in 2010, demonstrated that a permissive chromosomal background, together with epigenetic de-repression of the D4Z4 locus results in the (mis)expression of the transcription factor double homeobox 4 (DUX4)³ and is the main cause for the development of FHD. DUX4 is a so-called pioneer transcription factor^{4,5}, capable of regulating its target genes independent of their chromatin state. The network of genes activated by pioneer factors is therefore less affected by cellular identity. Indeed, Jones and colleagues have demonstrated that DUX4 activates the same downstream target genes in B lymphocytes as were previously identified in skeletal muscle myoblasts^{6,7}.

To explore the molecular mechanisms that trigger DUX4-mediated cytotoxicity, and to explore potential ways to mitigate this toxicity, we generated a cell line in which DUX4 expression can be induced by the addition of doxycycline. We introduced the DUX4 intron-exon structure (exons 1-3) involved in FHD pathogenesis into the adherent KBM7 cell line⁸⁻¹⁷, under control of a doxycycline-inducible promoter, and identified a clone that displayed robust DUX4-dependent cell death upon addition of doxycycline. Using this cell line, we determined the temporal molecular events that are triggered upon doxycycline-mediated DUX4 induction and demonstrate that the transcriptome changes induced by DUX4 in our inducible system are highly similar to those previously identified in myoblasts from FHD patients¹⁸⁻²⁰.

To test the feasibility of using this system to screen for factors that could mitigate DUX4 cytotoxicity, and at the same time test the role of a small list of genes downstream of the DUX4 transcription factor, we developed a small-scale CRISPR/Cas9 screening assay. Genes that were significantly upregulated after inducing DUX4 expression for 4.5 hours were considered early targets of DUX4. We hypothesize that targeting early-induced DUX4 target genes could interfere with the induction of the DUX4 cytotoxic transcriptional network thereby delaying or even abrogating DUX4-induced cell death. Early transcription and co-factors are of particular interest because of their potential role in perpetuating the toxic cascade of events. Our small-scale CRISPR/Cas9 screen allowed us to test this hypothesis, in a fast and cost-effective manner.

Results

Generation and validation of a DUX4 inducible cell line

To perform large screens, cells should preferably be highly proliferative, be easily transfectable and display a robust and screenable phenotype. An adherent clone of the KBM7 cell-line possess most of these characteristics and has already been used extensively in a wide variety of functional screens⁸⁻¹⁷. The cells were initially near-haploid^{8,21}, but subsequently rediploidized^{22,23}. These adherent diploid KBM7 cells were used for the generation of our FSHD cell model.

Low levels of DUX4 expression can efficiently induce apoptosis^{19,24,25}, which interfered with the generation of our FSHD cell model. To circumvent premature DUX4 toxicity, caused by the leaky expression of the Tet-On system^{26,27}, we inserted a LoxP-DsRed-LoxP-stop-cassette (LSL) in between the Tet-operator and the DUX4 transgene (Fig. 1A). The DUX4 transgene element itself consisted of the first three exons (starting with the translational start site) and the two introns of the DUX4 gene, including the polyA sequence. This is the same sequence found in the most common pathology-associated haplotype, 4A161²⁸. This construct was introduced in the adherent KBM7 cells in combination with a constitutive rtTA expression construct. After stable integration of the construct, the stop-cassette was removed using CRE recombinase, placing DUX4 under the control of the Tet operator (Fig. 1A). Eighty monoclonal lines were derived by single cell flow-cytometry sorting, of which one displayed tight doxycycline-dependent DUX4 induction and robust cell death upon doxycycline addition (Fig. 1B). A monoclonal cell line was derived from this positive clone, which we named the 'DUX4 Inducible Expression' (DIE) cell line.

To further characterize the DIE cells, we determined the sites of integration of the rtTA/BlastR and DUX4/PuroR transgenes. Targeted locus amplification (TLA)²⁹ was performed and confirmed single integration sites for both the rtTA and DUX4 transgenes (Fig. 1C). The DUX4 cassette integrated into the p-arm of chromosome 19 within the *MAST1* gene, and the rtTA cassette integrated into the *MGAT4B* gene, which is located at the end of chromosome 5q. To further analyze the functional effect of DUX4 induction, DIE cells were stained with AnnexinV-Alexa Fluor 488 and propidium iodine (PI) (Fig. S1). As shown in the supplementary videos, DIE cells stained positive for the apoptotic marker AnnexinV during 12 hours of doxycycline exposure. To show that the apoptotic phenotype was dependent on induction of the DUX4 transgene, we knocked out the DUX4 transgene using CRISPR/Cas9. To target the DUX4 transgene, 2 independent guide RNAs (gRNAs) were used, one targeting the C-terminal domain of the DUX4 open reading frame (ORF) and the other close to the polyA tail of DUX4. RT-qPCR and Western blot (WB) analysis of the DIE and the DIE-DUX4 knockout (DIE-KO) cells demonstrated successful knockout of the DUX4 transgene at RNA and protein levels (Fig. 1D). In addition, CRISPR/Cas9 targeting of the DUX4 transgene successfully rescued the DIE cells from apoptosis upon doxycycline administration, demonstrating that apoptosis upon doxycycline induction in the DIE cell line is mediated by DUX4 (Fig. 1E). DUX4 induction in the DIE cells also resulted in induction of its known downstream target genes (LEUTX, ZSCAN4, PRAMFE1 and ZNF217) (Fig. 1F), demonstrating that inducing DUX4 expression induces downstream target genes that are also activated by endogenous DUX4.

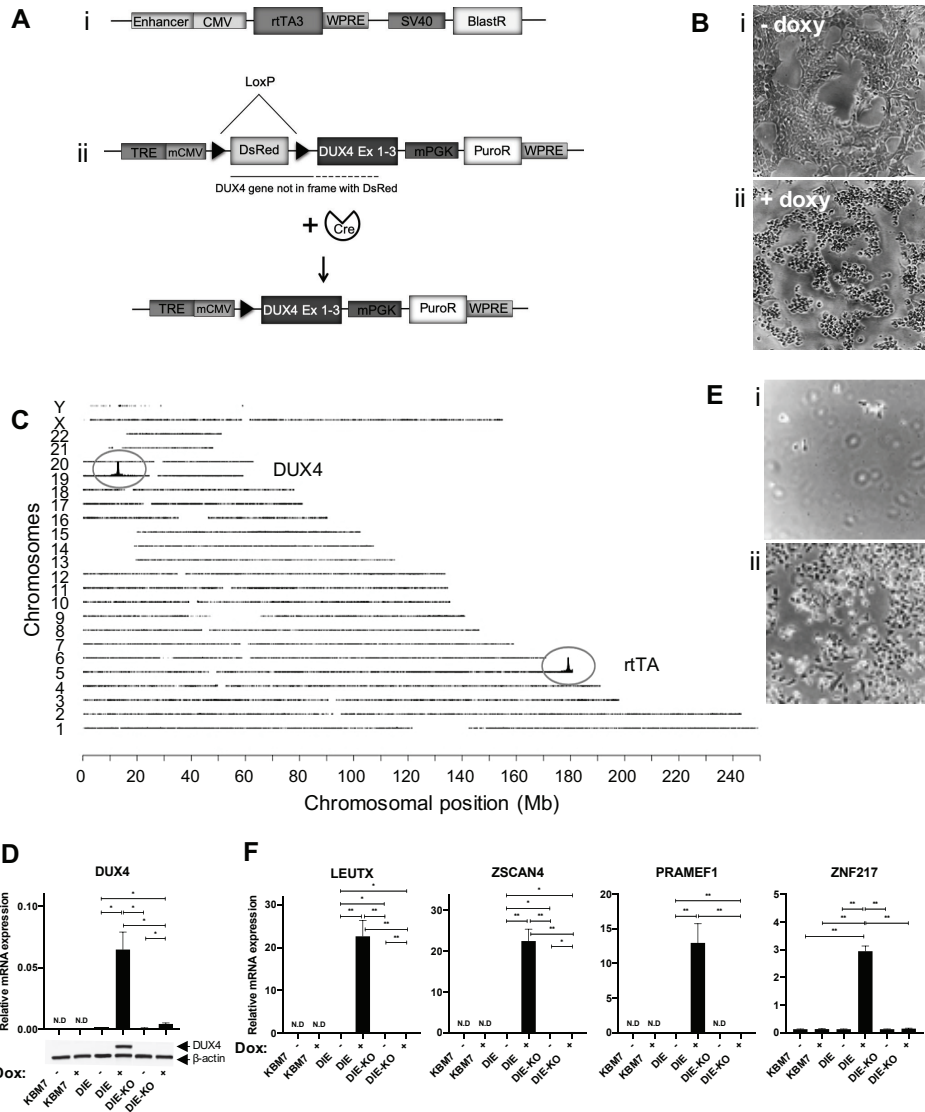


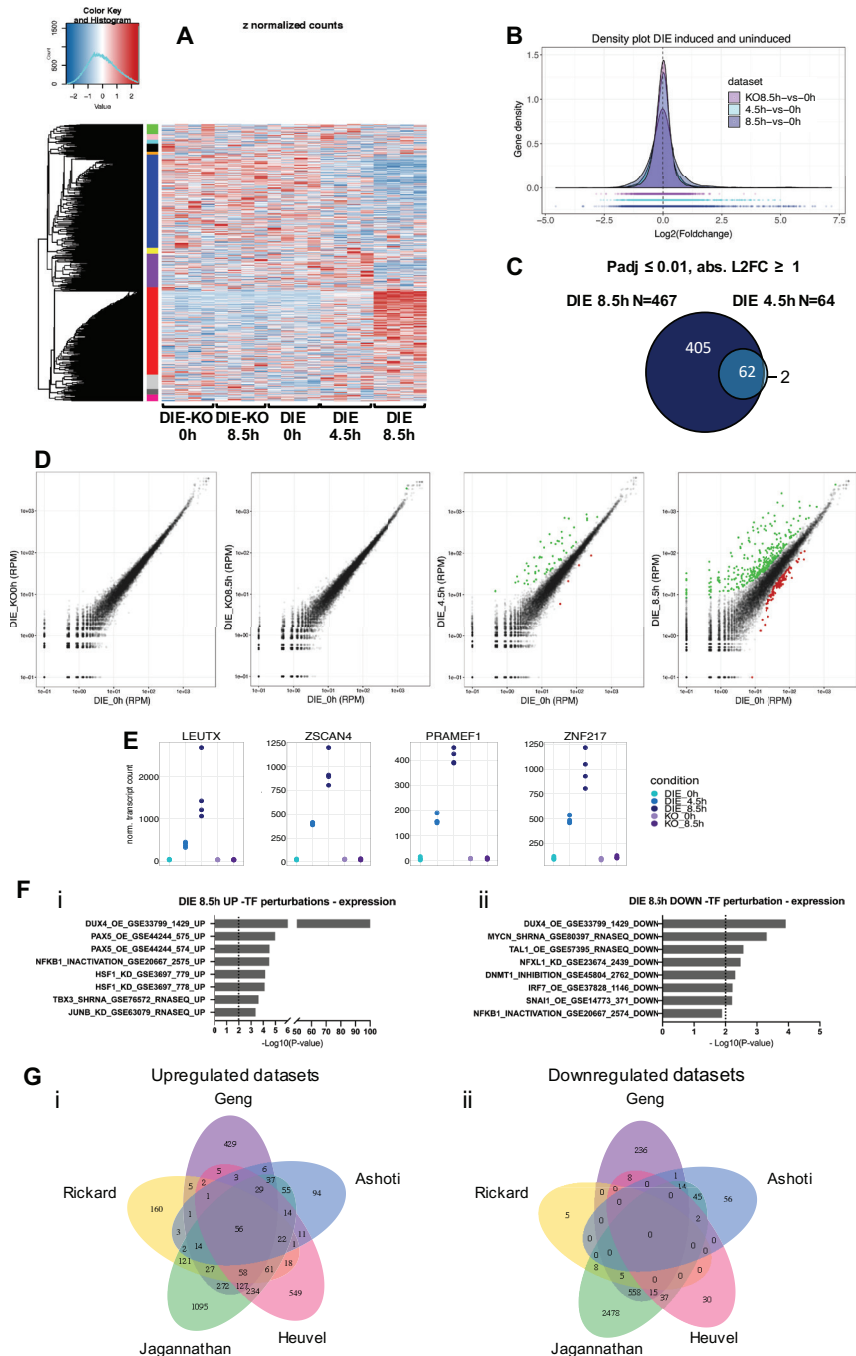
Figure 1. Creation and validation of the DIE cell line. (A) Schematic representation of the i) rTA construct, ii) the inducible LSL-DUX4 cassettes, and the inducible DUX4 cassette upon removal of the LSL through CRE recombinase. (B) Phase contrast images of DIE cells i) without doxycycline exposure and ii) with 24h of doxycycline exposure. (C) Schematic representation of transgene integration sites within human genome, by TLA analysis. The inducible DUX4 cassette maps back to the p-arm of chromosome 19, and the rTA transgene maps back to the end of the q-arm of chromosome 5. (D) Expression of DUX4 mRNA and protein in the parental KBM7 cells, DIE and DIE-KO cells with or without doxycycline admission, as detected by qRT-PCR (top panel) and western blot analysis (bottom panel), with β -actin serving as a loading control in the western blot. (E) Phase contrast images of doxycycline exposed DIE cells which were transduced with either i) only Cas9 protein, or ii) Cas9 protein and DUX4 sgRNAs. Dead cells were removed by a DPBS wash to expose the surviving population. (F) mRNA expression of known downstream targets of DUX4 in KBM7, DIE and DIE-KO cells with or without doxycycline admission, as measured by qRT-PCR. The statistical significance in all qRT-PCR data was determined by a two-tailed Student t-test. LSL: LoxP-DsRed-stop-LoxP, KO: Knock-out, N.D: not detected.

DUX4 gene expression signature in DIE cells

Next we analyzed the downstream transcriptional changes that were induced by DUX4 in the DIE cells by RNA sequencing. We compared 4 induced and uninduced technical replicates of two lines; the DIE line, and the DIE-KO line. As shown in Fig. 2A, DUX4 induction resulted in progressive temporal changes in gene expression. Figure 2B shows the magnitude of the combined transcriptional changes induced by DUX4 at different time intervals and schematically emphasizes both the increasing number of differentially transcribed genes as well as the speed at which these transcriptional changes occur over time. Indeed, DUX4 induction results in a profound and progressive increase in the number of differentially expressed genes; with 64 differentially expressed transcripts at 4.5 hours post DUX4 induction and 467 differentially expressed transcripts at 8.5 hours after induction (Fig. 2B and 2C). The number of induced genes is greater than those with reduced expression levels as can be seen in Fig. 2B. Differential expression analysis reflects this, demonstrating more differentially upregulated genes in both induced DIE samples compared to uninduced DIE sample [Padj value ≤ 0.01 , absolute Log₂foldchange ≥ 1] (Fig. 2B and D, Supplementary Table S1 and S2). Most differentially expressed genes are shared between the two induced samples (Fig. 2C). Among the upregulated genes are well known downstream targets of DUX4, including LEUTX, ZSCAN4, PRAMEF1 and ZNF217 (Fig. 2E). We next used Enrichr^{30,31} to search for other similar studies that show similarities in their transcriptome. Based on Enrichr entries, the upregulated genes in induced DIE cells are linked to DUX4 activity [-Log₁₀(P-value) = 100.3], as are the downregulated genes [-Log₁₀(P-value) = 3.8] (Fig. 2F and Tables S3-4). It shows high similarity between data from our study and one other DUX4 study that has been entered into the Enrichr database (GSE33799)¹⁸. We next compared the list of differentially expressed genes (DIE_8.5h) with 4 other studies that have previously identified the DUX4 transcriptional network in myoblast models or patients derived muscle biopsies (Geng, Rickard, Jagannathan and Heuvel)^{18-20,32} and observed a high percentage of overlap between datasets. 72% of the upregulated genes and 52.8% of the downregulated genes overlap with at least one of 4 datasets (Fig. 2G and Table S5 and S6). In addition, overlapping percentages mentioned here are likely an underrepresentation, due to the presence of gene families containing paralogs and pseudogenes in either reference genome databases, which can lead to multi-mapped or ambiguous reads³³. To conclude, data shown here strongly suggest that in our DIE cell system, DUX4 induces transcriptional changes similar to those found in myoblasts from FSHD patients.

Figure 2. RNA-sequencing data revealing differentially expressed genes upon DUX4 expression. (A) Heatmap showing differentially expressed genes between samples, with gene clusters (color coded) on y-axis, and samples on the x-axis. **(B)** Gene density plot demonstrating the effects of DUX4 activation on the transcriptome of the DIE cell line. DUX4 activation results in an increase of differentially expressed genes compared to uninduced DIE cells, as indicated by the bell shape widening and shortening. **(C)** Venn diagram showing the overlap and the number of differentially expressed genes at 4.5h and 8.5h of doxycycline induction (Adjusted P-value ≤ 0.01 , and absolute Log₂FC ≥ 1). **(D)** Scatter plots of gene expression (RPM: reads per million) of induced DIE cells versus uninduced DIE cells. Left two panels demonstrate uninduced DIE cells (DIE_0h) on the x-axis versus uninduced or induced DIE-KO samples (KO_0h and KO_8.5h) on the y-axis. Right two panels compare the uninduced DIE cells with induced DIE samples (4.5h and 8.5h). Green and red dots represent the differentially expressed genes with an Adjusted P-value ≤ 0.01 , and absolute Log₂FC ≥ 1 . Green dots represent upregulated genes, and the red dots represent downregulated genes. **(E)** Count plots showing unique molecular identifier (UMI) and between sample normalized transcript counts of 4 known DUX4 targets genes: LEUTX, ZSCAN4, PRAMEF1 and ZNF217, in uninduced and induced DIE and DIE-KO cells. Every sample shows 4 dots, representative of the 4 technical replicates. **(F)** Transcription factor (TF) perturbations analysis identifying transcription factors that are linked to the i) upregulation and ii) downregulation of the differentially expressed genes found in this study. Activation: OE or ACTIVATION.

Inhibition: KO, KD, SIRNA, SHRNA, INACTIVATION, or INHIBITION. **(G)** Quintuple Venn diagram comparing DUX4 i) upregulated and ii) downregulated genes found in this study (Ashoti) to those found in previous transcriptomic studies (Geng with P-value ≤ 0.01 , FDR ≤ 0.05 , abs L2FC ≥ 1 ; Rickard with Padj value of < 0.005 and abs L2FC > 2 ; Jagannathan with P-value ≤ 0.01 , FDR ≤ 0.05 , abs L2FC ≥ 1 , Heuvel with P-value ≤ 0.005 , FDR ≤ 0.05 , abs L2FC ≥ 1). See supplementary material Tables 5 and 6.



2

DUX4 induces an early embryonic transcription factor network

We noticed that the list of early DUX4-affected genes contains a relatively high number of transcription- and co-factors (19 out of 64 differentially-expressed genes, Table 1), more than could be expected based on random distribution, since transcription factors and co factors comprise only between 11-13.5% of all protein coding genes in the human genome. DUX4 is a pioneer factor that is normally expressed during early, preimplantation embryonic development^{34,35}. Figure 3A displays the expression of DUX4 in oocytes, zygotes and cleavage-stage embryos as well as later stages of pre-implantation development. As shown, DUX4 peaks at the 4-cell stage, and quickly wanes thereafter. When analyzing the DUX4 target genes identified in our transcriptome analysis, we noticed that many of them are also specifically expressed in preimplantation embryos³⁶. 57 out of 60 genes which were upregulated after 4.5hrs of DUX4 expression, overlapped with the single-cell sequencing (SCS) dataset of pre-implantation embryo development³⁶ and are specifically expressed at distinct, early stages of embryonic development (Fig. 3B and S2). Moreover, the expression of 43 DUX4-induced genes increase or peak at the 8-cell stage, which suggest and corroborates that DUX4 induction at the 4-cell stage is regulating the expression of many of these early genes. Figure S3 also validates that this increase or peak in expression is linked to DUX4 induction and is not due to a general increase in transcription around the 8-cell stage, as common housekeeping genes demonstrate a different expression pattern throughout development.

Table 1. DUX4 differentially expressed transcription- and co-factors

Gene	Factor	Expression peak Stage
ZSCAN4	transcription	8-cell
ZNF217	transcription	8-cell, Morulae, Epiblast
ZNF296	transcription	8-cell
LEUTX	transcription	8-cell
ZNF622	transcription	Morulae
ZNF574	transcription	4-cell
DUXA	transcription	8-cell, Morulae
HOXB2	transcription	8-cell, Primitive-endoderm
SNAI1	transcription	8-cell
ZNF705A	transcription	8-cell
OSR2	transcription	Oocyte, 4-cell, Morulae
CCNA1	co	8-cell
HSPA1A	co	Oocyte, 4-cell
GTF2F1	co	Oocyte to Morulae
HSPA1B	co	4-cell
MED26	co	8-cell
ID1*	transcription	hESC
ID3*	transcription	hESC
HES7*	transcription	Morulae

* Significant down regulated factors in induced DIE cells

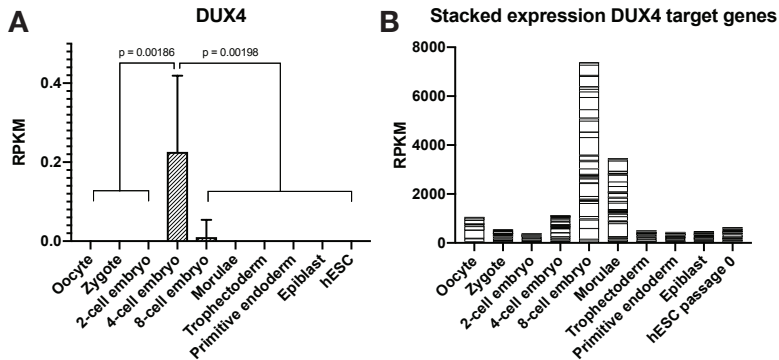


Figure 3. Expression of DUX4 and DUX4 early target genes during pre-implantation development. (A) DUX4 expression in preimplantation embryos. DUX4 expression is significantly upregulated at the 4-cell embryonic stage. Single cell RNA-seq data from preimplantation embryo's is from Yan et al³⁶. Statistical significance was determined by a two-tailed Student t-test. **(B)** A graph demonstrating the stacked expression of 54 DUX4 activated target genes identified by RNA-seq, in reads per kilobase per million (RPKM). The genes KHDC1L, DPPA3 and RGS2 were excluded due to disproportional high expression, which would otherwise skew the data (see Fig. S2C). hESC: human embryonic stem cells.

Screening for factors which modulate DUX4 cytotoxicity

To test if any of these differentially upregulated genes contribute to DUX4-induced cytotoxicity, we developed a small-scale CRISPR/Cas9 screening assay to quickly, efficiently and cost-effectively screen up to a few hundred sgRNAs. To set up this screening platform, a CRISPR/Cas9 reporter line was used for an easy and quantifiable read out to track the effectiveness of screening conditions. The reporter line consists of a constitutively expressed out of frame non-fluorescent dTomato gene, with an AAVSI target site directly upstream of the dTomato coding sequence (Fig. 4). Targeting the AAVSI sequence with CRISPR/Cas9 induces a frame shift, restoring the reading frame between the ATG start and the dTomato transgene, thereby activating dTomato fluorescence in the target cells³⁷. AAVSI sgRNA was generated by *in-vitro* transcription from a dsDNA template that was created by annealing two ssDNA oligomers and filling in the overhangs by PCR (Fig 5A&B). The dsDNA template was then used to generate the sgRNA by *in-vitro* transcription from the T7 RNA-polymerase promoter. Since the sgRNA mixture containing both PCR and IVT components was to be directly transduced into the reporter cells together with Cas9 protein (Fig. 5C), we tested different ratios of PCR and IVT mixtures, to identify which mixture minimally impacted growth and survival of the cells upon transduction (Table 2). Figure 5D shows the toxic effect of the indicated mixtures on adherent KBM7 reporter cells. Cell growth and survival could be further enhanced, by reducing the overall osmolarity of the transduction mixture (Figure 5D, bottom panel). Quantification of dTomato expression by FACS analysis confirmed these results, demonstrating improved survival and efficiencies when the PCR mixture was diluted more. Reduction of the osmolarity of the transduction mixture and reduction of the transduction time, from 45 to 35 minutes further enhanced cell survival (Figure 5E). The optimized experimental setup was subsequently used to screen the 60 upregulated genes found during RNA seq analysis. Short 57nt oligomers with gene specific spacer sequences were used to generate 3 different sgRNA per gene (Table S7). The spacer sequences were designed using the GPP sgRNA designer tool^{38,39}. Guides targeting the DUX4 transgene were used as positive controls. While the screen worked technically, based on cell survival seen in

the positive controls, no increased viability was observed in the other targeted DIE samples (Data not shown). These results imply that knocking-out any of these 60 “early” genes individually was insufficient to mitigate DUX4 cytotoxicity.

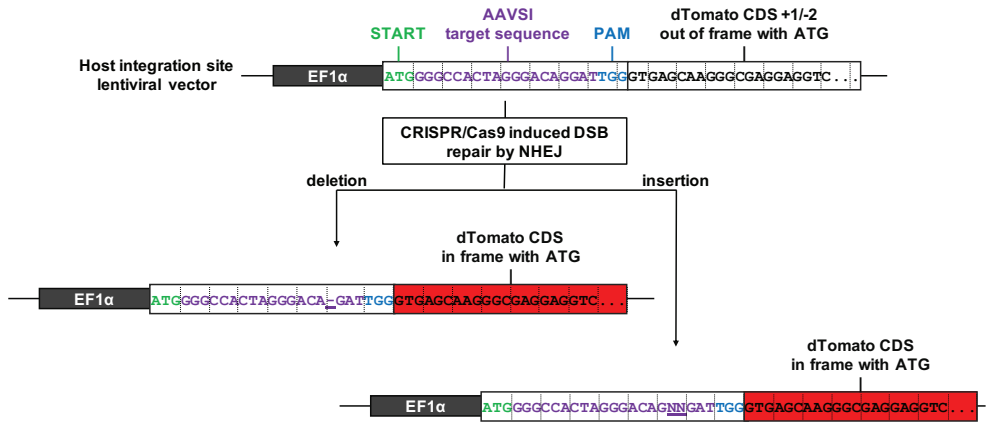


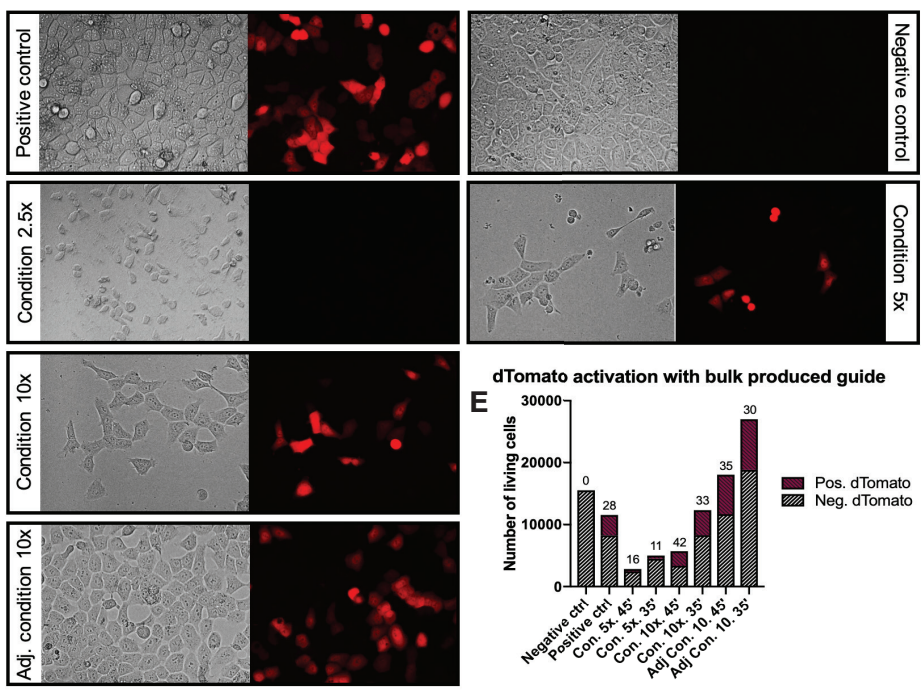
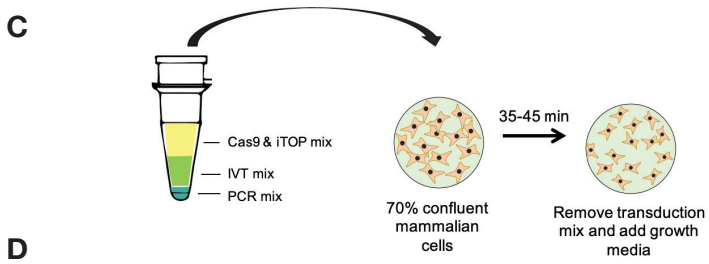
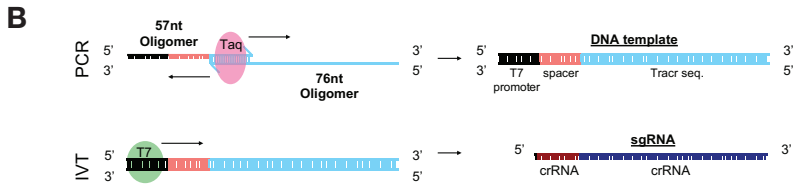
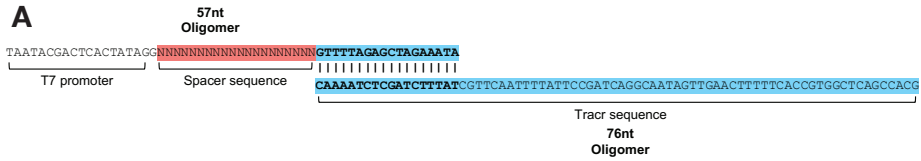
Figure 4. Schematic representation of the CRISPR/Cas9 reporter line. The stable integrated lentiviral vector contains an elongation factor-1 alpha (EF1 α) promoter, which regulates the expression of the dTomato gene. The out of frame dTomato sequence contains an AAVSI targeting sequence right between the start codon (ATG), and the dTomato coding sequence (CDS). Due to the presence of one additional nucleotide or the absence of 2 nucleotides (+1/-2) in the AAVSI sequence, the last nucleotide of the AAVSI sequence forms a codon with the first two nucleotides of dTomato CDS, putting the dTomato out of frame. Cas9 targeting of the AAVSI sequence can cause indels of different sizes, which can put the dTomato gene back in frame.

Table 2. IV-RT reactions with different amount of PCR mixture

	PCR reaction dilutions			
	1.5x	2.5x	5x	10x
Nuclease free water	0	10	20	25
T7 10x reaction buffer	5	5	5	5
25mM NTP mix	10	10	10	10
5 μ M T7	5	5	5	5
Template DNA (PCR mix)	30	20	10	5
Total volume (in ul)	50	50	50	50

Figure 5. Set up of a fast and efficient CRISPR/Cas9 small-scale screen. **A and B)** A schematic representation of the generation of the dsDNA template used for the production of sgRNAs. **A)** The 57nt single stranded (ss) DNA oligo containing a T7 promoter sequence, a guanine nucleotide, a spacer sequence, and a short piece of the tracr sequence, is annealed to a 76nt ssDNA oligomer containing the full tracr sequence. **B)** A PCR reaction fills up the 5' overhangs, generating a double stranded DNA template for in-vitro transcription (IVT). **C)** A schematic representation of the CRISPR/Cas9 iTOP transduction procedure. The dsDNA template and the sgRNA are made in a consecutive manner in the same tube/container. The PCR/IVT mixture is subsequently supplemented with transduction media and protein to produce the transduction solution. The transduction solution is added to cells. After the 35-45 min transduction period, the size of the cells will be temporarily reduced due to the loss of water. **D)** Bright field and fluorescent microscopy images of the KBM7 dTomato reporter cells transduced with Cas9 and sgRNA. The top panels show the controls. The positive controls are the dTomato reporter cells transduced with

Cas9 and purified sgRNA. The negative control are the same reporter cells transduced with Cas9 only. Panels below show the reporter cells which were transduced with purified Cas9 and various conditions of unpurified IVT sgRNA. E) FACS analysis showing the amount of living cells and percentage of dTomato positive cells of each of the tested condition.



Discussion

The lack of effective treatment for FSHD is, among other things, due to the complex nature of FSHD and incomplete understanding of the underlying molecular mechanism that initiates upon DUX4 expression and the sequential chain of events that ensues. To uncover the cytotoxic mechanisms induced by DUX4, we generated a DUX4-inducible cell line (DIE cell line). Since even sporadic, intermittent and low DUX4 expression has been shown to be sufficient to cause profound muscle degeneration^{19,24,25}, the potent cytotoxic effect of DUX4 is evident. This toxicity proved to be a big hurdle when generating the inducible DUX4-expressing line, ultimately resulting in a single clone out of 80 that displayed tight, DUX4-mediated apoptosis in a doxycycline-dependent manner. Many known DUX4 target genes are induced upon DUX4 induction in these DIE cells, despite their myeloid leukemia cellular background, supporting DUX4's role as a pioneer factor. Furthermore, many induced target genes encode transcription factors and cofactors, which in turn can activate their transcriptional program, potentially continuing and exacerbating the cytotoxic cascade.

With our inducible in vitro model for DUX4 cytotoxicity, we set out to investigate the underlying mechanism by which DUX4 expression leads to cell death. We developed and employed a small-scale CRISPR/Cas9 screening assay. This assay allowed us to quickly and cost-effectively screen 183 sgRNAs, targeting 61 genes, including DUX4 itself (Table S7). Because the guides are screened individually, it allowed us to directly assess the effect of individual genes downstream of DUX4. The developed screening method has the advantage that it does not involve any cloning steps, or the generation of a viral library. Combined, these properties make this screening tool fast, effective and cost-efficient. The small-scale screening assay was deployed to screen 60 early DUX4 targets, to test the hypothesis that knock-out of early genes holds more potential for modulating DUX4 toxicity, as interference would occur early in the DUX4 induced cascade. However, other than knocking out DUX4 itself, none of the other 60 targeted genes showed an effect on the DUX4 induced cytotoxicity. These results are indicative of the complex nature of FSHD, which is most likely mediated by more than one gene, acting up and/or downstream of DUX4.

This cell model did however demonstrate the effectiveness of these types of screening assays, which has provided us another tool in the toolkit for future small-scale screening purposes, or follow up experiments.

Here, we established a system that will allow us to identify cellular events and gene expression over time. This inducible system allows us to simultaneously control the amount of DUX4 expression, the timing of DUX4 induction and the length of the induction time. We have shown that the downstream transcriptional changes that follow DUX4 expression in the DIE cell line, greatly overlap with reported data of FSHD myoblasts, demonstrating that the DIE cell system recapitulates the molecular events underlying the disease. Given the complex nature of FSHD, this system enables us to obtain a more thorough insight into the temporal sequence of events that occur downstream of DUX4, which is the focus in chapter 3. This versatile model can furthermore be used to develop targeting strategies aimed at the DUX4 gene, as this system contains the genomic sequence (three exons and two introns) of DUX4, which is the focus in chapter 5. Ultimately, this system will allow us to develop targeting strategies and/or identify molecular events that are relevant for the pathogenesis of FSHD.

Methods

Cloning and generating the DIE cell line

To generate the inducible DsRed/DUX4 system, the third generation lenti-viral plasmid pRRLsincPPT-wpre⁴⁰ was used as the backbone. The linearized viral backbone was created by restriction digestion using HpaI and Sall (NEB). All inserts were generated with PCR amplification using phusion DNA polymerase (Fischer Scientific). Inserts were created with 15bp adapter sequences, matching the backbone or neighboring fragments, for in-fusion cloning (Clontech). The first fragment consisted of cPPT/CTS-TRE-mCMV sequences, and the second fragment contained the LoxP-DsRed-LoxP (LSL) sequence. After inserting these two fragments into the pRRLsincPPT-wpre backbone, this newly cloned construct was transformed into chemically competent Stbl3 Escherichia coli (E.coli). The plasmid was isolated and purified from the Stbl3 cells using the HiPure plasmid kits from Invitrogen (Fischer scientific). This TRE-LSL plasmid was then digested with XbaI and EcoRI (NEB) after which the remaining three inserts: DUX4 (exon1-3), mPGK and PuroR-WPRE, were cloned downstream from the LoxP-DsRed-LoxP in similar fashion.

The DIE cell line was obtained by transducing diploid KBM7 cells with lentiviral particles containing the inducible DsRed/DUX4 cassette mentioned above. 2 days after lentiviral transduction, transfected cells were selected with puromycin. After establishing a stable line by puromycin selection, lentiviral particles containing CMV-rtTA3-BlastR were added to the DsRed/DUX4 containing KBM7 cells. Positively transfected cells were subsequently selected with blasticidin, and FACs sorted for DsRed expression after exposure to doxycycline. The pLenti CMV rtTA3 Blast (w756-1) plasmid was a gift from Eric Campeau (Addgene plasmid #26429, <http://www.addgene.org/browse/article/3669/>).

Cell culturing

The KBM7 cells that were used to create the DIE line were near-haploid^{8,21}. Haploid cells are however unstable and do not remain haploid (reviewed in Yilmaz et al.)⁴¹ and rediploidize^{22,23}. KBM7 cells were cultured in IMDM media (Fischer Scientific) and 10% FBS. The DIE cells were cultured in IMDM media with 10% Tet system-approved FBS (Clontech), supplemented with 5µg/ml Puromycin and 6µg/ml Blasticidin. For transduction experiments, wells from a 96-wells plate were coated with Matrigel coated wells (Matrigel in PBS 1:250). 15,000 cells were seeded on top of the coated wells and incubated overnight (O/N) at 5% CO₂ and 37°C, until 70-80% confluency was reached.

Doxycycline titration curve

200,000 cells were seeded into wells of a 24-wells plates and kept at 5% CO₂ and 37°C until they reached a confluency of 90-100%. Different concentrations of doxycycline were added to the vertical lanes (100 ng/ml, 250ng/ml, 500ng/ml, 750ng/ml, 1000 ng/ml), with the horizontal lanes experiencing different exposure times (48h, 36h, 24h, 12h). After the doxycycline exposure, wells were washed with DPBS and were given a recovery period of 96. Surviving cells were subsequently stained using Giemsa modified staining solution (See paragraph viability staining).

Viability staining

Tissue culture (TC) plates containing cultured cells were washed with DPBS, and fixed with

100% Methanol for 10 minutes. Giemsa stain modified solution (Sigma) was subsequently added for 45 minutes, after which it was removed and the wells were washed with demineralized water.

Protein extraction and Western blot

DIE cells were harvested by trypsinization and lysed in RIPA buffer. Total protein concentrations were determined using a Pierce BCA protein assay kit (Fischer Scientific). 20ug protein was denatured using 4x Laemmli sample buffer (Bio-rad) with 10% BME (Sigma), and boiled for 5 minutes. Samples were run on a 15% SDS-polyacrylamide gel and transferred to a PVDF membrane (Merck). Membranes were blocked for an hour using 5% BSA in TBST, and were subsequently incubated overnight with anti-DUX4 antibody [E5-5] (Abcam, ab124699) in blocking solution (5% BSA in TBST), at 4°C. Membranes were then incubated for an hour with Secondary goat anti-rabbit-HRP antibody (Santa Cruz, sc-2004), and primary rabbit mAb β -Actin HRP conjugated antibody (Cell signaling, 5125s) in blocking buffer. Chemiluminescent signal was detected using GE ImageQuant LAS 4000 imager, using Pierce ECL Plus Western Blotting substrate (Fischer Scientific).

RNA extraction and RT-qPCR

Cultured cells were rinsed with DPBS just prior to the additional of TRIzol reagent (Thermo Scientific). Total RNA samples were subsequently extracted by addition chloroform, and isopropanol precipitation, and finally treated with RNase free DNase I (Promega). Reverse transcription was performed using the Superscript III kit (Invitrogen) and random primers (Promega), generating cDNA. Quantitative PCR was then initiated using IQ SYBR Green Supermix (Bio-Rad 1708880), 50 ng of cDNA, and the following gene-specific primers:

- DUX4: 5'-CCCAGGTACCAGCAGACC-3',
5'-TCCAGGAGATGTAAGTCTAATCCA-3'⁴² ;
- ZSCAN4: 5'-GTGGCCACTGCAATGACAA-3',
5'-AGCTTCCTGTCCCTGCATGT-3'⁴²;
- ZNF217: 5'-AAGCCCTATGGTGGCTCC-3',
5'-TTGATATGACACAGGCCCTTTTC-3'⁴²;
- PRAMEF1: 5'-CTCCAAGGACGGTTAGTTGC-3',
5'-AGTTCTCCAAGGGTTCTGG-3'⁴²;
- LEUTX: 5'-GGCCACGCACAAGATTTCTC-3',
5'-TCTTGAACCAGATCTTTACTACGGA-3';

Data were normalized to HPRT expression by using the following primer pair: 5'-CCTGGCGTCGTGATTAGTGA-3', 5'-CGAGCAAGACGTTTCAGTCCT-3'⁴³.

Live imaging

DIE cells were seeded into an 8-chamber coverslip slide (Ibidi) 24-36 hours prior to imaging. Growth media was supplemented with 1:50 Annexin5-Alexa Fluor 488, 1:100 Propidium Iodide and 1ug/ml doxycycline, and imaged for 12h with a Confocal Zeiss LSM 700 microscope at 37°C and % CO₂.

RNAseq sample preparation and sequencing

Cultured cells were rinsed with DPBS just prior to the additional of TRIzol reagent (Thermo Scientific). Total RNA samples were subsequently extracted by addition chloroform, and

isopropanol precipitation. The library prep was performed using CEL-seq1 primers⁴⁴ and the Life technologies Ambion kit (AM1751)⁴⁵, and were processed using CEL-seq2 protocol⁴⁶. Samples were sequenced using Illumina Nextseq 500, 2x75 kit, high output. Four technical replicates per samples were sent for sequencing, and were sequenced to an average of 600.000 reads per replicate (combined read count of 2.4 million reads per sample). Differential expression analysis was done using the DESeq2 package⁴⁷.

Small-scale CRISPR-Cas9 screen by iTOP

Large quantities of 180 sgRNAs were generated by producing a dsDNA template for each sgRNA. For this, a short ssDNA fragment containing the gene-specific spacer sequence was annealed to a longer ssDNA fragment containing the complete complementary sequence of the spCas9 tracr. The short ssDNA piece contains a T7 promotor, an additional guanine nucleotide, the 20nt spacer sequence, and the first 19nt of the spCas9 tracr sequence: 5'-TAATACGACTCACTATAGG-20nt-GTTTTAGAGCTAGAAATAG-3'. This short ssDNA fragment was annealed to a longer ssDNA piece containing the complete complementary sequence of the spCas9 tracr:5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTAAAAAGTGGCACCGAGTCGGTGC-3'. The shorter variable pieces of ssDNA were supplied in a 96-well plate. The longer ssDNA fragment containing the anti-sense tracr sequence was mixed in a 1:1 ration to all wells containing the variable short ssDNA Oligomer. The two fragments are able to anneal to each other due to 19nt tracr sequence that is complementary between the two ssDNA fragments, which serves as a primed region for amplification. Taq DNA polymerase (Thermo scientific) was used to fill up the ss overhangs, thereby creating a dsDNA template for IV-RT reaction. IV-RT kit used for the generation of sgRNA was supplied by NTRANS technologies. The total volume of the PCR reaction did not exceed 5µl, and was diluted 10x in the IV-RT mixture, that consisted of nuclease free water, 1x T7 RNA polymerase reaction buffer, 5nM of each NTP (Jenabioscience), and 500nM of T7 RNA polymerase. The IV-RT reaction mixture was incubated overnight (12-15h) at 37°C. Residual DNA was removed by the addition of 2U Turbo DNase (Fischer scientific) per sample, and incubation at 37°C for 30 min. The DNase was inactivated by an incubation step of 65°C for 10 min. Cas9 protein (in 5x Transduction buffer) and adjusted CRISPR/Cas9 transduction media was added to the newly synthesized sgRNA in appropriate volumes (See Table 3 for the composition of the adjusted CRISPR-Cas9 media). Half of the transduction mixture was added to 70-80% confluent cells, that were plated out a day before on Matrigel coated 96-wells plates. The cells were exposed to the transduction mixture for 40-45min, after which the mix was removed and normal growth media was added gently, completely filling up the well to dilute out remaining transduction mixture. After a recovery period of 72-96 hours, 1000ng/ml of doxycycline was added for 24h. The doxycycline media was subsequently removed, the wells were washed with DPBS to remove the majority of dead/floating cells, normal growth media was added to the cells and all plates were then placed back into the 5% CO₂ and 37°C incubator for 2-4 days. This allowed the remaining cells that have started the apoptotic process to perish, or let surviving cells grow out and therefore become more visible.

Flowcytometry sorting (FACS) and analysis

dTomato reporter cells were trypsinized using 0.25% Trypsin-EDTA, then resuspended in iMDM media with 10% FBS and DAPI nuclear stain. Cells were subsequently strained using Cell-strainer capped tubes (Falcon) and analyzed using the BD FACSCanto II.

Data Resources

Data containing the bulk RNA sequencing samples in quadruplicate are available from the GEO data base, accession number: GSE154649.

Table 3. Adjusted CRISPR-Cas9 transduction media (5 ml)

Compound	amount
GABA	208 mg
5M NaCl	550 ul
100x Glutamine	75 ul
100x non-essential amino acids	75 ul
100x N2 supplement	75 ul
50x B27 supplement	150 ul
Opti-MEM	3780 ul
100ug/ml EGF	5ul
100 ug/ml bFGF	10ul
MiliQ	98ul

References

1. Deenen, J. C. W. *et al.* Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* **83**, 1056–1059 (2014).
2. Tawil, R. & Van Der Maarel, S. M. Facioscapulohumeral muscular dystrophy. *Muscle and Nerve* **34**, 1–15 (2006).
3. Lemmers, R. J. L. F. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329**, 1650–1653 (2010).
4. Choi, S. H. *et al.* DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **44**, 5161–5173 (2016).
5. Vuoristo, S. *et al.* DUX4 regulates oocyte to embryo transition in human. *Biorxiv* (2019) doi:<http://dx.doi.org/10.1101/732289>.
6. Jones, T. I., Himeda, C. L., Perez, D. P. & Jones, P. L. Large family cohorts of lymphoblastoid cells provide a new cellular model for investigating facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* **27**, 221–238 (2017).
7. Banerji, C. R. S., Zammit, P. S. & Panamarova, M. Lymphocytes contribute to DUX4 target genes in 1 FSHD muscle biopsies. *Biorxiv* (2019) doi:<http://dx.doi.org/10.1101/717652>.
8. Mulherkar, N. *et al.* Ebola Virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340–343 (2012).
9. Papatheodorou, P. *et al.* Lipolysis-stimulated lipoprotein receptor (LSR) is the host receptor for the binary toxin Clostridium difficile transferase (CDT). *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16422–16427 (2011).
10. Jae, L. T. *et al.* Deciphering the Glycosylome of Dystroglycanopathies Using Haploid Screens for Lassa Virus Entry. **340**, 479–483 (2014).
11. Jae, L. T. *et al.* Lassa virus entry requires a trigger-induced receptor switch. *Science* **344**, 1506–1510 (2014).
12. Blomen, V. A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–6 (2015).
13. Rong, Y. *et al.* Genome-wide screening of genes required for glycosylphosphatidylinositol biosynthesis. *PLoS One* **10**, 1–18 (2015).
14. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
15. Gerhards, N. M. *et al.* Haploid genetic screens identify genetic vulnerabilities to

- microtubule-targeting agents. *Mol. Oncol.* **12**, 953–971 (2018).
16. Luteijn, R. D. *et al.* A Genome-Wide Haploid Genetic Screen Identifies Heparan Sulfate-Associated Genes and the Macropinocytosis Modulator TMED10 as Factors Supporting Vaccinia Virus Infection. *J. Virol.* **93**, (2019).
 17. Mezzadra, R. *et al.* SLFN11 can sensitize tumor cells towards IFN- γ -mediated T cell killing. *PLoS One* **14**, 1–16 (2019).
 18. Geng, L. N. *et al.* DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy. *Dev. Cell* **22**, 38–51 (2012).
 19. Rickard, A. M., Petek, L. M. & Miller, D. G. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum. Mol. Genet.* **24**, 5901–5914 (2015).
 20. Van Den Heuvel, A. *et al.* Single-cell RNA sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. *Hum. Mol. Genet.* **28**, 1064–1075 (2019).
 21. Andersson, B. S. *et al.* Ph-positive chronic myeloid leukemia with near-haploid conversion in vivo and establishment of a continuously growing cell line with similar cytogenetic pattern. *Cancer Genet. Cytogenet.* **24**, 335–343 (1987).
 22. Carette, J. E. *et al.* Generation of iPSCs from cultured human malignant cells. *Blood* **115**, 4039–4042 (2010).
 23. Olbrich, T. *et al.* A Chemical Screen Identifies Compounds Capable of Selecting for Haploidy in Mammalian Cells. *Cell Rep.* **28**, 597–604.e4 (2019).
 24. Snider, L. *et al.* Facioscapulohumeral dystrophy: Incomplete suppression of a retrotransposed gene. *PLoS Genet.* **6**, 1–14 (2010).
 25. Bosnakovski, D. *et al.* Muscle pathology from stochastic low level DUX4 expression in an FSHD mouse model. *Nat. Commun.* **8**, 1–9 (2017).
 26. Agha-Mohammadi, S. *et al.* Second-generation tetracycline-regulatable promoter: Repositioned tet operator elements optimize transactivator synergy while shorter minimal promoter offers tight basal leakiness. *J. Gene Med.* **6**, 817–828 (2004).
 27. Urlinger, S. *et al.* Exploring the sequence space for tetracycline-dependent transcriptional activators: Novel mutations yield expanded range and sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7963–7968 (2000).
 28. Lemmers, R. J. *et al.* Deep characterization of a common D4Z4 variant identifies biallelic DUX4 expression as a modifier for disease penetrance in FSHD2. *Eur. J. Hum. Genet.* **26**, 94–106 (2018).
 29. De Vree, P. J. P. *et al.* Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat. Biotechnol.* **32**, 1019–1025 (2014).
 30. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, (2013).
 31. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
 32. Jagannathan, S. *et al.* Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum. Mol. Genet.* **25**, ddw271 (2016).
 33. Robert, C. & Watson, M. Errors in RNA-Seq quantification affect genes of relevance to human disease. *Genome Biol.* **16**, 1–16 (2015).
 34. Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
 35. Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **49**, 935–940 (2017).
 36. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1131–1139 (2013).
 37. D'Astolfo, D. S. *et al.* Efficient intracellular delivery of native proteins. *Cell* **161**, 674–690 (2015).
 38. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
 39. Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat. Commun.* **9**, 1–15 (2018).
 40. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**, 8463–71 (1998).

41. Yilmaz, A., Peretz, M., Sagi, I. & Benvenisty, N. Haploid Human Embryonic Stem Cells: Half the Genome, Double the Value. *Cell Stem Cell* **19**, 569–572 (2016).
42. Ferreboeuf, M. *et al.* DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. *Hum. Mol. Genet.* **23**, 171–181 (2014).
43. van Attekum, M., Terpstra, S., Reinen, E., Kater, A. & Eldering, E. Macrophage-mediated chronic lymphocytic leukemia cell survival is independent of APRIL signaling. *Cell Death Discov.* **2**, (2016).
44. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification. *Cell Rep.* **2**, 666–673 (2012).
45. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
46. Hashimshony, T. *et al.* CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 1–7 (2016).
47. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

Supplementary movie legend

Movie 1. Related to Figure 1. Adherent KBM7 cells in growth media supplemented with doxycycline and AnnexinV-Alexa Fluor 488 conjugated antibody. Live imaging was done using a confocal Zeiss LSM 700 microscope.

Movie 2. Related to Figure 1. DIE cells in growth media supplemented with AnnexinV-Alexa Fluor 488 conjugated antibody. Live imaging was done using a confocal Zeiss LSM 700 microscope.

Movie 3. Related to Figure 1. DIE cells in growth media supplemented with doxycycline and AnnexinV-Alexa Fluor 488 conjugated antibody. Live imaging was done using a confocal Zeiss LSM 700 microscope.

Supplementary Material

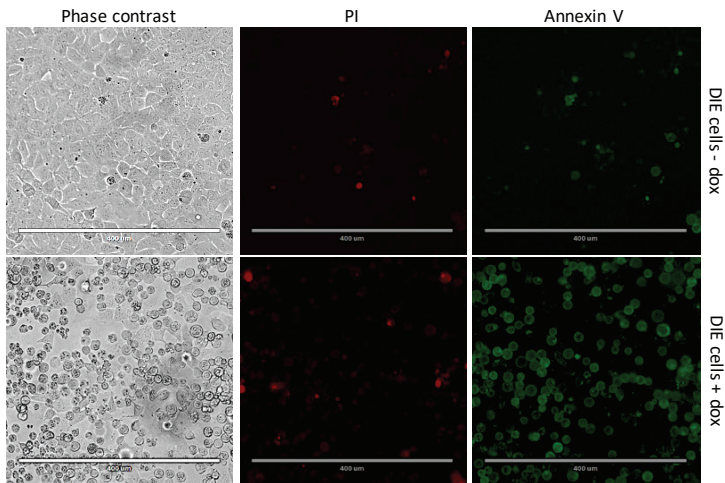
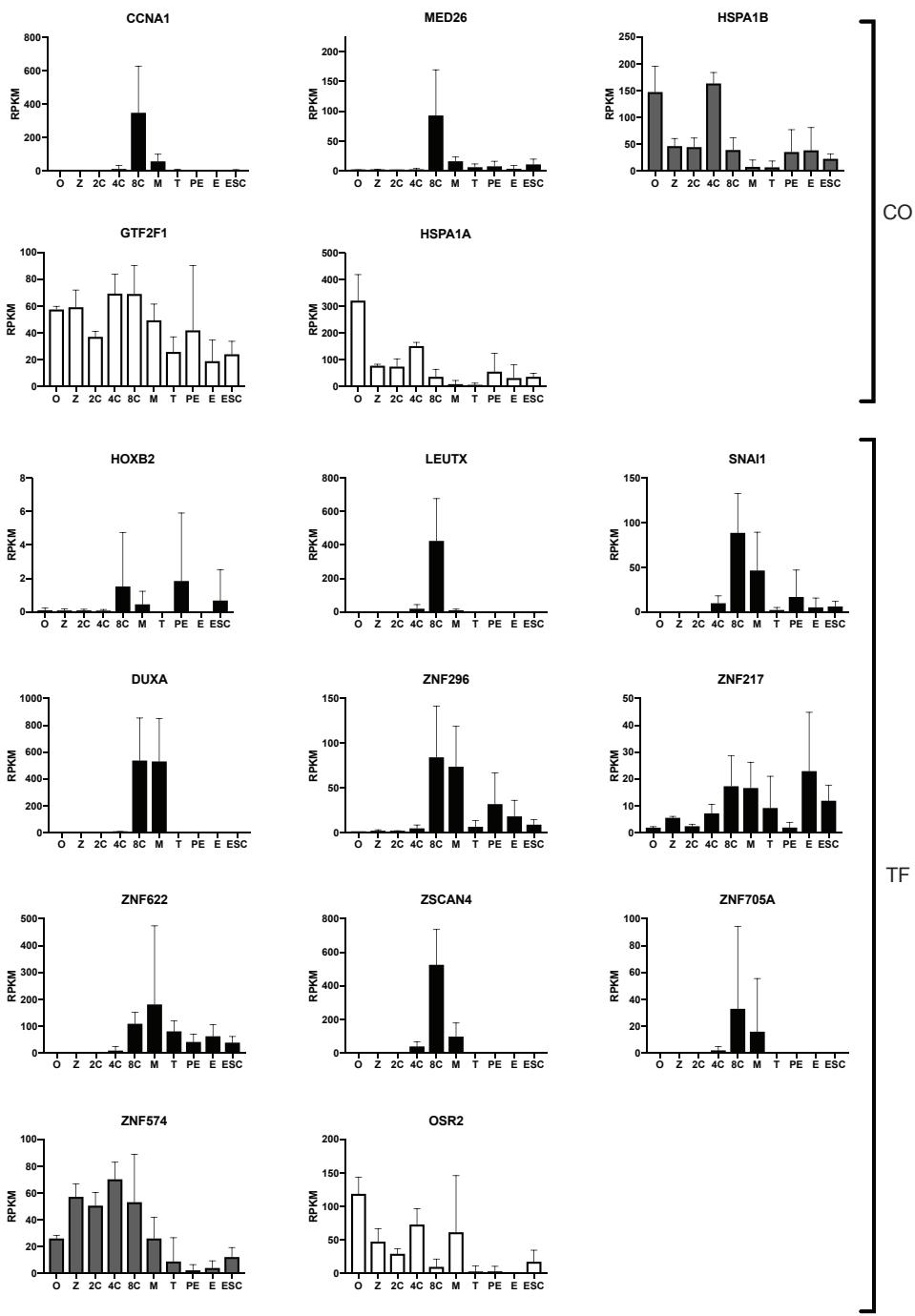
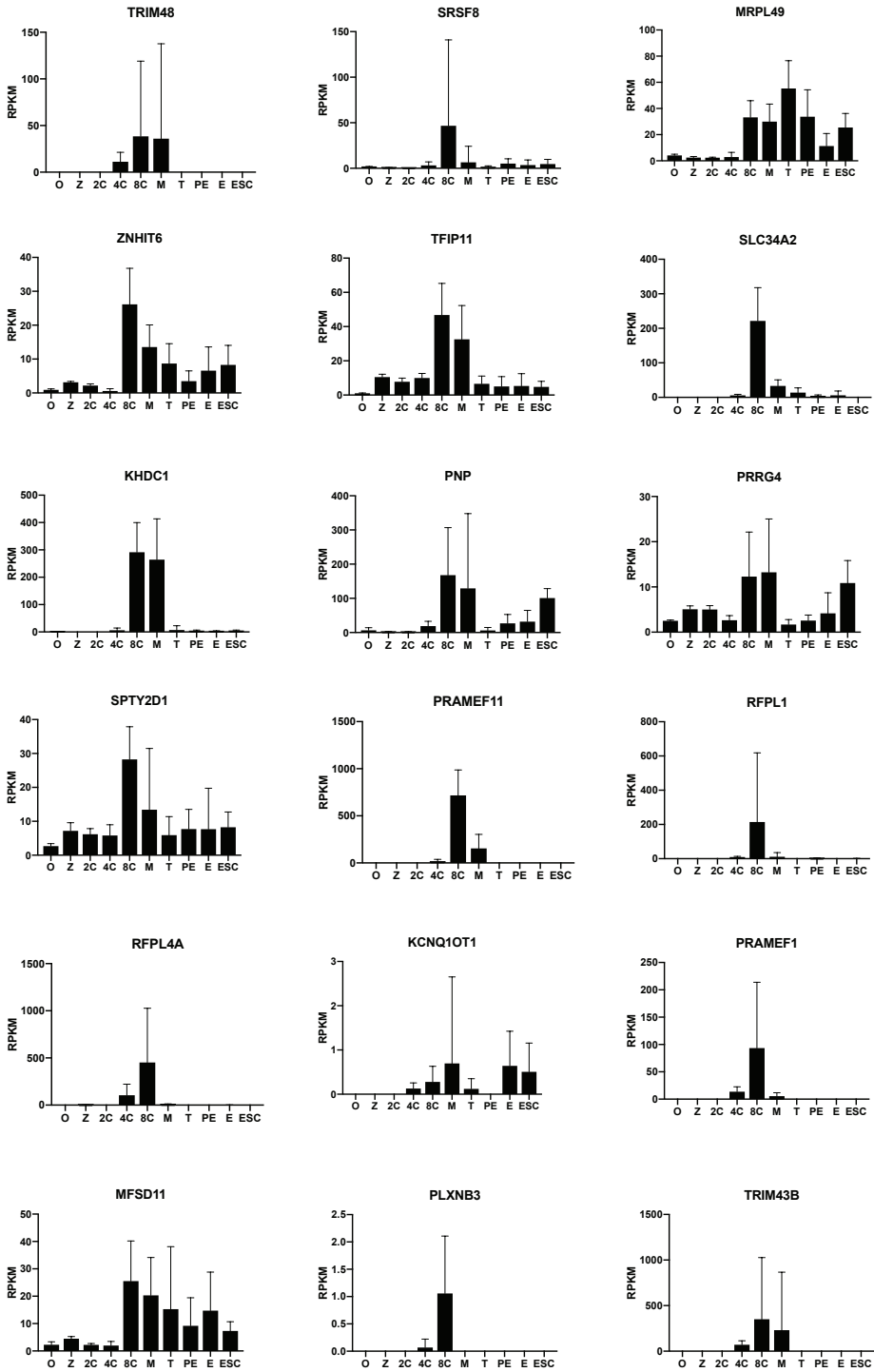


Figure S1. Living and dying DIE cells. Uninduced (top panel) and doxycycline-induced (bottom panel) DIE cells, stained with Propidium Iodide (PI) (middle panel) and AnnexinV FITC (right panel), with a phase contrast image in the left panel. DIE cells in the bottom panel are stained positive for AnnexinV, with no increasing PI signal compared to uninduced DIE cells (top panel).

A

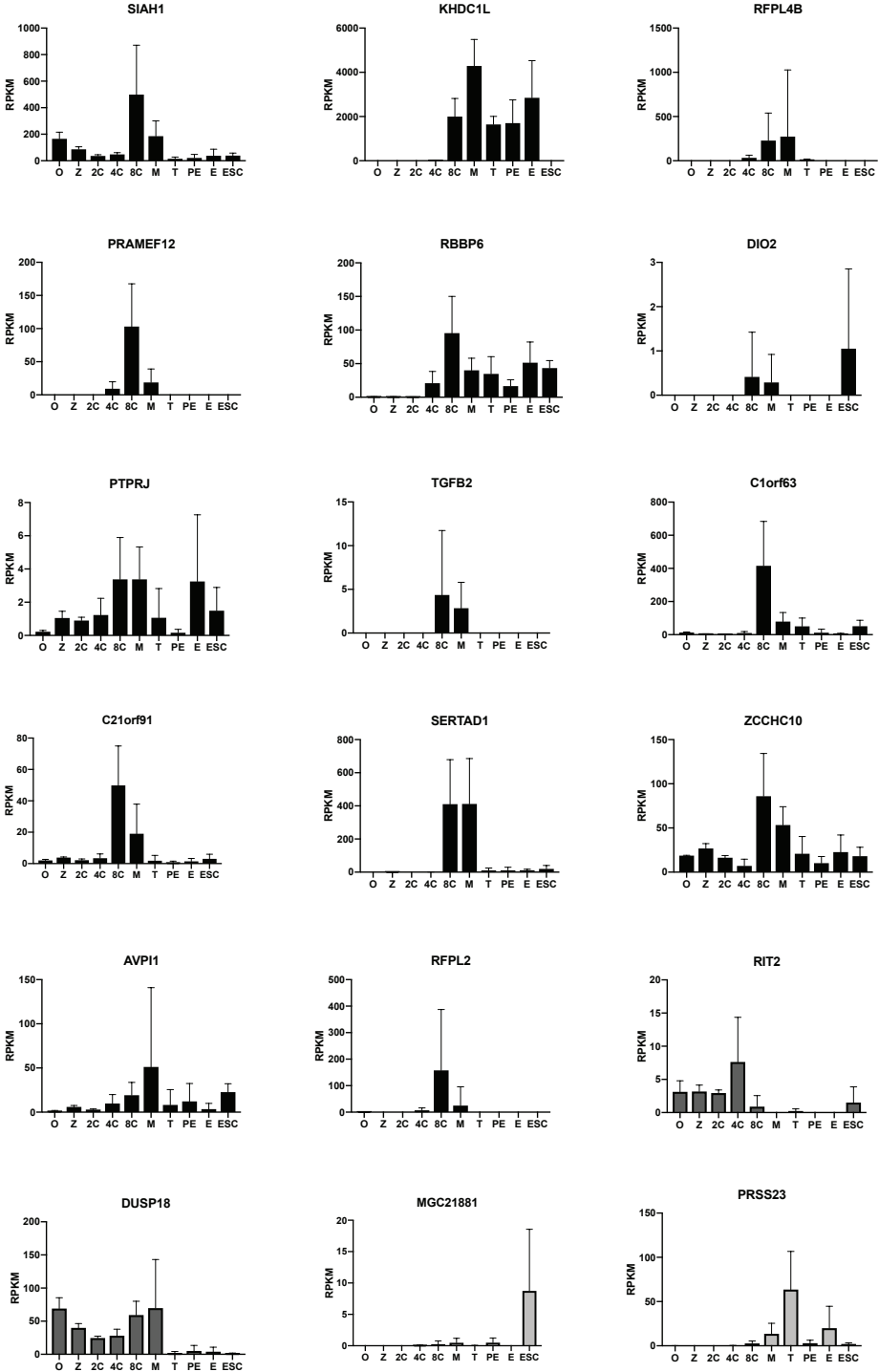


2

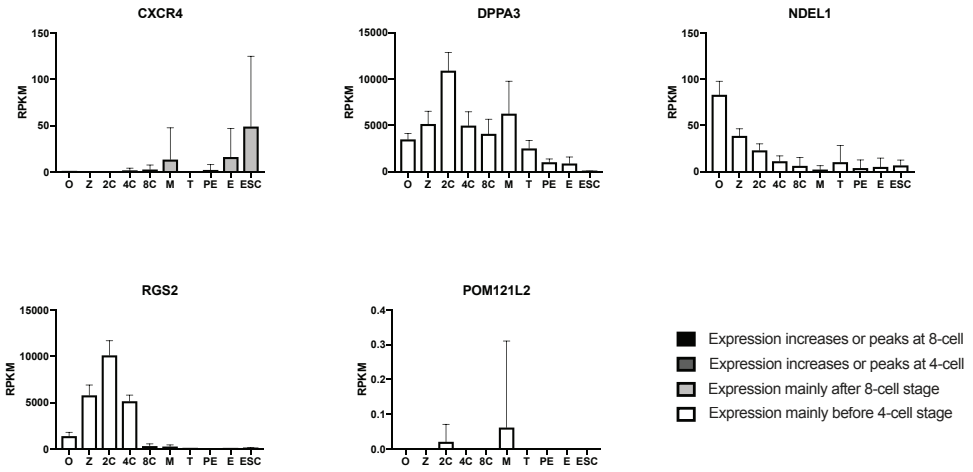
B

2

B (continued)



B (continued)



C

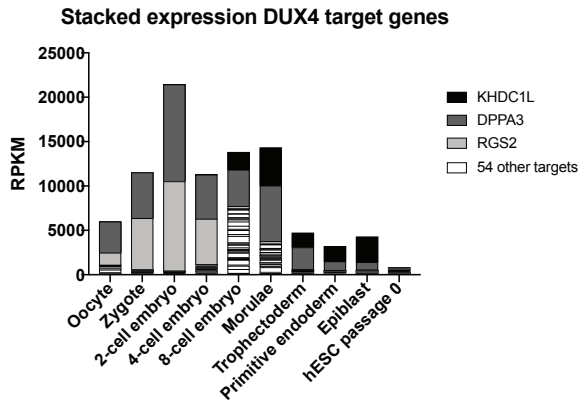
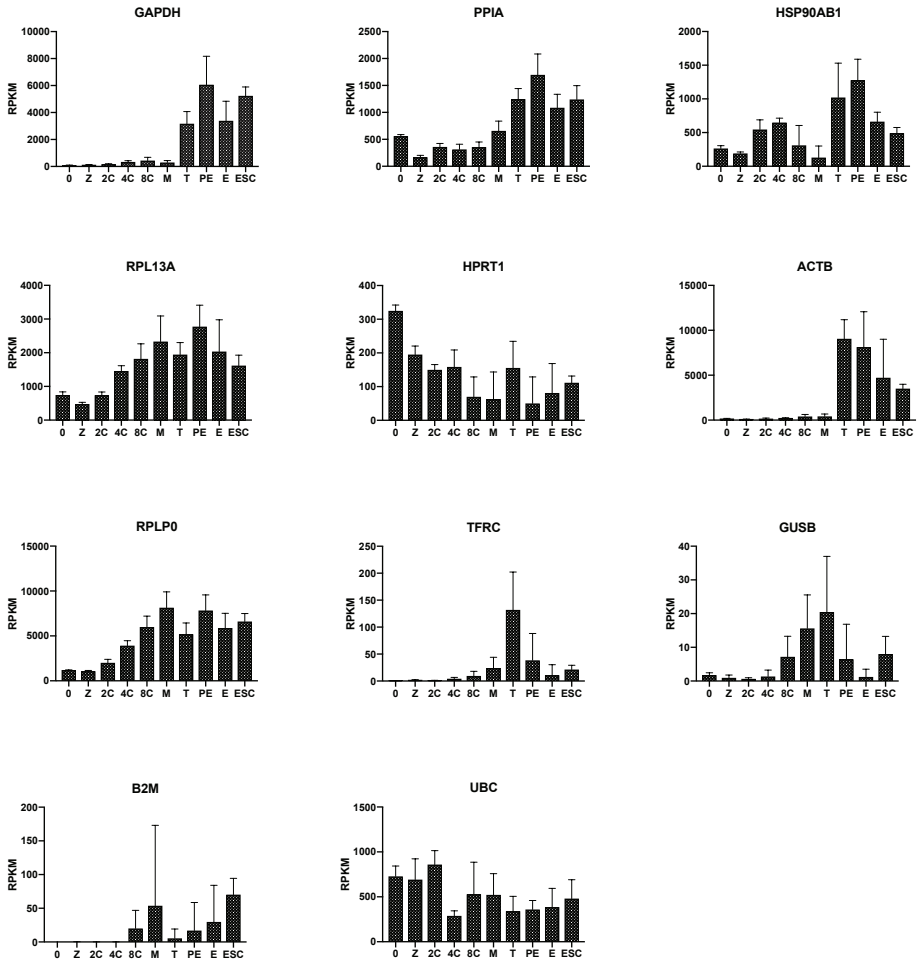


Figure S2. Expression of DUX4-induced early genes during embryonic development. (A) DUX4 induced cofactors (CO) and transcription-factors (TF) during the early stages of embryonic development in reads per kilobase per million (RPKM) mapped reads. (B) Expression of the other DUX4 induced early genes in RPKM. (C) Stacked expression of all 57 DUX4 induced early genes. KHDC1L, DPPA3, and RGS2 expression is disproportionately greater than the other 54 genes, and are individually color coded and annotated. O: Oocyte, Z: Zygote, 2C: 2-cell embryo, 4C: 4-cell embryo, 8C: 8-cell embryo, M: Morulae, T: Trophoctoderm, PE: Primitive endoderm, E: Epiblast, hESC: human embryonic stem cells. Single cell RNA-seq data from preimplantation embryo's is from Yan et al.³⁶.

A



B Stacked expression household genes

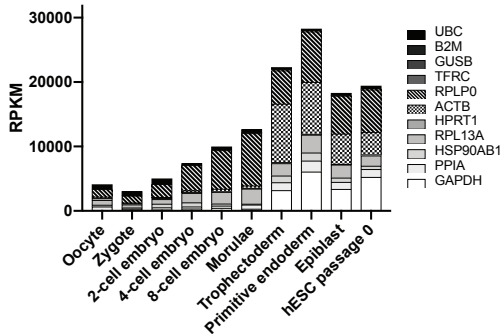


Figure S3. Expression housekeeping genes in preimplantation embryos. (A) Expression in RPKM mapped reads of 11 housekeeping genes in the early stages of embryonic development. **(B)** Stacked Expression of the housekeeping genes at same stages of early embryonic development, in RPKM. O: Oocyte, Z: Zygote, 2C: 2-cell embryo, 4C: 4-cell embryo, 8C: 8-cell embryo, M: Morulae, T: Trophectoderm, PE: Primitive endoderm, E: Epiblast, hESC: human embryonic stem cells. Single cell RNA-seq data from preimplantation embryo's is from Yan et al.³⁶

Table S1: Differentially expressed genes after 4.5h of doxycycline induction* Adjusted p value ≤ 0.01 , absolute $\log_2FC \geq 1$

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
ZSCAN4	285.62	4.08	0.18	23.25	1.51E-119	1.26E-115
LINC00633	137.90	3.26	0.18	18.02	1.43E-72	5.96E-69
ZNF217	358.21	2.26	0.13	17.44	4.24E-68	1.18E-64
SRSF8	133.28	2.41	0.14	17.29	5.57E-67	1.16E-63
PRAMEF1	120.48	4.25	0.25	17.10	1.40E-65	2.34E-62
RBBP6	221.34	1.90	0.12	15.39	2.03E-53	2.83E-50
PNP	260.82	1.76	0.12	14.69	8.01E-49	9.54E-46
ZNF296	105.83	2.73	0.19	14.28	2.94E-46	3.06E-43
SIAH1	89.75	2.93	0.21	14.12	2.74E-45	2.53E-42
TRIM51	68.72	3.39	0.25	13.68	1.42E-42	1.18E-39
RFPL4B	87.89	4.47	0.33	13.50	1.60E-41	1.22E-38
KHDC1L	131.92	2.86	0.22	13.08	4.20E-39	2.92E-36
CCNA1	208.50	2.17	0.17	12.64	1.27E-36	7.58E-34
TFIP11	83.33	2.73	0.22	12.64	1.26E-36	7.58E-34
PRAMEF12	63.82	3.69	0.30	12.32	7.12E-35	3.96E-32
LEUTX	415.52	3.62	0.30	12.02	2.83E-33	1.48E-30
ZNF622	66.31	2.26	0.19	11.78	5.22E-32	2.56E-29
HSPA1A	281.27	1.03	0.11	9.60	7.76E-22	3.59E-19
RFPL1	47.03	3.86	0.41	9.48	2.57E-21	1.13E-18
PTPRJ	33.57	3.15	0.34	9.28	1.70E-20	7.09E-18
RFPL2	60.64	3.32	0.36	9.22	2.94E-20	1.17E-17
RFPL4A	51.75	3.34	0.36	9.17	4.87E-20	1.85E-17
ID1	100.44	-1.47	0.16	-8.95	3.40E-19	1.23E-16
SLC34A2	57.57	2.19	0.25	8.61	7.45E-18	2.59E-15
TRIM43B	23.75	4.10	0.52	7.88	3.25E-15	1.08E-12
GTF2F1	58.72	1.54	0.20	7.70	1.32E-14	4.24E-12
TRIM48	29.06	2.95	0.40	7.36	1.81E-13	5.58E-11
DIO2	26.35	3.73	0.51	7.35	2.04E-13	6.07E-11
MRPL49	94.37	1.06	0.15	6.90	5.27E-12	1.46E-09
LOC441081	33.58	3.75	0.55	6.83	8.78E-12	2.36E-09
ZNF574	73.36	1.03	0.16	6.32	2.62E-10	6.83E-08
C1orf63	31.18	1.87	0.30	6.17	6.75E-10	1.71E-07
DUXA	29.80	2.60	0.42	6.14	8.50E-10	2.08E-07
HSPA1B	50.54	1.06	0.17	6.07	1.31E-09	3.13E-07
C21orf91	17.50	2.36	0.40	5.94	2.93E-09	6.78E-07
DUSP18	10.48	3.14	0.53	5.90	3.74E-09	8.21E-07

Table S1 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
ZNHIT6	60.07	1.14	0.19	5.90	3.71E-09	8.21E-07
DPPA3	32.41	2.47	0.42	5.83	5.49E-09	1.17E-06
AVPI1	32.14	1.35	0.24	5.73	1.01E-08	2.06E-06
CXCR4	15.51	3.89	0.70	5.58	2.44E-08	4.84E-06
ID3	42.15	-1.10	0.20	-5.55	2.90E-08	5.63E-06
NDEL1	34.42	1.19	0.23	5.22	1.82E-07	3.30E-05
HOXB2	28.18	1.42	0.28	5.16	2.45E-07	4.34E-05
MGC21881	36.17	1.22	0.24	5.07	3.89E-07	6.75E-05
MFSD11	20.82	1.71	0.34	5.07	4.07E-07	6.92E-05
PLXNB3	61.80	1.58	0.31	5.06	4.27E-07	7.12E-05
SNAI1	12.65	3.47	0.71	4.88	1.05E-06	1.68E-04
KHDC1	18.95	2.69	0.56	4.84	1.28E-06	2.02E-04
C20orf112	9.87	-2.37	0.50	-4.77	1.85E-06	2.86E-04
PRAMEF11	8.00	3.71	0.79	4.67	3.04E-06	4.53E-04
HES7	21.19	-1.34	0.29	-4.56	5.09E-06	7.30E-04
TMEM254-AS1	44.46	1.51	0.34	4.44	9.12E-06	1.23E-03
PRRG4	8.04	2.68	0.61	4.41	1.06E-05	1.37E-03
SPTY2D1	20.68	1.48	0.34	4.40	1.07E-05	1.38E-03
RIT2	8.91	4.75	1.08	4.38	1.16E-05	1.47E-03
ZNF705A	6.03	4.75	1.09	4.34	1.40E-05	1.74E-03
KCNQ1OT1	29.91	1.05	0.25	4.20	2.67E-05	3.05E-03
SERTAD1	17.17	1.28	0.31	4.14	3.54E-05	3.99E-03
TGFB2	13.62	1.94	0.47	4.13	3.59E-05	4.00E-03
RGS2	23.97	1.18	0.29	4.03	5.61E-05	6.00E-03
MED26	8.95	1.87	0.47	4.01	6.10E-05	6.36E-03
OSR2	7.12	2.32	0.58	4.01	6.08E-05	6.36E-03
PRSS23	8.83	1.91	0.48	3.94	8.02E-05	8.16E-03
ZCCHC10	18.53	1.22	0.31	3.88	1.03E-04	9.83E-03
PRSS23	7.14	1.91	0.49	3.92	8.80E-05	9.46E-03

Table S2: Differentially expressed genes after 8.5h of doxycycline induction

* Adjusted p value ≤ 0.01 , absolute $\log_2FC \geq 1$

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
ZSCAN4	285.62	5.33	0.17	30.85	5.93E-209	5.72E-205
CCNA1	208.50	4.55	0.16	28.08	1.79E-173	8.65E-170
LINC00633	137.90	4.56	0.17	26.11	3.04E-150	9.76E-147
ZNF217	358.21	3.30	0.13	26.07	8.67E-150	2.09E-146

Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
KHDC1L	131.92	5.34	0.21	25.98	8.35E-149	1.61E-145
PNP	260.82	2.75	0.12	23.74	1.35E-124	2.18E-121
PRAMEF1	120.48	5.59	0.24	22.96	1.09E-116	1.51E-113
ZNF296	105.83	4.07	0.18	22.27	7.75E-110	9.35E-107
SIAH1	89.75	4.29	0.20	21.65	5.98E-104	6.41E-101
SRSF8	133.28	2.83	0.14	20.76	9.70E-96	9.36E-93
TPRX1	61.43	4.67	0.24	19.39	8.53E-84	7.48E-81
TFIP11	83.33	4.00	0.21	19.25	1.29E-82	1.04E-79
PLXNB3	61.80	5.13	0.27	18.99	2.08E-80	1.54E-77
LEUTX	415.52	5.66	0.30	18.94	5.56E-80	3.83E-77
TRIM51	68.72	4.51	0.24	18.79	8.39E-79	5.40E-76
RFPL4B	87.89	6.04	0.33	18.55	8.01E-77	4.83E-74
SLC34A2	57.57	4.25	0.23	18.16	1.14E-73	6.46E-71
PRAMEF12	63.82	5.21	0.29	17.95	4.58E-72	2.46E-69
NXF1	107.66	2.18	0.12	17.61	1.96E-69	9.94E-67
RBBP6	221.34	2.11	0.12	17.34	2.45E-67	1.18E-64
RFPL2	60.64	5.79	0.34	16.86	9.15E-64	4.20E-61
ZNF622	66.31	3.09	0.18	16.83	1.48E-63	6.47E-61
PTP4A1	160.48	1.87	0.11	16.55	1.53E-61	6.41E-59
HNRNPF	350.21	1.97	0.12	16.45	7.99E-61	3.21E-58
TMEM254-AS1	44.46	4.83	0.29	16.42	1.27E-60	4.91E-58
RFPL4A	51.75	5.44	0.35	15.54	1.86E-54	6.90E-52
ZNHIT6	60.07	2.69	0.17	15.44	8.64E-54	3.09E-51
GTF2F1	58.72	2.85	0.18	15.42	1.18E-53	4.06E-51
RFPL1	47.03	5.85	0.39	14.85	6.90E-50	2.30E-47
MRPL49	94.37	2.04	0.14	14.30	2.15E-46	6.92E-44
DPPA3	32.41	5.33	0.39	13.70	9.53E-43	2.96E-40
RYBP	47.53	2.43	0.18	13.28	3.04E-40	9.17E-38
DUXA	29.80	5.12	0.39	13.05	6.10E-39	1.78E-36
PTPRJ	33.57	4.27	0.33	13.05	6.31E-39	1.79E-36
TRIM48	29.06	4.92	0.38	12.96	2.11E-38	5.83E-36
EXOSC10	86.08	1.82	0.14	12.95	2.41E-38	6.45E-36
TFAP2C	52.23	2.15	0.17	12.82	1.20E-37	3.14E-35
C1orf63	31.18	3.52	0.28	12.64	1.34E-36	3.40E-34
ANXA5	256.82	1.09	0.09	12.21	2.64E-34	6.53E-32
ALYREF	220.09	1.27	0.11	12.12	8.47E-34	2.04E-31
LOC441081	33.58	6.37	0.53	12.06	1.71E-33	4.03E-31
ZNF574	73.36	1.83	0.15	12.03	2.43E-33	5.59E-31



Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
INO80C	80.41	1.65	0.14	11.96	5.50E-33	1.23E-30
LINC00493	63.11	1.96	0.17	11.68	1.59E-31	3.48E-29
MGC21881	36.17	2.50	0.22	11.43	3.11E-30	6.66E-28
DIO2	26.35	5.54	0.49	11.26	1.97E-29	4.13E-27
ID1	100.44	-1.79	0.17	-10.79	3.95E-27	8.12E-25
TRIM43B	23.75	5.45	0.51	10.75	6.21E-27	1.25E-24
ALPPL2	19.14	3.30	0.31	10.72	8.59E-27	1.69E-24
AVPI1	32.14	2.30	0.22	10.56	4.47E-26	8.62E-24
KHDC1	18.95	5.35	0.52	10.38	3.18E-25	6.02E-23
RNF11	45.92	1.85	0.18	10.32	5.56E-25	1.03E-22
SPTY2D1	20.68	3.10	0.30	10.25	1.20E-24	2.18E-22
HOXB2	28.18	2.60	0.25	10.23	1.43E-24	2.56E-22
SNUPN	140.64	2.47	0.24	10.17	2.63E-24	4.61E-22
LOC100216545	25.93	2.52	0.25	10.12	4.33E-24	7.47E-22
RGS2	23.97	2.66	0.26	10.12	4.52E-24	7.66E-22
CCNJ	28.13	2.35	0.24	9.95	2.42E-23	4.03E-21
NDEL1	34.42	2.08	0.21	9.84	7.23E-23	1.18E-20
TCEB3	44.12	1.83	0.19	9.81	9.79E-23	1.57E-20
PNN	80.38	1.33	0.14	9.74	2.05E-22	3.25E-20
MFSD11	20.82	3.00	0.31	9.64	5.54E-22	8.63E-20
ADPGK	32.75	2.10	0.22	9.56	1.13E-21	1.73E-19
RFK	42.75	1.85	0.19	9.50	2.08E-21	3.14E-19
C21orf91	17.50	3.54	0.37	9.48	2.62E-21	3.88E-19
KDM5B	27.91	2.22	0.24	9.42	4.63E-21	6.76E-19
MMRN2	13.16	4.38	0.47	9.35	8.68E-21	1.25E-18
ODC1	132.72	1.04	0.11	9.25	2.21E-20	3.14E-18
ID3	42.15	-2.13	0.23	-9.20	3.60E-20	5.04E-18
DDX10	73.50	1.28	0.14	9.17	4.55E-20	6.27E-18
ARL4D	37.76	1.85	0.20	9.08	1.06E-19	1.44E-17
DYX1C1	33.45	1.85	0.21	8.99	2.39E-19	3.20E-17
TGFB2	13.62	3.81	0.43	8.96	3.31E-19	4.38E-17
C8orf33	76.19	1.35	0.15	8.85	8.44E-19	1.09E-16
CDC42EP1	122.65	-1.18	0.13	-8.84	9.56E-19	1.21E-16
ELOF1	115.88	1.01	0.12	8.79	1.54E-18	1.93E-16
LOC284551	12.05	4.12	0.47	8.77	1.75E-18	2.17E-16
SLC2A3	11.49	3.76	0.43	8.71	3.08E-18	3.76E-16
TRAPPC6B	29.92	1.88	0.22	8.67	4.17E-18	5.04E-16
CXCR4	15.51	5.83	0.68	8.62	6.51E-18	7.75E-16

Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
SUPT6H	50.14	1.46	0.17	8.59	8.94E-18	1.05E-15
TTC23	22.59	2.39	0.28	8.54	1.32E-17	1.53E-15
ST6GAL1	70.93	-1.36	0.16	-8.45	3.00E-17	3.45E-15
LOC100188947	21.78	2.26	0.27	8.44	3.15E-17	3.58E-15
ZRANB2	93.18	1.13	0.14	8.33	8.33E-17	9.35E-15
C1D	39.32	1.63	0.20	8.31	9.75E-17	1.08E-14
PPP1R18	77.02	-1.38	0.17	-8.26	1.42E-16	1.56E-14
SNAI1	12.65	5.54	0.68	8.13	4.34E-16	4.60E-14
RBM25	98.59	1.13	0.14	8.04	9.11E-16	9.45E-14
CWC15	79.29	1.21	0.15	7.97	1.62E-15	1.66E-13
RHOBTB1	14.62	2.61	0.33	7.92	2.33E-15	2.36E-13
YPEL5	25.81	1.86	0.24	7.90	2.77E-15	2.79E-13
CLK1	19.40	2.30	0.30	7.79	6.54E-15	6.44E-13
PSMD9	98.75	1.05	0.13	7.77	7.59E-15	7.40E-13
DBR1	14.99	2.55	0.33	7.63	2.31E-14	2.23E-12
ZSCAN5A	15.28	2.53	0.33	7.60	2.87E-14	2.72E-12
ACAP2	18.44	2.06	0.27	7.60	2.95E-14	2.77E-12
YTHDC1	49.90	1.35	0.18	7.54	4.66E-14	4.33E-12
ALG13	36.76	1.44	0.19	7.47	7.89E-14	7.25E-12
ATF3	13.69	2.77	0.37	7.47	8.12E-14	7.39E-12
PNRC1	13.55	2.49	0.33	7.43	1.08E-13	9.70E-12
SHC1	54.31	1.20	0.16	7.42	1.20E-13	1.07E-11
MEX3A	74.86	-1.25	0.17	-7.30	2.90E-13	2.54E-11
PANX2	19.32	2.09	0.29	7.26	3.78E-13	3.28E-11
ALDH9A1	46.05	1.27	0.18	7.21	5.52E-13	4.75E-11
KIAA1551	14.82	2.31	0.32	7.20	5.84E-13	4.98E-11
SERTAD1	17.17	2.04	0.29	7.09	1.37E-12	1.16E-10
GLUL	57.83	1.11	0.16	7.09	1.39E-12	1.17E-10
SIRT1	26.62	1.60	0.23	7.02	2.19E-12	1.82E-10
SAMD8	8.41	4.23	0.60	7.01	2.30E-12	1.90E-10
DYNC2H1	18.84	2.03	0.29	6.99	2.66E-12	2.17E-10
DUSP18	10.48	3.61	0.52	6.94	3.97E-12	3.22E-10
BIRC2	22.55	1.72	0.25	6.91	4.80E-12	3.86E-10
MELK	43.85	1.32	0.19	6.89	5.59E-12	4.46E-10
EFNB1	39.53	-1.48	0.22	-6.85	7.48E-12	5.87E-10
RIT2	8.91	7.17	1.05	6.83	8.58E-12	6.63E-10
MPHOSPH8	50.27	1.19	0.18	6.66	2.69E-11	2.04E-09
C20orf203	6.74	4.32	0.65	6.61	3.97E-11	3.00E-09



Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
ZCCHC10	18.53	1.92	0.29	6.57	5.13E-11	3.81E-09
SHISA3	73.52	-1.00	0.15	-6.54	6.12E-11	4.51E-09
PRRG4	8.04	3.77	0.58	6.52	7.03E-11	5.10E-09
SAPCD2	70.23	-1.15	0.18	-6.50	7.84E-11	5.64E-09
HHLA2	36.67	1.25	0.19	6.50	8.26E-11	5.86E-09
ZBTB24	17.44	1.81	0.28	6.50	8.21E-11	5.86E-09
FAM58A	46.70	1.13	0.17	6.49	8.34E-11	5.87E-09
PRAMEF11	8.00	5.01	0.77	6.49	8.61E-11	6.02E-09
LOC100507557	4.20	6.94	1.08	6.43	1.28E-10	8.90E-09
PRSS23	8.83	2.91	0.45	6.43	1.29E-10	8.90E-09
MAD2L1BP	17.98	1.69	0.26	6.42	1.36E-10	9.34E-09
MCM9	9.33	2.71	0.42	6.40	1.53E-10	1.03E-08
PRELP	14.09	2.10	0.33	6.37	1.91E-10	1.28E-08
TRIM23	9.52	2.51	0.40	6.34	2.24E-10	1.49E-08
IER5	19.30	1.77	0.28	6.33	2.52E-10	1.67E-08
PIM1	14.23	1.87	0.30	6.31	2.76E-10	1.81E-08
GLIS2	31.39	-1.59	0.25	-6.29	3.26E-10	2.10E-08
NAT8L	48.61	-1.30	0.21	-6.29	3.27E-10	2.10E-08
NKIRAS1	34.49	1.33	0.21	6.25	4.23E-10	2.71E-08
NRDE2	8.26	3.17	0.51	6.22	4.96E-10	3.11E-08
HDAC9	39.67	-1.34	0.22	-6.21	5.27E-10	3.28E-08
TIPARP	15.01	1.88	0.31	6.12	9.12E-10	5.54E-08
NUDT10	16.60	1.76	0.29	6.11	9.94E-10	5.99E-08
SOX12	30.90	-1.48	0.24	-6.11	1.02E-09	6.11E-08
ZNF705A	6.03	6.47	1.06	6.08	1.20E-09	7.14E-08
TC2N	9.35	2.24	0.37	6.07	1.30E-09	7.68E-08
ZC3H4	47.94	-1.16	0.19	-6.06	1.35E-09	7.92E-08
C2orf69	46.21	1.14	0.19	6.06	1.40E-09	8.19E-08
PHF23	29.92	-1.51	0.25	-6.04	1.59E-09	9.11E-08
ART3	5.63	4.71	0.78	6.01	1.88E-09	1.08E-07
KLHL15	9.45	2.41	0.41	5.94	2.77E-09	1.56E-07
MBD3L2	3.72	6.46	1.09	5.93	3.03E-09	1.69E-07
PRAMEF5	5.05	3.99	0.68	5.91	3.42E-09	1.87E-07
ACSL3	23.34	1.42	0.24	5.90	3.71E-09	2.01E-07
LOC152217	426.72	1.19	0.20	5.89	3.79E-09	2.05E-07
SOGA1	25.04	-1.72	0.29	-5.86	4.58E-09	2.41E-07
ELL2	17.64	1.62	0.28	5.85	4.77E-09	2.50E-07
AHCYL2	7.69	2.49	0.43	5.79	7.16E-09	3.71E-07

Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
HES7	21.19	-1.74	0.30	-5.76	8.39E-09	4.33E-07
SGCG	3.11	6.35	1.10	5.75	8.82E-09	4.53E-07
PELI2	8.26	2.34	0.41	5.75	8.93E-09	4.56E-07
NRBF2	11.97	2.06	0.36	5.74	9.64E-09	4.89E-07
PRDM7	7.39	3.00	0.52	5.73	9.80E-09	4.95E-07
EPB41L2	40.95	1.09	0.19	5.72	1.05E-08	5.25E-07
ISOC1	15.26	1.70	0.30	5.72	1.05E-08	5.25E-07
PLEKHG3	30.00	-1.52	0.27	-5.70	1.19E-08	5.88E-07
OXR1	16.65	1.74	0.31	5.69	1.29E-08	6.35E-07
ZNF10	11.15	2.02	0.36	5.66	1.49E-08	7.31E-07
MTAP	24.72	1.32	0.23	5.63	1.77E-08	8.56E-07
GOLGB1	18.99	1.54	0.27	5.63	1.78E-08	8.61E-07
LINC00652	5.43	3.38	0.60	5.59	2.28E-08	1.09E-06
KIN	23.66	1.33	0.24	5.58	2.42E-08	1.15E-06
SIX5	28.74	-1.43	0.26	-5.57	2.56E-08	1.20E-06
PAPOLG	27.90	1.30	0.23	5.56	2.67E-08	1.25E-06
FAM155B	22.83	-1.74	0.31	-5.54	3.01E-08	1.40E-06
RNF213	23.65	1.37	0.25	5.47	4.53E-08	2.06E-06
TESK2	5.55	3.44	0.63	5.46	4.63E-08	2.10E-06
OSR2	7.12	3.02	0.55	5.46	4.81E-08	2.17E-06
KITLG	9.18	2.21	0.41	5.45	5.10E-08	2.29E-06
STIL	15.20	1.67	0.31	5.43	5.72E-08	2.53E-06
KDM5A	20.14	1.38	0.25	5.41	6.28E-08	2.77E-06
PPP1R9B	23.55	-1.61	0.30	-5.40	6.66E-08	2.92E-06
SNIP1	16.23	1.56	0.29	5.40	6.71E-08	2.93E-06
KIAA0040	7.19	2.30	0.43	5.36	8.50E-08	3.68E-06
RARG	20.40	-1.60	0.30	-5.36	8.54E-08	3.68E-06
JUN	6.88	2.53	0.47	5.35	8.94E-08	3.82E-06
CASP6	14.88	1.69	0.32	5.33	1.00E-07	4.25E-06
ZSCAN5B	3.42	5.80	1.09	5.32	1.04E-07	4.38E-06
FAM57A	48.24	-1.04	0.20	-5.30	1.14E-07	4.76E-06
NR2F2	22.61	-1.51	0.29	-5.30	1.17E-07	4.87E-06
NR2F6	41.02	-1.11	0.21	-5.30	1.18E-07	4.87E-06
SRPK3	3.80	4.27	0.81	5.29	1.22E-07	5.04E-06
SGK1	5.78	2.96	0.56	5.29	1.23E-07	5.07E-06
DAB2	7.59	2.17	0.41	5.28	1.32E-07	5.38E-06
LOC256021	4.84	5.58	1.06	5.27	1.36E-07	5.56E-06
N4BP2L2	19.05	1.42	0.27	5.27	1.38E-07	5.61E-06



Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
ITPRIPL2	29.13	-1.32	0.25	-5.27	1.40E-07	5.64E-06
MED15	24.18	1.34	0.25	5.26	1.41E-07	5.66E-06
BCAS2	36.36	1.02	0.19	5.25	1.50E-07	5.99E-06
NFYA	8.17	2.33	0.44	5.25	1.52E-07	6.07E-06
RAB11FIP1	37.80	1.06	0.20	5.22	1.80E-07	7.10E-06
TRIM35	25.50	1.28	0.24	5.22	1.84E-07	7.20E-06
GABPB1-AS1	19.09	1.40	0.27	5.21	1.87E-07	7.32E-06
ATXN1L	10.29	2.02	0.39	5.21	1.88E-07	7.32E-06
RILPL1	24.29	-1.38	0.27	-5.19	2.06E-07	7.95E-06
GNG11	10.13	2.19	0.42	5.17	2.35E-07	8.99E-06
LOC400027	28.50	1.10	0.21	5.13	2.89E-07	1.10E-05
ARL4C	44.75	-1.01	0.20	-5.12	3.06E-07	1.15E-05
ASH1L-AS1	8.76	2.06	0.40	5.09	3.53E-07	1.32E-05
TOPORS	17.01	1.42	0.28	5.06	4.09E-07	1.52E-05
SLC35E4	8.94	2.16	0.43	5.06	4.19E-07	1.54E-05
ITGB8	4.94	3.29	0.65	5.05	4.37E-07	1.60E-05
SNX33	30.35	-1.23	0.24	-5.05	4.41E-07	1.61E-05
ATG14	18.38	1.38	0.27	5.04	4.66E-07	1.70E-05
LGALS3	4.93	3.45	0.69	5.03	4.92E-07	1.78E-05
KDM4E	2.61	5.66	1.13	5.02	5.23E-07	1.88E-05
MIDN	26.26	-1.26	0.25	-5.01	5.56E-07	1.99E-05
IRX5	18.35	1.32	0.26	5.00	5.87E-07	2.10E-05
C3	2.78	5.50	1.10	4.99	5.97E-07	2.12E-05
TRIM47	26.57	-1.28	0.26	-4.99	5.99E-07	2.13E-05
B3GNT2	4.55	3.42	0.69	4.99	6.06E-07	2.13E-05
PHOX2B	2.39	5.67	1.14	4.99	6.06E-07	2.13E-05
TGIF2	21.56	-1.41	0.28	-4.98	6.20E-07	2.18E-05
ASF1A	31.54	1.01	0.20	4.97	6.56E-07	2.29E-05
IGDCC3	18.29	-1.55	0.31	-4.97	6.78E-07	2.36E-05
NARG2	22.32	1.29	0.26	4.96	7.05E-07	2.45E-05
GSC	4.24	3.70	0.75	4.95	7.43E-07	2.57E-05
DDN	18.40	-1.59	0.32	-4.92	8.50E-07	2.91E-05
ZNF280A	3.27	4.05	0.82	4.92	8.47E-07	2.91E-05
ACSM2A	2.86	5.41	1.10	4.92	8.62E-07	2.94E-05
TGS1	28.36	1.07	0.22	4.92	8.70E-07	2.96E-05
SETD1B	30.15	-1.33	0.27	-4.91	9.22E-07	3.08E-05
HARS2	18.47	1.32	0.27	4.90	9.41E-07	3.13E-05
KIAA2018	5.02	2.77	0.57	4.90	9.58E-07	3.18E-05

Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
CLP1	29.80	1.04	0.21	4.90	9.79E-07	3.23E-05
STK17B	10.16	1.79	0.37	4.89	9.93E-07	3.27E-05
PTCSC3	4.01	5.20	1.06	4.88	1.06E-06	3.45E-05
FOSL1	17.87	-1.72	0.35	-4.87	1.10E-06	3.57E-05
LOC644656	7.63	2.11	0.43	4.87	1.14E-06	3.71E-05
MED26	8.95	2.19	0.45	4.86	1.15E-06	3.72E-05
STX6	24.08	1.11	0.23	4.85	1.21E-06	3.88E-05
HOXA11	19.29	-1.46	0.30	-4.85	1.25E-06	3.98E-05
C16orf55	13.90	1.41	0.29	4.84	1.29E-06	4.10E-05
UTP23	21.58	1.24	0.26	4.84	1.29E-06	4.10E-05
FAM195A	34.44	-1.03	0.21	-4.82	1.41E-06	4.46E-05
ATF7IP	27.84	1.09	0.23	4.82	1.41E-06	4.46E-05
OTUD3	18.71	1.39	0.29	4.81	1.52E-06	4.75E-05
MBD3L5	3.30	4.45	0.93	4.80	1.55E-06	4.85E-05
DNTT	1.66	5.58	1.17	4.79	1.69E-06	5.24E-05
C12orf50	3.42	5.14	1.08	4.78	1.73E-06	5.33E-05
GADD45A	5.73	2.41	0.50	4.78	1.74E-06	5.35E-05
EFNA2	10.19	-2.02	0.42	-4.78	1.75E-06	5.37E-05
RBM5	31.71	1.03	0.22	4.76	1.89E-06	5.77E-05
BAMBI	7.96	2.02	0.42	4.76	1.90E-06	5.78E-05
PANX1	9.53	1.70	0.36	4.76	1.91E-06	5.81E-05
DLC1	7.47	2.15	0.45	4.75	2.06E-06	6.17E-05
KCNQ10T1	29.91	1.15	0.24	4.74	2.18E-06	6.49E-05
SCAPER	7.02	2.17	0.46	4.73	2.28E-06	6.78E-05
SIKE1	26.60	1.13	0.24	4.71	2.45E-06	7.23E-05
E2F2	24.80	-1.32	0.28	-4.71	2.54E-06	7.45E-05
CITED4	34.75	-1.03	0.22	-4.70	2.62E-06	7.66E-05
ZIM3	1.49	5.53	1.18	4.69	2.70E-06	7.88E-05
BEND4	31.73	-1.05	0.22	-4.69	2.75E-06	7.99E-05
NAA35	11.43	1.65	0.35	4.69	2.78E-06	8.04E-05
FBXL12	26.06	1.10	0.23	4.69	2.80E-06	8.05E-05
SAMD10	21.23	-1.42	0.30	-4.68	2.80E-06	8.05E-05
WHAMM	20.80	1.26	0.27	4.68	2.91E-06	8.33E-05
BIK	26.05	1.06	0.23	4.67	3.02E-06	8.61E-05
BHLHE22	3.50	5.13	1.10	4.66	3.18E-06	9.06E-05
FAM78A	7.27	-2.91	0.63	-4.65	3.29E-06	9.35E-05
SOX13	21.63	-1.38	0.30	-4.65	3.33E-06	9.39E-05
SCG3	3.48	5.00	1.08	4.65	3.39E-06	9.52E-05



Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
TMED7	15.33	1.31	0.28	4.64	3.57E-06	1.00E-04
ORAI1	36.53	-1.07	0.23	-4.62	3.84E-06	1.07E-04
TAF4B	2.89	5.12	1.11	4.62	3.85E-06	1.07E-04
TMEM185A	51.31	1.20	0.26	4.61	3.97E-06	1.10E-04
ZNF776	26.87	1.03	0.22	4.61	4.07E-06	1.12E-04
GRAMD1C	3.26	4.96	1.08	4.61	4.12E-06	1.13E-04
FBXO33	17.67	1.32	0.29	4.60	4.31E-06	1.17E-04
USP38	11.59	1.63	0.36	4.58	4.62E-06	1.25E-04
PLSCR1	8.81	1.89	0.41	4.58	4.74E-06	1.27E-04
BTG1	11.90	1.54	0.34	4.57	4.84E-06	1.29E-04
DUSP16	19.08	-1.35	0.29	-4.57	4.99E-06	1.33E-04
MKRN9P	1.59	5.36	1.17	4.56	5.00E-06	1.33E-04
HEXIM1	9.97	1.72	0.38	4.56	5.10E-06	1.35E-04
FERMT2	18.50	1.24	0.27	4.55	5.25E-06	1.39E-04
LOC401557	1.45	5.36	1.18	4.55	5.43E-06	1.43E-04
C15orf60	3.57	4.91	1.08	4.53	5.84E-06	1.53E-04
NT5DC3	11.48	1.49	0.33	4.53	5.87E-06	1.53E-04
KLF17	2.77	4.90	1.08	4.52	6.12E-06	1.59E-04
CCNL1	8.85	1.83	0.41	4.50	6.67E-06	1.73E-04
VEPH1	1.41	5.28	1.17	4.50	6.79E-06	1.75E-04
TGIF1	39.96	-1.03	0.23	-4.50	6.91E-06	1.78E-04
GUSBP1	10.33	1.60	0.36	4.48	7.56E-06	1.94E-04
CEBPB	15.31	-1.68	0.38	-4.47	7.71E-06	1.97E-04
ARRDC3	24.43	1.02	0.23	4.47	7.83E-06	2.00E-04
LAMTOR3	21.08	1.23	0.28	4.47	7.93E-06	2.02E-04
C12orf43	18.14	1.22	0.27	4.46	8.12E-06	2.05E-04
TRIM49	1.54	5.23	1.17	4.46	8.12E-06	2.05E-04
RWDD1	33.11	1.04	0.23	4.46	8.16E-06	2.05E-04
ZNF134	7.86	1.85	0.42	4.45	8.58E-06	2.16E-04
LDB2	30.13	-1.10	0.25	-4.45	8.75E-06	2.19E-04
EN2	19.03	-1.47	0.33	-4.44	9.14E-06	2.27E-04
SRSF5	19.21	1.14	0.26	4.43	9.64E-06	2.38E-04
CDO1	12.71	1.57	0.35	4.42	9.74E-06	2.40E-04
MCMDC2	1.33	5.23	1.18	4.42	9.72E-06	2.40E-04
PLCL2	20.31	-1.31	0.30	-4.42	9.83E-06	2.41E-04
ATXN2	30.87	-1.06	0.24	-4.42	9.90E-06	2.42E-04
PLK2	4.67	2.65	0.60	4.42	1.01E-05	2.46E-04
CCDC174	13.46	1.43	0.32	4.41	1.01E-05	2.47E-04

Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
RADIL	17.71	1.26	0.29	4.41	1.03E-05	2.49E-04
CNOT8	10.23	1.54	0.35	4.41	1.04E-05	2.51E-04
ANK3	15.26	1.34	0.30	4.40	1.06E-05	2.53E-04
GRPEL2	24.82	1.08	0.25	4.40	1.07E-05	2.56E-04
H19	25.80	-1.26	0.29	-4.38	1.17E-05	2.79E-04
LRFN1	17.04	-1.45	0.33	-4.37	1.22E-05	2.87E-04
CROT	12.10	1.49	0.34	4.37	1.23E-05	2.90E-04
LOC285540	2.67	4.77	1.09	4.37	1.24E-05	2.91E-04
TAL1	9.88	-1.88	0.43	-4.37	1.26E-05	2.95E-04
LOC100130557	4.38	2.48	0.57	4.37	1.26E-05	2.95E-04
DNAJC3	12.24	1.38	0.32	4.36	1.28E-05	2.99E-04
SH3KBP1	24.89	1.09	0.25	4.34	1.44E-05	3.35E-04
BTG2	5.47	-3.42	0.79	-4.32	1.58E-05	3.64E-04
HIPK1	9.69	1.62	0.37	4.31	1.62E-05	3.73E-04
GNA14	2.27	4.91	1.14	4.30	1.68E-05	3.82E-04
PPIL3	11.84	1.52	0.35	4.30	1.73E-05	3.91E-04
TLX1	21.58	-1.21	0.28	-4.30	1.73E-05	3.91E-04
TNS3	28.97	-1.08	0.25	-4.29	1.75E-05	3.94E-04
ILF3-AS1	18.45	1.16	0.27	4.29	1.76E-05	3.96E-04
GPBAR1	1.89	4.87	1.14	4.28	1.84E-05	4.13E-04
C6orf147	6.80	1.96	0.46	4.28	1.87E-05	4.17E-04
LINC00310	2.50	3.62	0.85	4.27	1.94E-05	4.32E-04
CASP10	24.94	1.03	0.24	4.27	1.99E-05	4.41E-04
AJUBA	18.32	-1.34	0.31	-4.27	2.00E-05	4.42E-04
SALL2	17.01	-1.41	0.33	-4.25	2.16E-05	4.73E-04
NBPF3	25.18	-1.08	0.25	-4.24	2.19E-05	4.78E-04
NINJ1	28.87	-1.03	0.24	-4.23	2.30E-05	5.00E-04
KAT6B	6.89	1.94	0.46	4.21	2.55E-05	5.47E-04
C2CD4B	1.86	4.81	1.14	4.21	2.57E-05	5.51E-04
OTX1	13.06	1.45	0.35	4.20	2.69E-05	5.75E-04
RGMA	7.97	-2.21	0.53	-4.19	2.74E-05	5.80E-04
SLU7	19.24	1.17	0.28	4.19	2.79E-05	5.89E-04
FAM124A	14.42	-1.52	0.36	-4.18	2.97E-05	6.26E-04
BCL9	25.23	-1.05	0.25	-4.17	3.02E-05	6.34E-04
TMUB1	17.58	-1.27	0.30	-4.17	3.06E-05	6.38E-04
HCG27	3.93	2.80	0.68	4.13	3.57E-05	7.36E-04
TEFM	7.31	1.82	0.44	4.13	3.65E-05	7.52E-04
IMPACT	22.03	1.03	0.25	4.13	3.68E-05	7.56E-04



Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
KIAA0317	20.97	1.12	0.27	4.12	3.73E-05	7.63E-04
TRIM36	4.65	2.26	0.55	4.12	3.81E-05	7.77E-04
TRIM24	42.40	-1.24	0.30	-4.11	3.90E-05	7.94E-04
ELOVL1	17.94	-1.23	0.30	-4.11	3.96E-05	8.05E-04
NAV1	15.80	-1.46	0.36	-4.10	4.11E-05	8.32E-04
CELSR2	19.58	-1.18	0.29	-4.09	4.28E-05	8.63E-04
MLLT6	30.21	-1.01	0.25	-4.09	4.39E-05	8.81E-04
ZNF256	5.92	1.80	0.44	4.08	4.59E-05	9.19E-04
LOC256880	1.72	4.69	1.15	4.07	4.65E-05	9.28E-04
STK11	21.09	-1.15	0.28	-4.07	4.64E-05	9.28E-04
PHF15	9.73	-1.98	0.49	-4.07	4.78E-05	9.50E-04
ANKRD34A	3.06	-4.57	1.12	-4.06	4.89E-05	9.65E-04
PDGFRA	13.39	1.28	0.31	4.06	4.90E-05	9.65E-04
ZNF789	5.98	1.96	0.48	4.06	4.90E-05	9.65E-04
KIAA1210	1.07	4.97	1.23	4.05	5.11E-05	1.00E-03
TLE3	24.43	-1.01	0.25	-4.04	5.24E-05	1.02E-03
SHB	12.39	-1.76	0.43	-4.04	5.26E-05	1.02E-03
RYK	12.63	1.35	0.33	4.04	5.40E-05	1.05E-03
KIAA0907	11.57	1.36	0.34	4.03	5.46E-05	1.05E-03
SERTAD3	6.58	-2.25	0.56	-4.03	5.67E-05	1.09E-03
ZSCAN16	16.55	-1.30	0.32	-4.03	5.67E-05	1.09E-03
KIAA0020	16.54	1.21	0.30	4.02	5.71E-05	1.09E-03
ELF4	30.76	-1.06	0.26	-4.02	5.76E-05	1.10E-03
DMRTA2	16.89	-1.24	0.31	-4.00	6.31E-05	1.20E-03
DHRS3	17.09	-1.24	0.31	-3.99	6.58E-05	1.24E-03
NOG	10.34	-1.62	0.41	-3.98	6.76E-05	1.27E-03
FAM174B	11.87	-1.60	0.40	-3.97	7.33E-05	1.37E-03
C20orf112	9.87	-1.57	0.40	-3.96	7.34E-05	1.37E-03
DIS3	20.01	1.05	0.26	3.96	7.51E-05	1.39E-03
SPSB4	9.01	-1.61	0.41	-3.95	7.87E-05	1.46E-03
BBS4	12.40	1.40	0.36	3.94	8.16E-05	1.50E-03
SUSD2	2.24	3.81	0.97	3.93	8.40E-05	1.54E-03
FADD	23.05	-1.02	0.26	-3.92	8.71E-05	1.59E-03
CDKN1A	8.39	1.60	0.41	3.92	8.78E-05	1.60E-03
ZMAT3	7.03	1.95	0.50	3.91	9.24E-05	1.67E-03
NPHP3	4.54	2.27	0.58	3.91	9.31E-05	1.68E-03
SFTPB	3.91	2.60	0.67	3.89	9.89E-05	1.77E-03
CITED2	3.17	2.98	0.77	3.87	1.07E-04	1.90E-03

Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
IQCE	5.12	-2.84	0.73	-3.87	1.08E-04	1.91E-03
BBC3	7.02	-2.07	0.54	-3.87	1.10E-04	1.94E-03
TTC14	8.96	1.44	0.37	3.86	1.13E-04	1.97E-03
CTH	9.14	1.42	0.37	3.86	1.14E-04	1.99E-03
RTN4R	13.12	-1.51	0.39	-3.84	1.22E-04	2.11E-03
SAMD5	8.57	1.52	0.40	3.84	1.24E-04	2.14E-03
STK31	1.32	4.54	1.18	3.84	1.24E-04	2.14E-03
LOC220729	6.73	1.58	0.41	3.83	1.29E-04	2.21E-03
EPC1	20.23	-1.12	0.29	-3.81	1.38E-04	2.34E-03
DDX20	15.53	1.08	0.28	3.81	1.39E-04	2.35E-03
ERF	10.51	-1.72	0.46	-3.79	1.52E-04	2.55E-03
NAB2	5.07	-3.42	0.91	-3.77	1.62E-04	2.69E-03
TRIM54	19.77	-1.07	0.28	-3.76	1.67E-04	2.76E-03
PLEKHO1	11.55	-1.43	0.38	-3.73	1.90E-04	3.09E-03
DDIT4	5.07	-3.38	0.91	-3.73	1.90E-04	3.09E-03
EPN2	16.45	1.05	0.28	3.71	2.10E-04	3.36E-03
SH3GL2	2.78	3.54	0.96	3.70	2.14E-04	3.41E-03
RLF	9.26	1.47	0.40	3.69	2.22E-04	3.54E-03
ARHGAP42	3.30	2.56	0.69	3.69	2.24E-04	3.56E-03
CHKA	20.89	-1.04	0.28	-3.68	2.29E-04	3.62E-03
C9orf66	1.87	4.30	1.17	3.68	2.35E-04	3.70E-03
FAM120C	13.23	1.14	0.31	3.68	2.36E-04	3.71E-03
SIPA1L2	12.33	1.23	0.33	3.67	2.39E-04	3.75E-03
SETD5-AS1	12.66	1.20	0.33	3.67	2.41E-04	3.77E-03
TNFRSF10D	8.68	1.46	0.40	3.67	2.42E-04	3.78E-03
BCOR	14.20	-1.16	0.32	-3.67	2.47E-04	3.83E-03
DDX58	5.49	1.91	0.52	3.65	2.59E-04	4.00E-03
CDKN2B	3.51	2.38	0.65	3.65	2.61E-04	4.03E-03
THOC5	11.96	1.17	0.32	3.65	2.62E-04	4.03E-03
PPP3CC	10.16	-1.72	0.47	-3.65	2.65E-04	4.07E-03
CHST3	9.81	-1.69	0.46	-3.64	2.69E-04	4.12E-03
FAM189A2	2.33	2.90	0.80	3.64	2.70E-04	4.13E-03
GIT1	23.57	-1.02	0.28	-3.64	2.70E-04	4.13E-03
LINC00115	5.14	1.82	0.50	3.64	2.72E-04	4.14E-03
PDPK1	14.78	1.18	0.32	3.64	2.75E-04	4.19E-03
CNNM4	8.09	1.39	0.38	3.63	2.80E-04	4.24E-03
SGMS1	8.31	1.47	0.40	3.63	2.80E-04	4.24E-03
ARL14EP	9.59	1.36	0.38	3.62	2.94E-04	4.40E-03



Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
CD97	11.12	-1.46	0.40	-3.61	3.02E-04	4.49E-03
C5orf44	8.94	1.38	0.38	3.61	3.09E-04	4.57E-03
PDSS1	2.33	3.10	0.86	3.61	3.10E-04	4.57E-03
IRX3	11.61	-1.40	0.39	-3.60	3.16E-04	4.63E-03
TMEM189	12.62	-1.33	0.37	-3.60	3.18E-04	4.65E-03
DKFZP434I0714	6.90	1.60	0.45	3.60	3.23E-04	4.70E-03
MTF1	10.79	1.25	0.35	3.60	3.23E-04	4.70E-03
TRIM43	1.95	4.07	1.13	3.60	3.23E-04	4.70E-03
GSE1	19.25	-1.09	0.30	-3.59	3.26E-04	4.72E-03
GLMN	5.35	1.87	0.52	3.58	3.45E-04	4.98E-03
CENPI	4.92	1.83	0.51	3.57	3.53E-04	5.07E-03
SYNJ1	3.94	2.08	0.58	3.57	3.56E-04	5.10E-03
SGSH	8.51	1.42	0.40	3.57	3.58E-04	5.12E-03
RBM7	4.21	2.06	0.58	3.57	3.59E-04	5.12E-03
TBX5	13.12	-1.29	0.36	-3.56	3.65E-04	5.19E-03
ZBTB10	4.24	2.00	0.56	3.55	3.82E-04	5.41E-03
TP53BP2	12.27	1.16	0.33	3.55	3.90E-04	5.51E-03
PACRGL	5.03	1.69	0.48	3.53	4.14E-04	5.78E-03
TMTC1	3.73	2.30	0.65	3.53	4.19E-04	5.82E-03
KIAA1429	12.70	1.16	0.33	3.52	4.27E-04	5.93E-03
DSEL	8.39	1.46	0.42	3.51	4.52E-04	6.20E-03
PTCH1	8.24	-1.55	0.44	-3.50	4.62E-04	6.33E-03
LMO4	15.74	1.10	0.32	3.50	4.72E-04	6.42E-03
ZNF48	15.07	-1.18	0.34	-3.48	4.93E-04	6.67E-03
TPRN	13.03	-1.23	0.35	-3.48	4.95E-04	6.68E-03
CCDC85B	11.03	-1.29	0.37	-3.48	4.99E-04	6.72E-03
TIGD5	9.69	-1.63	0.47	-3.48	5.04E-04	6.77E-03
INSR	19.55	-1.03	0.30	-3.47	5.12E-04	6.86E-03
ZNF473	13.03	1.07	0.31	3.47	5.12E-04	6.86E-03
ZNF827	17.64	1.00	0.29	3.47	5.27E-04	7.00E-03
LOC100505659	2.12	3.08	0.89	3.47	5.29E-04	7.02E-03
MT-ND2	2153.05	1.18	0.34	3.46	5.33E-04	7.03E-03
SFT2D2	7.82	1.44	0.41	3.46	5.33E-04	7.03E-03
WDR89	15.53	-1.15	0.33	-3.46	5.34E-04	7.04E-03
ZNF674-AS1	13.75	1.10	0.32	3.45	5.57E-04	7.30E-03
MED31	6.70	1.52	0.44	3.45	5.63E-04	7.37E-03
FAM188A	9.60	1.32	0.38	3.43	5.93E-04	7.68E-03
LOC100507373	2.63	2.60	0.76	3.43	5.93E-04	7.68E-03

Table S2 continued

Gene.Symbol	baseMean	log2FC	lfcSE	stat	pvalue	padj
CETN3	9.38	1.34	0.39	3.43	6.15E-04	7.91E-03
ST3GAL1	11.35	-1.43	0.42	-3.42	6.31E-04	8.09E-03
GAS2L1	16.34	-1.13	0.33	-3.41	6.59E-04	8.38E-03
LEF1	12.74	-1.40	0.41	-3.40	6.63E-04	8.40E-03
MAN1A1	4.19	1.96	0.58	3.40	6.72E-04	8.50E-03
JAG2	15.56	-1.10	0.33	-3.37	7.50E-04	9.31E-03
LRFN4	13.80	-1.11	0.33	-3.35	8.05E-04	9.92E-03

Table S3: TF perturbations followed by expressionFiltered for: HUMAN, upregulated genes (UP), P-value ≤ 0.01

Input: differentially upregulated genes (358) in DIE_8.5h dataset

Term	Overlap	P.value	Adj. P.value	Odds Ratio	Comb. Score
DUX4_OE_GSE33799_CREEDSID_GENE_1429	113/451	5.1E-101	9.9E-98	14.36	3315.88
PAX5_OE_GSE44244_CREEDSID_GENE_575	21/407	1.1E-05	4.2E-03	2.96	33.84
PAX5_OE_GSE44244_CREEDSID_GENE_574	24/541	3.1E-05	1.0E-02	2.54	26.41
NFKB1_INACTIVATION_GSE20667_CREEDSID_GENE_2575	17/309	3.3E-05	9.3E-03	3.15	32.50
HSF1_KD_GSE3697_CREEDSID_GENE_779	17/328	7.0E-05	1.7E-02	2.97	28.42
HSF1_KD_GSE3697_CREEDSID_GENE_778	17/331	7.8E-05	1.7E-02	2.94	27.83
TBX3_SHRNA_HFF_GSE76572_RNASEQ	17/364	2.4E-04	4.8E-02	2.68	22.27
JUNB_KD_FORESKIN_GSE63079_RNASEQ	15/313	4.2E-04	6.4E-02	2.75	21.31
TP63_KD_GSE20286_CREEDSID_GENE_2453	15/334	8.3E-04	9.6E-02	2.57	18.25
MYB_KD_GSE49286_CREEDSID_GENE_1842	13/268	9.1E-04	9.9E-02	2.78	19.47
KLF9_OE_GBM1A_GSE62212_RNASEQ	17/411	9.6E-04	9.9E-02	2.37	16.47
LEF1_KD_GSE42637_CREEDSID_GENE_1775	15/351	1.4E-03	1.3E-01	2.45	16.15
GATA2_OE_HESC_GSE57395_RNASEQ	19/505	1.5E-03	1.3E-01	2.17	14.02
NFKB1_INACTIVATION_GSE20667_CREEDSID_GENE_2577	11/220	1.7E-03	1.5E-01	2.87	18.20
MYB_KD_GSE49286_CREEDSID_GENE_1835	14/333	2.3E-03	1.5E-01	2.41	14.67
SETDB1_KO_HELA_GSE86813_RNASEQ	9/170	3.1E-03	2.0E-01	3.03	17.56
MYC_OE_MCF7_GSE101738_RNASEQ	16/441	4.9E-03	2.2E-01	2.08	11.06
FLI1_KD_GSE27524_CREEDSID_GENE_1612	12/289	5.0E-03	2.2E-01	2.38	12.60
MAF_OE_MACROPHAGE_GSE98368_RNASEQ	19/567	5.3E-03	2.3E-01	1.92	10.05
FLI1_KD_GSE27524_CREEDSID_GENE_1595	12/293	5.6E-03	2.3E-01	2.35	12.17
SREBF2_KD_GSE50588_CREEDSID_GENE_2823	12/294	5.7E-03	2.3E-01	2.34	12.07
FLI1_KD_GSE27524_CREEDSID_GENE_1611	12/297	6.2E-03	2.4E-01	2.32	11.77
FLI1_KD_GSE27524_CREEDSID_GENE_1596	13/335	6.3E-03	2.4E-01	2.22	11.29
NFKB1_INACTIVATION_GSE20667_CREEDSID_GENE_2574	11/262	6.5E-03	2.3E-01	2.41	12.10
HSF1_KD_GSE3697_CREEDSID_GENE_783	11/263	6.7E-03	2.3E-01	2.40	11.99
IRX6_SIRNA_BT549_GSE79586_RNASEQ	16/469	8.6E-03	2.5E-01	1.96	9.30

Table S4: TF perturbations followed by expressionFiltered for: HUMAN, downregulated genes (DOWN), P-value ≤ 0.01

Input: differentially downregulated genes (125) in DIE_8.5h dataset

Term	Overlap	P.value	Adj. P.value	Odds Ratio	Comb. Score
DUX4_OE_GSE33799_CREEDSID_GENE_1429	6/130	1.22E-04	0.027	7.823	70.489
MYCN_SHRNA_IMR575_GSE80397_12HR_RNASEQ	10/472	4.89E-04	0.064	3.591	27.372
TAL1_OE_HESC_GSE57395_RNASEQ	3/47	2.69E-03	0.175	10.819	64.039
NFXL1_KD_GSE23674_CREEDSID_GENE_2439	7/327	3.29E-03	0.208	3.628	20.746
DNMT1_INHIBITION_GSE45804_CREEDSID_GENE_2762	6/265	4.90E-03	0.234	3.838	20.411
IRF7_OE_GSE37828_CREEDSID_GENE_1146	5/195	6.03E-03	0.251	4.346	22.215
SNAI1_OE_GSE14773_CREEDSID_GENE_371	6/278	6.16E-03	0.246	3.658	18.620

Table S5: Shared differentially upregulated genes between 8.5h induced DIE cells and other datasets

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
ACAP2	yes	no	no	yes	yes	yes
ACSL3	no	no	no	yes	no	yes
ACSM2A	no	no	no	no	no	no
ADPGK	yes	yes	yes	yes	yes	yes
AHCYL2	no	no	no	yes	no	no
ALDH9A1	no	no	no	no	no	no
ALG13	yes	no	yes	no	no	no
ALPPL2	no	yes	yes	yes	yes	yes
ALYREF	no	yes	yes	yes	yes	yes
ANK3	no	no	no	no	no	no
ANXA5	no	no	no	no	no	no
ARHGAP42	no	no	yes	yes	yes	yes
ARL14EP	no	no	no	no	no	no
ARL4D	no	no	no	yes	no	yes
ARRDC3	no	no	no	yes	no	yes
ART3	yes	yes	yes	yes	yes	yes
ASF1A	no	no	no	no	no	no
ASHIL-AS1	no	no	yes	no	no	no
ATF3	yes	no	no	yes	no	yes
ATF7IP	no	no	no	yes	yes	yes
ATG14	no	no	no	yes	yes	yes
ATXN1L	yes	no	no	yes	yes	yes
AVPI1	yes	yes	yes	yes	yes	yes
B3GNT2	yes	no	yes	yes	yes	yes
BAMBI	yes	yes	no	yes	yes	yes
BBS4	no	no	no	no	no	no
BCAS2	no	no	yes	yes	yes	yes
BHLHE22	yes	no	no	yes	no	yes
BIK	no	yes	yes	yes	no	no
BIRC2	no	no	yes	no	no	no
BTG1	no	no	no	yes	no	no
C12orf43	yes	no	yes	no	no	no
C12orf50	yes	yes	yes	yes	yes	yes
C15orf60	yes	no	no	yes	no	no
C16orf55	no	no	no	no	yes	yes

Table S5 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
C1D	no	no	no	no	no	yes
C1orf63	yes	no	no	yes	yes	yes
C20orf203	no	yes	no	no	no	no
C21orf91	yes	no	yes	yes	yes	yes
C2CD4B	no	no	no	no	no	no
C2orf69	no	no	yes	no	no	yes
C3	no	no	no	no	no	no
C5orf44	no	no	no	no	no	no
C6orf147	no	no	no	no	no	no
C8orf33	yes	yes	yes	yes	yes	yes
C9orf66	no	no	no	no	no	no
CASP10	no	no	no	no	no	no
CASP6	yes	no	no	no	yes	yes
CCDC174	no	no	yes	no	no	no
CCNA1	yes	yes	yes	yes	yes	yes
CCNJ	yes	no	no	yes	no	yes
CCNL1	no	no	yes	yes	no	yes
CDKN1A	yes	no	no	no	no	no
CDKN2B	no	no	no	no	no	no
CDO1	yes	no	no	no	no	yes
CENPI	no	no	no	no	no	no
CETN3	no	no	no	no	no	no
CITED2	no	no	no	no	no	no
CLK1	yes	no	yes	yes	yes	yes
CLP1	yes	no	yes	yes	yes	yes
CNNM4	yes	yes	yes	yes	yes	yes
CNOT8	no	no	no	no	no	no
CROT	no	no	no	yes	no	yes
CTH	yes	no	no	yes	no	yes
CWC15	yes	yes	yes	no	no	no
CXCR4	yes	yes	no	yes	yes	yes
DAB2	no	no	no	no	no	no
DBR1	yes	yes	yes	yes	yes	yes
DDX10	no	no	no	yes	yes	yes
DDX20	no	no	no	yes	yes	yes
DDX58	no	no	no	no	no	no
DIO2	no	no	yes	yes	yes	no

Table S5 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
DIS3	no	no	no	yes	yes	yes
DKFZP434I0714	no	no	yes	no	no	no
DLC1	no	no	no	yes	no	yes
DNAJC3	no	no	no	yes	no	yes
DNTT	yes	yes	no	yes	no	yes
DPPA3	yes	yes	no	yes	yes	yes
DSEL	no	no	yes	no	no	no
DUSP18	no	no	no	no	no	no
DUXA	no	yes	yes	yes	yes	yes
DYNC2H1	yes	no	yes	yes	yes	no
DYX1C1	no	no	no	no	no	no
ELL2	no	no	no	yes	no	no
ELOF1	yes	no	yes	yes	yes	yes
EPB41L2	no	no	yes	no	no	no
EPN2	no	no	yes	no	no	no
EXOSC10	yes	yes	yes	yes	yes	yes
FAM120C	no	no	yes	no	no	no
FAM188A	no	no	no	yes	yes	yes
FAM189A2	no	yes	yes	yes	yes	yes
FAM58A	no	yes	yes	yes	yes	no
FBXL12	yes	yes	yes	yes	yes	yes
FBXO33	yes	yes	yes	yes	yes	yes
FERMT2	no	no	no	yes	no	no
GABPB1-AS1	no	no	no	no	no	no
GADD45A	no	no	no	yes	no	no
GLMN	yes	no	yes	yes	yes	yes
GLUL	no	no	yes	yes	no	no
GNA14	no	no	yes	no	yes	no
GNG11	no	no	yes	no	no	no
GOLGB1	yes	no	no	yes	yes	yes
GPBAR1	yes	yes	yes	yes	yes	yes
GRAMD1C	no	yes	yes	yes	yes	yes
GRPEL2	yes	no	no	yes	yes	yes
GSC	no	yes	yes	no	yes	no
GTF2F1	yes	yes	yes	yes	yes	yes
GUSBP1	no	no	no	yes	yes	yes
HARS2	no	no	no	no	yes	no

Table S5 continued

Gene	Geng	Rickard	Heuvel	Jaggnathan		
				iDUX4	enDUX4	vDUX4
HCG27	no	no	no	no	no	no
HEXIM1	no	no	no	yes	no	yes
HHLA2	no	no	no	no	no	no
HIPK1	no	no	no	no	no	yes
HNRNPF	yes	yes	yes	yes	yes	yes
HOXB2	yes	no	no	yes	yes	yes
IER5	no	no	no	no	no	no
ILF3-AS1	no	no	no	no	no	no
IMPACT	no	no	no	no	no	no
INO80C	yes	yes	yes	yes	yes	yes
IRX5	yes	no	no	yes	yes	no
ISOC1	yes	no	yes	yes	no	yes
ITGB8	no	no	no	no	no	no
JUN	no	no	no	yes	no	no
KAT6B	no	no	no	no	no	no
KCNQ10T1	no	no	no	no	no	no
KDM4E	no	yes	yes	yes	yes	yes
KDM5A	no	yes	yes	yes	yes	yes
KDM5B	yes	no	yes	yes	yes	yes
KHDC1	no	yes	yes	yes	yes	no
KHDC1L	yes	yes	yes	yes	yes	yes
KIAA0020	yes	no	no	yes	yes	yes
KIAA0040	no	yes	yes	yes	yes	yes
KIAA0317	no	no	no	no	yes	yes
KIAA0907	no	no	no	yes	no	no
KIAA1210	no	no	no	no	no	no
KIAA1429	yes	no	no	no	yes	yes
KIAA1551	no	yes	no	yes	yes	yes
KIAA2018	no	no	no	no	no	yes
KIN	no	no	no	yes	no	yes
KITLG	yes	no	no	yes	no	no
KLF17	yes	yes	yes	yes	yes	yes
KLHL15	yes	yes	yes	yes	yes	yes
LAMTOR3	no	no	no	no	no	no
LEUTX	no	yes	yes	yes	yes	yes
LGALS3	no	no	no	no	no	no
LINC00115	no	no	no	no	no	no

Table S5 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
LINC00310	no	no	no	no	no	no
LINC00493	no	no	no	no	no	no
LINC00633	no	no	no	no	no	no
LINC00652	no	no	no	no	no	no
LMO4	yes	no	no	yes	no	yes
LOC100130557	no	no	no	no	no	no
LOC100188947	no	no	no	no	no	no
LOC100216545	no	no	no	no	no	no
LOC100505659	no	no	no	no	no	no
LOC100507373	no	no	no	no	no	no
LOC100507557	no	no	no	no	no	no
LOC152217	no	no	no	no	no	no
LOC220729	no	no	no	no	no	no
LOC256021	no	no	no	no	no	no
LOC256880	no	no	no	no	no	no
LOC284551	no	no	no	no	no	no
LOC285540	no	no	no	no	no	no
LOC400027	no	no	no	no	no	no
LOC401557	no	no	no	no	no	no
LOC441081	yes	yes	no	no	no	no
LOC644656	no	no	no	no	no	no
MAD2L1BP	yes	no	yes	yes	yes	yes
MAN1A1	no	no	no	yes	no	yes
MBD3L2	yes	yes	yes	yes	yes	yes
MBD3L5	no	yes	yes	yes	yes	yes
MCM9	no	no	no	yes	no	no
MCMDC2	no	no	no	no	no	no
MED15	yes	no	no	no	no	no
MED26	yes	yes	yes	yes	yes	yes
MED31	yes	no	yes	yes	yes	yes
MELK	yes	yes	yes	no	yes	no
MFS11	no	no	no	yes	yes	yes
MGC21881	no	no	no	no	no	no
MKRN9P	no	yes	no	no	no	no
MMRN2	no	no	no	no	no	no
MPHOSPH8	no	no	yes	no	yes	yes
MRPL49	no	no	yes	yes	yes	yes

Table S5 continued

Gene	Geng	Rickard	Heuvel	Jaggnathan		
				iDUX4	enDUX4	vDUX4
MT-ND2	no	no	no	no	no	no
MTAP	yes	no	no	no	no	no
MTF1	no	no	no	yes	yes	yes
N4BP2L2	no	no	no	no	no	no
NAA35	no	no	no	no	yes	yes
NARG2	yes	no	no	no	no	no
NDEL1	yes	no	no	yes	no	yes
NFYA	yes	no	yes	yes	yes	yes
NKIRAS1	yes	no	yes	yes	yes	yes
NPHP3	no	no	no	no	no	no
NRBF2	yes	no	yes	yes	yes	yes
NRDE2	no	yes	yes	yes	yes	yes
NT5DC3	yes	no	yes	yes	yes	yes
NUDT10	no	yes	no	no	yes	no
NXF1	yes	yes	yes	yes	yes	yes
ODC1	yes	yes	yes	yes	yes	yes
OSR2	yes	no	no	yes	no	yes
OTUD3	no	no	no	yes	yes	yes
OTX1	no	no	no	no	no	no
OXR1	yes	no	yes	yes	no	yes
PACRGL	no	no	no	yes	no	no
PANX1	no	no	no	yes	no	no
PANX2	yes	yes	yes	yes	yes	yes
PAPOLG	no	no	no	yes	yes	yes
PDGFRA	no	no	no	yes	no	yes
PDPK1	no	no	yes	no	yes	no
PDSS1	yes	yes	no	yes	yes	yes
PELI2	yes	no	no	yes	no	yes
PHOX2B	no	no	no	no	no	no
PIM1	yes	no	yes	yes	no	yes
PLK2	no	no	no	yes	no	no
PLSCR1	no	no	no	no	no	no
PLXNB3	no	no	no	no	no	no
PNN	yes	no	no	yes	no	yes
PNP	no	yes	yes	yes	yes	yes
PNRC1	no	no	no	yes	no	no
PPIL3	no	no	no	no	yes	no

Table S5 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
PRAMEF1	yes	yes	yes	yes	yes	yes
PRAMEF11	yes	yes	yes	yes	yes	yes
PRAMEF12	yes	yes	yes	yes	yes	yes
PRAMEF5	yes	yes	yes	yes	yes	yes
PRDM7	no	no	no	no	no	no
PRELP	no	no	no	no	no	no
PRRG4	yes	no	yes	yes	yes	yes
PRSS23	no	no	yes	no	no	yes
PSMD9	no	no	yes	yes	yes	yes
PTCSC3	no	no	no	no	no	no
PTP4A1	yes	yes	yes	yes	yes	yes
PTPRJ	no	no	no	no	no	no
RAB11FIP1	yes	yes	yes	yes	yes	yes
RADIL	no	no	no	no	no	no
RBBP6	yes	yes	yes	yes	yes	yes
RBM25	yes	no	no	yes	yes	yes
RBM5	no	no	no	no	no	no
RBM7	yes	no	yes	yes	yes	yes
RFK	no	yes	yes	yes	yes	yes
RFPL1	yes	yes	yes	yes	yes	yes
RFPL2	yes	yes	yes	yes	yes	yes
RFPL4A	yes	yes	yes	yes	yes	yes
RFPL4B	yes	yes	yes	yes	yes	yes
RGS2	no	no	no	yes	no	no
RHOBTB1	yes	yes	no	yes	yes	yes
RIT2	no	no	no	no	no	no
RLF	yes	no	no	yes	yes	yes
RNF11	yes	no	no	yes	no	yes
RNF213	no	no	no	no	no	no
RWDD1	yes	no	no	no	no	yes
RYBP	yes	no	yes	yes	yes	yes
RYK	yes	no	yes	no	no	no
SAMD5	no	no	no	no	no	no
SAMD8	yes	yes	yes	yes	yes	yes
SCAPER	no	no	no	no	yes	no
SCG3	no	no	no	no	no	no
SERTAD1	yes	no	yes	yes	yes	yes

Table S5 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
SETD5-AS1	no	no	no	no	no	no
SFT2D2	yes	no	yes	yes	yes	yes
SFTP8	no	no	no	no	no	no
SGCG	yes	no	no	yes	no	yes
SGK1	no	yes	yes	yes	yes	yes
SGMS1	no	no	no	no	no	no
SGSH	yes	no	no	no	no	no
SH3GL2	yes	yes	no	yes	no	no
SH3KBP1	no	no	no	no	no	no
SHC1	no	no	no	no	no	no
SIAH1	yes	yes	yes	yes	yes	yes
SIKE1	no	no	no	no	no	no
SIPA1L2	no	no	no	yes	yes	no
SIRT1	yes	no	yes	yes	yes	yes
SLC2A3	yes	yes	yes	yes	yes	yes
SLC34A2	yes	yes	yes	yes	yes	yes
SLC35E4	no	no	yes	no	yes	no
SLU7	yes	no	no	yes	yes	yes
SNAI1	yes	yes	yes	yes	yes	yes
SNIP1	yes	no	no	yes	yes	yes
SNUPN	no	no	no	no	yes	no
SPTY2D1	yes	yes	yes	yes	yes	yes
SRPK3	no	no	no	no	no	no
SRSF5	no	no	no	no	no	no
SRSF8	no	no	yes	yes	yes	yes
STIL	yes	yes	no	yes	yes	yes
STK17B	no	no	no	no	no	no
STK31	no	no	no	no	no	no
STX6	yes	no	no	yes	yes	yes
SUPT6H	yes	no	yes	yes	yes	yes
SUSD2	no	no	no	no	no	no
SYNJ1	yes	yes	yes	yes	yes	yes
TAF4B	yes	yes	no	yes	yes	yes
TC2N	yes	yes	yes	yes	yes	yes
TCEB3	yes	no	yes	yes	yes	yes
TEFM	no	no	no	yes	yes	yes
TESK2	yes	yes	no	yes	yes	yes

Table S5 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
TFAP2C	yes	yes	no	no	yes	no
TFIP11	yes	yes	yes	yes	yes	yes
TGF β 2	no	no	yes	no	no	no
TGS1	no	no	no	no	no	no
THOC5	yes	yes	yes	yes	yes	yes
TIPARP	no	no	no	no	yes	no
TMED7	yes	no	yes	yes	yes	yes
TMEM185A	yes	no	yes	yes	yes	yes
TMEM254-AS1	no	yes	yes	no	no	no
TMTC1	no	no	no	no	no	no
TNFRSF10D	yes	no	no	no	no	yes
TOPORS	yes	no	no	yes	yes	yes
TP53BP2	yes	no	no	yes	yes	yes
TPRX1	yes	yes	yes	yes	yes	yes
TRAPP β 6B	yes	no	no	yes	yes	yes
TRIM23	yes	yes	yes	yes	yes	yes
TRIM35	no	no	no	no	no	no
TRIM36	yes	yes	no	yes	yes	yes
TRIM43	yes	yes	yes	yes	yes	yes
TRIM43B	no	yes	no	no	no	no
TRIM48	yes	yes	yes	yes	yes	yes
TRIM49	yes	yes	yes	yes	yes	yes
TRIM51	no	yes	yes	yes	yes	yes
TTC14	yes	no	no	no	no	no
TTC23	no	no	no	no	no	no
USP38	yes	no	yes	yes	yes	yes
UTP23	yes	no	no	no	no	yes
VEPH1	no	no	no	no	no	no
WHAMM	no	no	no	yes	no	no
WPEL5	no	no	no	yes	yes	yes
YTHDC1	yes	no	yes	yes	yes	yes
ZBTB10	no	no	no	yes	yes	yes
ZBTB24	no	yes	yes	yes	yes	no
ZCCHC10	no	no	no	yes	yes	yes
ZIM3	no	yes	yes	yes	yes	yes
ZMAT3	no	no	no	no	no	no
ZNF10	no	no	no	yes	yes	yes

Table S5 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
ZNF134	no	no	no	yes	yes	yes
ZNF217	yes	yes	yes	yes	yes	yes
ZNF256	yes	no	no	yes	yes	no
ZNF280A	yes	yes	no	yes	yes	yes
ZNF296	yes	yes	yes	yes	yes	yes
ZNF473	no	no	no	yes	yes	no
ZNF574	yes	yes	yes	yes	yes	yes
ZNF622	yes	no	no	yes	yes	yes
ZNF674-AS1	no	no	yes	no	no	no
ZNF705A	yes	yes	yes	yes	yes	yes
ZNF776	no	no	no	no	yes	no
ZNF789	no	no	no	no	no	no
ZNF827	no	no	no	no	no	no
ZNHIT6	yes	yes	yes	yes	yes	yes
ZRANB2	no	no	no	no	no	no
ZSCAN4	yes	yes	yes	yes	yes	yes
ZSCAN5A	yes	yes	no	yes	yes	no
ZSCAN5B	no	yes	yes	yes	yes	yes

Table S6: Shared differentially downregulated genes between 8.5h induced DIE cells and other datasets

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
AJUBA	no	no	yes	no	no	no
ANKRD34A	no	no	no	no	no	no
ARL4C	no	no	no	no	no	no
ATXN2	no	no	no	no	no	no
BBC3	no	no	no	no	yes	no
BCL9	no	no	no	no	no	no
BCOR	no	no	no	no	no	no
BEND4	no	no	no	no	no	no
BTG2	yes	no	yes	no	yes	no
C20orf112	no	no	no	no	no	no
CCDC85B	no	no	yes	no	yes	no
CD97	no	no	yes	no	yes	no

Table S6 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
CDC42EP1	no	no	yes	no	yes	yes
CEBPB	yes	no	yes	yes	yes	no
CELSR2	no	no	no	no	no	no
CHKA	no	no	no	no	no	no
CHST3	no	no	yes	yes	no	no
CITED4	no	no	no	no	no	no
DDIT4	yes	no	yes	yes	yes	no
DDN	no	no	no	no	no	no
DHRS3	no	no	no	no	no	no
DMRTA2	no	no	no	no	no	no
DUSP16	no	no	no	no	no	no
E2F2	no	no	no	no	no	no
EFNA2	no	no	no	no	no	no
EFNB1	no	no	no	yes	no	no
ELF4	no	no	no	no	yes	no
ELOVL1	no	no	yes	no	yes	no
EN2	no	no	no	no	no	no
EPC1	no	no	no	no	no	no
ERF	no	no	yes	no	no	no
FADD	no	no	yes	no	no	no
FAM124A	no	no	no	no	no	no
FAM155B	no	no	no	no	no	no
FAM174B	yes	no	no	no	no	no
FAM195A	no	no	no	no	no	no
FAM57A	yes	no	yes	yes	no	no
FAM78A	no	no	no	yes	no	no
FOSL1	no	no	yes	no	no	no
GAS2L1	no	no	yes	no	no	no
GIT1	no	no	yes	no	no	no
GLIS2	no	no	yes	yes	yes	no
GSE1	no	no	yes	no	no	no
H19	no	no	no	no	no	no
HDAC9	no	no	no	no	no	no
HES7	no	no	no	no	no	no
HOXA11	no	no	yes	no	yes	no
ID1	yes	no	yes	no	no	no
ID3	yes	no	yes	yes	no	no

Table S6 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
IGDCC3	no	no	no	no	no	no
INSR	no	no	no	no	no	no
IQCE	no	no	yes	no	yes	no
IRX3	no	no	yes	no	no	no
ITPRIPL2	no	no	yes	yes	no	no
JAG2	no	no	no	no	no	no
LDB2	no	no	no	no	no	no
LEF1	no	no	no	no	no	no
LRFN1	no	no	no	no	no	no
LRFN4	no	no	yes	no	no	no
MEX3A	no	no	no	no	no	no
MIDN	no	no	no	no	no	no
MLLT6	no	no	yes	no	yes	no
NAB2	no	no	yes	no	no	no
NAT8L	no	no	no	no	no	no
NAV1	no	no	yes	yes	yes	no
NBPF3	no	no	no	no	no	no
NINJ1	yes	no	yes	yes	yes	no
NOG	no	no	yes	no	no	no
NR2F2	no	no	yes	yes	no	no
NR2F6	no	no	no	no	yes	no
ORAI1	no	no	yes	no	no	no
PHF15	yes	no	yes	yes	yes	no
PHF23	no	no	no	no	no	no
PLCL2	no	no	no	yes	no	no
PLEKHG3	no	no	yes	no	no	no
PLEKHO1	no	no	yes	no	yes	no
PPP1R18	no	no	yes	yes	yes	no
PPP1R9B	no	no	yes	yes	yes	no
PPP3CC	yes	no	yes	no	yes	no
PTCH1	no	no	no	no	no	no
RARG	no	no	yes	yes	yes	no
RGMA	no	no	no	no	no	no
RILPL1	yes	no	yes	no	no	no
RTN4R	no	no	no	no	no	no
SALL2	no	no	yes	no	no	no
SAMD10	no	no	no	no	no	no

Table S6 continued

Gene	Geng	Rickard	Heuvel	Jagganathan		
				iDUX4	enDUX4	vDUX4
SAPCD2	no	no	yes	no	no	no
SERTAD3	no	no	no	no	no	no
SETD1B	no	no	no	no	yes	no
SHB	no	no	no	no	no	no
SHISA3	no	no	no	no	no	no
SIX5	yes	no	yes	no	yes	no
SNX33	no	no	yes	yes	yes	no
SOGA1	no	no	yes	no	no	no
SOX12	no	no	yes	no	yes	no
SOX13	no	no	no	yes	no	no
SPSB4	no	no	no	no	no	no
ST3GAL1	no	no	no	no	no	no
ST6GAL1	no	no	no	no	no	no
STK11	no	no	no	no	yes	no
TAL1	no	no	no	no	no	no
TBX5	no	no	no	no	no	no
TGIF1	no	no	no	yes	yes	yes
TGIF2	no	no	no	no	no	no
TIGD5	yes	no	yes	no	no	no
TLE3	no	no	yes	yes	no	no
TLX1	no	no	no	no	no	no
TMEM189	no	no	no	no	yes	no
TMUB1	no	no	no	no	no	no
TNS3	yes	no	yes	yes	yes	no
TPRN	no	no	no	no	no	no
TRIM24	no	no	no	no	no	no
TRIM47	yes	no	no	yes	yes	no
TRIM54	no	no	no	no	no	no
WDR89	no	no	no	no	no	no
ZC3H4	no	no	yes	no	no	no
ZNF48	no	no	no	no	no	no
ZSCAN16	no	no	no	no	no	no

Table S7. Gene specific spacers targeting the 4.5h upregulated genes

Gene	spacer 1	spacer 2	spacer 3
ZSCAN4	ACAGCAATAATTCATATGCA	CCATCACCATAGGGACACCT	AACCCTGTACTCACTAAGGC
ZNF217	CAAAATCTCACCTGAAACG	GGACACATAATGGCAAATCG	AACATGTCTTAATGCAACAC
ETDB	GTCATAAAGCAACTCTAGGG	TGACATCGACCTGTCAAGAT	TCAGTAAAACAGCACAACCG
SRSF8	CGGACGAAAGCGAAGCCCCG	CGCTACAGGGAATCTCGCTA	ACAGCCGATCTCCTTACAGC
PRAMEF1	ACTGGAGGTGTCCAGCCCCG	GAGTCTGGAAGTGTCTCCTG	AGACAGCAGAGGACCGTCCA
RBBP6	AAGTCGAACTGAACCAGCGA	GATATCATCGATCTAGGTCA	TGAGACACAACAATTCATCT
PNP	AGTGGTCAGAACCCTCTCAG	CCGGTCGTAGGCATCAGACA	TTGCCAGTACCTGTACTTCG
SIAH1	AAGTTGCGAATGGATCCCAA	CAGAAGACGCATATTTACAG	TAAGTCCATTACAACCCCTAC
ZNF296	CTGTGGCAAACAGTTCACAG	GGCCGCTGCCACTTGCACGG	GTGAGCGCATGTGCACTTTG
TRIM51	CTGAATGCAAGAAGACAACG	ACTCACCTCTGAATCCACTG	AAATCTCAGAAATCTGAACA
RFPL4B	CGTGCGACTTTGTCTCAAGT	TCTGCTTTGATTGCATCCAG	GAGGTGAAGTCATGGTCCCT
KHDC1L	TGTGCAGCTCAATGCAGCGA	GGGAACGAGTGTCTCAGCA	TCTTCCATGTGGAACACCAT
CCNA1	ACAAACTCGTCTACTTCAGG	AGATGAATCTACCAGCATAG	AGGCATGCGCAGCATTCTGG
TFIP11	CAAATACTCTTACAAGACCG	GGGTGCACATTATTCCTGTG	CATCGTCTATCCACTCATGA
LEUTX	CAGAAGGCTCAGTAPAGTCA	CCAGGATGAAACCTCGCAG	TTGGGAACAGACCTTTACTA
PRAMEF12	TAGATTACATGACTTCATCA	AAGGTCTAGGATCACCATCA	CATTACGACTGTCTCCAAGG
ZNF622	GTGTTGGCAAGATTTGCTTG	GTTCTCGTAGGCGTTGAAAG	TGATGAAGAATTGGAATGTG
HSPA1A	GGTGTGTGACAAGTGTCAAG	AATCTACCTCCTCAATGGTG	GTCAGGCCCCACCATTGAGG
RFPL1	TCTGTGTGATGCACCCACTT	GACAGCGCATCCACACTCCA	CAAAGTAGATCCTCCCCATG
RFPL4A	AGTGGATATGACGTTTCGATG	AGGAATCTGTGAACCGACAG	CCTGAGGAGTTTCCGAAGTG
PTPRJ	TTACTGTTGTGCATCAACCA	CTATACCTACAAGATACATG	ATGGGTCCACAGTCCCACG
RFPL2	CCACACCCTCTAACCTGATG	GCAGTGAATTAATGCACTTG	TGTGGGGAAGGGGCACACGA
SLC34A2	TATGATCTCGAGGTAATGGG	GCTGACAACGATGGACGTTG	GGGTGTAACCTACCAATCAG
DIO2	CCTGTTTGTAGGCATCGAGG	GGTGAAGAGTTCTCCTCAG	AGCCGCTCCAAGTCCACTCG
TRIM43B	GGAAATGTGTCATAAACCCAG	CATATCCCTACAGGGCGATG	AGAGGCAGCTGAGGAAGACC
GTF2F1	GAGGTGGACTACATGTGAGA	GTTCACCCTCAAGCTTCGGG	ACAAAGTCAACTTTTGCTACG
TRIM48	GCAAATGTGTGGCATTACACA	GAACCTTCAAAGAACCAG	AATCTCGCAAGTCTTCCAGA
MRPL49	GTCTACAAGGACATCAGCGA	GAACATTATCCTACCCTTAG	ATGTTCCGGGCTACGCTGCG
LOC441081	GGATCTGAGTTGGAGAACAG	TACCCAGCTCTAAGCGATG	AAATTGAAGACGGAATCACC
ZNF574	CCAGCCGATGCACAAAACGT	AGGGCCCGAGGAACAACAG	TGTGGAGCACTCATACCGAA
HSPA1B	GGTCTTGACAAGTGTCAAG	GTCAGGCCCTACCATTGAGG	TGGGTGAGGCCCTACCATTG
HSPA1A/B	CAAGGTGCGAAGCTGCTGC	CGGCTGATTGGCCGCAAGTT	CAAGGGCAAGATCAGCGAGG
DUXA	TCAGTTACACTCTCATCA	AGATGGTAAAAACAAATCAT	TTACTTTGAAACTCCACACC
DUSP18	TGCTGACCATATCCACAGCG	CACTGAGACATTGATGACCA	GAGTTAGGGGAGTCAGCCAC
C21orf91	GTGTGCAAGAGATCAGACTT	GGTACTTGGGAGTCAAACCTG	GTGGCCTCATAGTCACAACC



Table S7 continued

Gene	spacer 1	spacer 2	spacer 3
ZNHIT6	CCTCCGATAAATCACCCGGT	CAGGAATATCGCATAACAACG	GAGGTGAAGGATGAGAACGC
DPPA3	TTAATCCAACCTACATCCCA	TTGAGATACATGTTACTCGG	TGTAGGAGCAGCAGTCCCTCA
AVPI1	CAAGCCCTGTTTCAACGCAG	GATCATCTGGGAATGTGCAG	GGCTGCAGTGGTGTAGCGAG
CXCR4	CAACCACCACAAGTCATTG	TGACATGGACTGCCTTGCAT	CAGGACAGGATGACAATACC
HOXB2	GCAAGGCCGCGATCTCGACG	CGACCCTGCCGAGGAACCCG	ACCGGCCGATGAAGCACAAG
GLIDR	CAGGCTCAAAGCAACAAGG	CTTCAGTTAAAGACTACCAG	GCCACGACTCAGATCTACAC
MFSD11	TGCCCAAGAACAATCTGACAA	TAACGGTGATTAGCCTTGTG	GATTTCCAAGAACAAGCTG
NDEL1	AAGATGATTTAAGTCAGACT	CTCCGTTCCCTTGCCAACAG	GTTGGAGGCACAATTAGTAC
PLXNB3	GCAGTCGTACAGGATCACTA	GCCCTCTCTTACCACACGT	CCGGAGACTCTGCTACACGG
SNAI1	GGGACTCTCCTGGAGCCGAA	GGCTTCGGATGTGCATCTTG	GCTGACCTCCCTGTCAGATG
KHDC1	CACATACCTTCGCTGCATTG	TGACTGTAGTCGGACCACAC	GACTCCTATCATCATGCTCG
PRAMEF11	AGTTCTACAAAACACAGTCAA	AAGTTCTACAAAACACAGTCA	GCTTCTGAAGATTCCTCAAG
TGFB2	AGAAAACATAAAGTCCACT	TAGGGTCTGTAGAAAAGTGGG	AACAGCATCAGTTACATCGA
SPTY2D1	ACCACTGACTGTCCGCCTAG	CACAGGGCCAGAATAACTG	ACTCTGCGTGATTGTACTCG
TMEM254-AS1	GAAGAGGAAGTGAAAACCG	ATGCTGATCTACATCGACAA	ATGTGAAAGGCCCCCTAG
RIT2	ATGCGGGAGCAGTACATGCG	GAGTACAAGTGGAATGCT	AGAAGATGCTTATAAGACCC
PRRG4	TTTGGACCTCTTGCGCAATG	GATCAAATCTATTATACAGA	CAGCAATTAATCCAGTCAGA
ZNF705A	CAGCGTTTGTAACTCACAG	AAGGCGAAAGCAATTAGTAT	AATTTGTTTATGTGGTTTAG
KCNQ1OT1	TAGACAAAAGCTCCCAACG	CAGTTATTGAAACCTCTACG	GTATCCATGTGCAACCAATG
SERTAD1	GAGGTCAAAGAGGGAGCTAG	GCCACCAGGCCGTAGCATCG	CTCTGGCAGTCGACTCCTGG
MED26	GGTTGTAGGAAACACGACTT	GCGGTCGCACGGCGTTGACG	CAGAGCTTGATGCACCCAA
OSR2	CTTAGGCGGATCCTCTTGCG	AGGGGAAGCGCGCTCCACG	TGGGATACCCAGCGTCCAG
ZCCHC10	CATTGGACTTATGAATGCAC	CCATGCATCGGCTAATAGCC	ACAGATTATTATTGCAACAA
RGS2	TCAACACGACTGCAGACCCA	TTGTAAGAAGTAGCTCAAAC	ACTCCTGGGAAGCCAAAAC
PRSS23	GGAACCCAGAAGCTTCGAGT	GCTAAATTGAGGGTAGACTG	AAACGCACCCATGTGCCAA
DUX4	TGCCAGCGCGGAGCTCCTGG	GCTCCGCGCTGGCAGCTGGG	GGCAGGCGGCTGTGCAGCG

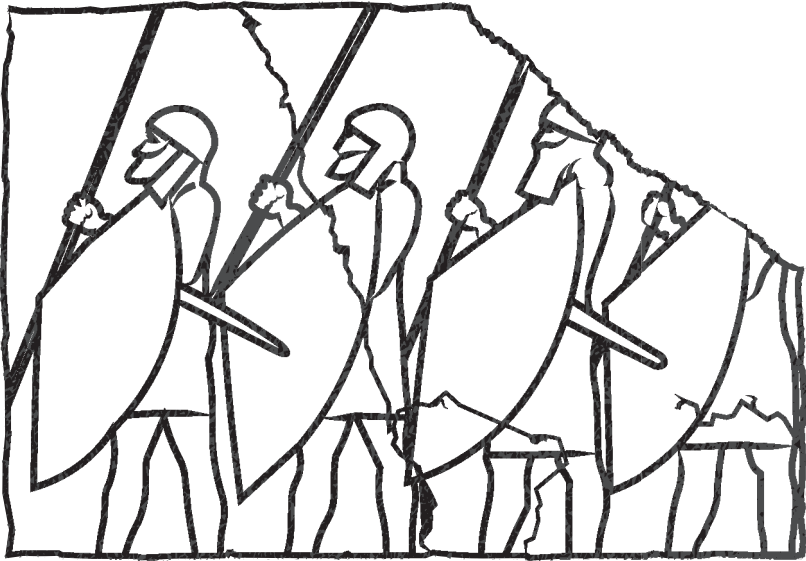


Illustration based on a stone carving on display at the Pergamon

Chapter 3

DUX4 induces a homogeneous sequence of molecular changes, culminating in cellular apoptosis

Ator Ashoti, Anna Alemany, Fanny Sage and Niels Geijsen

This chapter is available in an adapted form on bioRxiv

Ashoti, A., Alemany, A., Sage, F. & Geijsen, N. DUX4 induces a homogeneous sequence of molecular changes , culminating in the activation of a stem-cell-like transcriptional network and induction of apoptosis in somatic cells. bioRxiv (2021). doi.org/10.1101/2021.05.04.442407

Abstract

Facioscapulohumeral muscular dystrophy (FSHD) is a muscle degenerative disease that disproportionately affects the muscles of the face, shoulder girdle and upper arms. FSHD is caused by the ectopic expression of Double Homeobox 4 (DUX4), which has been derepressed due to aberrant genetic and/or epigenetic events. The expression of DUX4 in FSHD-affected tissue is low, with both transcript and protein proven difficult to detect. Yet when mis-expressed, this low expression can have great implications, which is evident in patients suffering from FSHD. This suggest that there might be more of these elusive genes, perhaps regulated by DUX4 itself, that can have great implications in the development of FSHD, but that have remained elusive due to stringent parameters set in transcriptional studies. Given that the earlier the intervention in the DUX4 induced cytotoxic cascade, the greater the impact on disease development and progression, we focused on finding subtle but robust changes in gene expression patterns early after DUX4 induction, by single cell RNA sequencing.

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is the third most prevalent muscular dystrophy worldwide¹. The disease is autosomal dominant, caused by a gain-of-function event, which leads to the ectopic expression of Double Homeobox 4 (DUX4)²⁻⁴ in affected skeletal muscle, primarily in the face, shoulders and upper arms. DUX4 is a transcription factor normally expressed during early embryonic development^{5,6}, in the adult testis⁷ and in the thymus⁸. It induces the expression of a network of genes involved in many different cellular processes, including embryonic development^{7,9,10}, RNA processing^{9,11,12}, protein homeostasis^{12,13}, germline development^{7,9,10}, stress response^{12,14}, and cell adhesion and migration^{11,15}, among many others. Expression of DUX4 is stochastic and low, yet potent enough to induce apoptosis in muscle tissue^{7,11,16,17}. Which genes and pathways play a defined role in inducing apoptosis downstream of DUX4 is not yet known.

Like many other muscular dystrophies, there is no effective treatment for FSHD. Many interconnected genetic and epigenetic events play a role, but at its core, the aberrant expression of DUX4 is the causal event that leads to muscle deterioration in FSHD patients. Genes involved in the expression and activation of DUX4, and genes that directly contribute to DUX4-induced cytotoxicity remain largely unknown. DUX4 is stochastically expressed in a burst-like fashion in only around 0.1-0.5% of myonuclei^{7,17,18}. Therefore, identifying key players, such as downstream key transcription factors, by performing RNA transcriptomics on primary material might fall short. Genes might be missed, especially ones that are as lowly expressed as DUX4 is in muscle fibers^{7,11,16}. This is evident in a study of Heuvel et al.¹⁹, where muscle tissue derived from 4 FSHD patients was analyzed by single-cell RNA sequencing. Out of the 5133 cells that were collected and analyzed from these patients, only 23 cells were classified as DUX4-affected. Cells were considered DUX4-affected if 5 or more DUX4 biomarker genes (out of a list of 67 biomarker genes²⁰) were differentially expressed compared to healthy control samples. This reinforces the idea that potential key players might be missed due to the low number of DUX4 affected cells not reaching a critical number needed to detect DUX4-induced transcriptional changes, and/or due to stringency in the analysis. This explains in part why there is little known about the underlying mechanism induced by DUX4 that leads these cells to apoptosis.

We have generated a transgenic cell line, in which DUX4 expression can be induced through the addition of doxycycline²¹. These so called DUX4-inducible expression (DIE) cells allow for precise titration and timing of the DUX4 response. The response in the DIE cells is robust, as 99-100% of the induced cells enter apoptosis²¹. Using this line, we interrogated whether DUX4 induction led to the induction of defined and orderly molecular changes, or whether it induced a stochastic disruption of gene expression networks before ultimately triggering apoptosis. In order to address this question, we performed single-cell RNA sequencing (SCS) on induced DIE cells, as early as 2 hours after induction. By mapping the early molecular changes that follow DUX4 activation at high temporal resolution, we demonstrate that DUX4 induction homogeneously triggers a series of sequential molecular changes that ultimately lead to apoptosis.

Results

Single cell analysis induced DIE cells

We previously demonstrated that the transcriptomic changes induced by DUX4 in DIE cells are very similar to those reported in FSHD-patient cells and other cellular models²¹. A robust DUX4 expression profile could be seen after only 4.5 hours of induction. However, the manner in which DUX4 expression leads to apoptosis, and what sort of paths are taken is not yet understood. Does DUX4 initiate a defined sequence of transcriptional events every time, or does it initiate a stochastic response that causes a disproportional amount of disruption in the cells that will eventually lead to cell death? To explore this question, we decided to analyze at single-cell resolution the transcriptional changes that occur shortly after DUX4 induction. Our inducible system allows us to examine the immediate effect of DUX4 induction, and track the changes overtime. Due to the robust induction of DUX4, >99% of the DIE cells enter apoptosis withing 48hours of DUX4 induction (Chapter 5), these changes can be tracked at a high resolution. This type of data would be difficult to attain with primary material, due to the low frequency of DUX4 expression, that occurs in a burst like fashion in a small subset of myonuclei. Furthermore, this inducible system allowed us to time the induction of DUX4, creating a clear timeline trajectory, which is not possible when working with primary material. SCS, will also allow us to detect subtle and perhaps rare early transcriptional changes in specific cell populations that could otherwise be drowned out and missed in bulk RNA sequencing. DIE cells were induced for 2, 3, 4 and 6 hours with doxycycline before sampling and processing for SCS. By reducing the dimensionality using t-Distributed Stochastic Neighbor Embedding (t-SNE) mapping, we were able to have a 2-D visualization of the cell clustering^{22,23}. Each point in the t-SNE map is a cell of which its position is determined based on its transcriptome. Cells with a similar transcriptome are more likely to cluster together than cells with large variations in their transcriptomes. Generally, the larger the differences in transcriptomics between cells, the further apart they will be in a dimensionality-reduced map. Our results show separate embedding of uninduced cells (0h) and 6h-induced cells, but mixed populations of the intermediate states (2h, 3h and 4h) (Fig. 1A). The cells do however orientate themselves on the y-axis, from the uninduced cells at the top, to the maximum of 6h-induced cells at the bottom (Fig. 1B). This is evident when the expression of known DUX4 target genes were projected onto the t-SNE map (Fig. 1C). LEUTX, PRAMEF1 and ZSCAN4 are genes that have previously shown to increase in expression in FSHD models or FSHD-affected muscle cells^{9,11,19,24}. This can indeed also be seen in Fig. 1C, where the expression of these genes is significantly upregulated in 6h-induced cells. This also holds true for genes that are downregulated upon DUX4 expression, such as ID1^{9,19}. Figure 1C also shows that as DUX4 induction persists, the expression of ID1 decreases significantly. As the cells are organized on the y-axis, we manually divided the vertical axis of the t-SNE into 10 clusters of equal size (Fig. 1D). The mean induction state was calculated for each cluster by considering all cells and their time of induction (Table 1). To avoid confusion, the mean induction state of each cluster will from here on be referred to as the experimental induction times used. Clusters 1 and 2 will therefore be referred to as 0h, 3&4 as 2h, 5&6 as 3h, 7&8 as 4h and 9&10 as 6h. Using this type of clustering, differential gene expression analysis was performed to identify differentially expressed genes between the uninduced cell clusters (clusters 1&2) and the induced cell clusters, as schematically indicated in figure 1D (right).

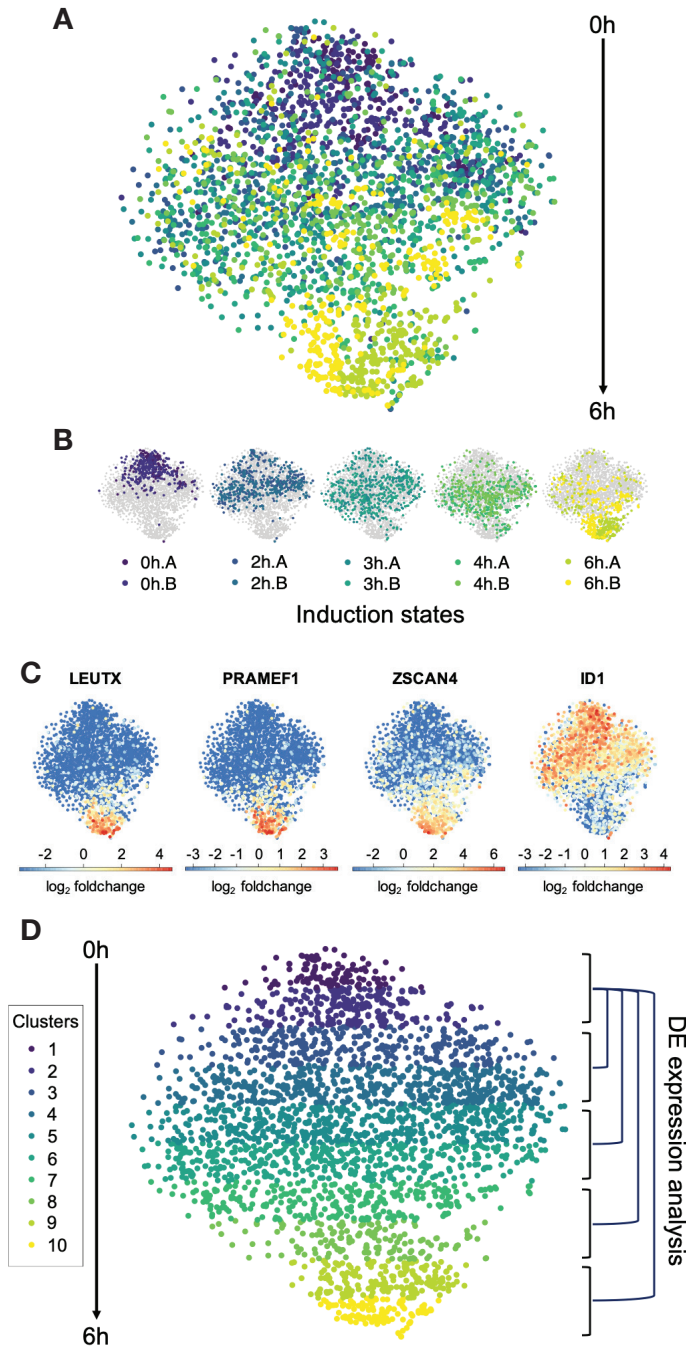


Figure 1. SCS data of DIE cells analyzed with RaceID. A) Induced and uninduced DIE single cell data represented in a t-distributed stochastic neighbour embedding (t-SNE) map. Each point represents a single cell, with the induction time of the cells indicated by color. **B)** Individual t-SNE maps for each of the induction time. Each sample is indicated in a different color. Induction states are shown from left to right, starting with the two 0h replicates. **A** and **B** annotations indicate the two replicates. **C)** t-SNE maps highlighting the expression of DUX4 marker genes. The fold change in gene expression is shown on a \log_2 scale as a linear color scale. **D)** Clustering of the cells in the t-SNE map based on the y-axis coordinates. Clusters are numbered (1-10) and color-coded. Clusters 1 and 2 contain the most uninduced DIE cells, and will be used as the control situation for differential gene expression (DE) analysis with the induced clusters (3&4, 5&6, 7&8, and 9&10).

Table 1. Cell make up of t-SNE clusters

cluster	0h A	0h B	2h A	2h B	3h A	3h B	4h A	4h B	6h A	6h B	cells/ cl.	mean state	
cl.1	33	23	15	10	10	5	0	4	5	1	106	1.39	0h
cl.2	61	77	22	17	25	14	7	11	6	1	241	1.28	
cl.3	38	106	48	29	30	23	9	17	11	1	312	1.57	2h
cl.4	34	57	83	71	69	58	38	49	18	29	506	2.61	
cl.5	10	20	54	108	58	51	69	90	15	40	515	3.14	3h
cl.6	1	3	33	41	58	65	64	69	23	48	405	3.64	
cl.7	0	0	12	14	33	41	76	37	31	39	283	4.05	4h
cl.8	0	0	2	1	15	14	20	11	52	42	157	4.97	
cl.9	0	1	0	0	2	6	25	5	84	61	184	5.51	6h
cl.10	1	0	1	1	1	2	3	1	71	33	114	5.73	

Differential gene expression analysis

In order to detect subtle but significant changes in expression, differentially expressed genes were filtered for an adjusted p-value (Padj) of $< 10^{-6}$, and a $\log_2(\text{foldchange})$ ($\log_2\text{FC}$) of > 0.5 and < -0.5 (Tables S1-4). This analysis demonstrated that as the induction time increases, so does the number of differentially expressed genes, with a core group of differentially expressed genes being shared between induction states, (Fig. 2A, Table 2 and Tables S1-4). This suggests that a deterministic chain of events is induced early on, if not immediately after DUX4 expression. Interestingly, even though DUX4 expression itself was not detected in the induced DIE cells, a DUX4 expression profile is readily detected only 3h post DUX4 induction, as can be seen when expression data is entered through the Enrichr database^{25,26} (Fig. 2B). This DUX4 profile even becomes more apparent as induction time increases. The Enricher database can match the entered lists of genes with previously entered studies, matching our gene lists with one other study from Geng et al.⁹ in which DUX4 had been overexpressed in human primary myoblasts. Such an early induced DUX4 expression profile has (to the best of our knowledge) not been seen before, with other studies measuring the effects of DUX4 6h²⁷ or 14h²⁴ post induction, or 24-36h post lentiviral transfection^{9,24}. As no other study has examined the effects of DUX4 at such early time points, our datasets uniquely show the earliest DUX4 affected genes. Furthermore, previously identified DUX4 affected genes such as RFPL4B, GOLGB1, ZNF296, SRSF8, ID1 and ID3^{9,11,19,24}, could too be classified as potential early marker genes, as they have been identified as being differentially expressed after a mere 3h of DUX4 induction.

Remarkably, a high percentage (29-36%) of upregulated transcripts encode for transcription factors and cofactors, in addition to a number of differentially expressed kinases (Table 2, Table S5), which was also observed in previous bulk-seq experiments²¹. This suggests that DUX4 induces a network of downstream transcription factors that in turn induces a cascade of secondary transcriptional events, ultimately leading to apoptosis. Since the expression of transcription factors can be low, many additional factors might fall under the detection limit

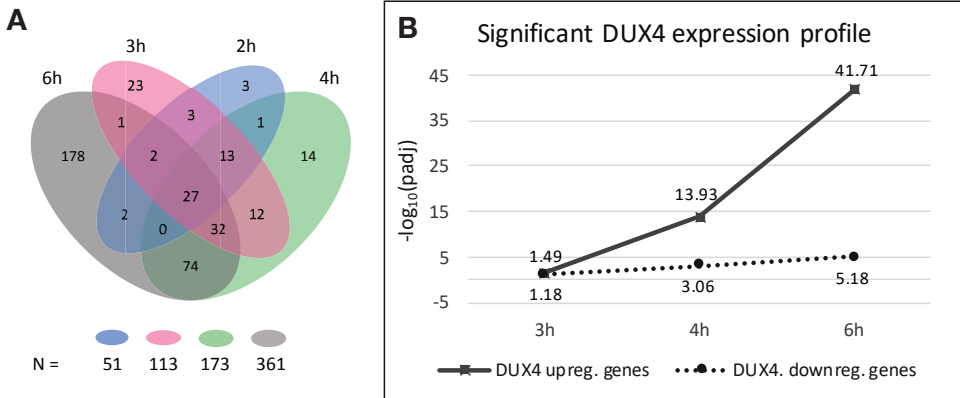


Figure 2. DUX4 differential gene expression profile between induction states. A) Venn diagram demonstrating the number of shared differentially expressed genes between different induction times. Total number of differentially expressed genes in each state is shown below the graph, in the same color-coding. **B)** A graph demonstrating the increase in significance of the upregulated and downregulated expression profile of DUX4 in induced DIE cells, according to entries in the Enrichr database. Induction samples (each containing two biological replicates) are plotted on the x-axis, and the y-axis displays the negative \log_{10} of the adjusted p-value (padj), of the detected DUX4 expression profile.

in single-cell sequencing data. Enrichr^{25,26} was therefore used to analyze our datasets for the presence of signature gene expression profiles, or “transcriptional footprints”, that are indicative of the activity of specific transcription factors. Analysis of our DIE cell data using Enrichr yielded a list of potential transcription factors that can explain the observed changes in gene expression (Table S6). Some of the identified profiles did indeed match differentially expressed transcription factors in the induced DIE cells (SOX3, NR2F2, ZNF217 and OTX2). In addition, the Enrichr algorithm detected several other transcription factor profiles of transcription factors which themselves are not found to be differentially expressed in our dataset (Table S6). However, a number of these “transcriptional footprints” were found in all 4 timepoints examined (LIN28, SOX5, ZIC3, JUNB, KLF10, MEIS2, MYCN, PITX, SETB1, ZEB2, ZNF503, MYB, WT1, NR2F2), suggesting that these factors are induced early after DUX4 induction and persist with continued DUX4 expression. Of particular interest is the identification of several transcription factors which are represented in both the upregulated as well as the downregulated gene set (e.g. ZIC3, JUNB, KLF10, MEIS2, MYCN and SETB1). As both upregulated and downregulated genes corresponded to the activity status of these transcription factors, it does strongly suggest their role in the DUX4-induced cytotoxic cascade, as opposed to only finding a one-sided effect.

In summary, we have found that the induction of DUX4 promotes changes in the transcriptional landscape of DIE cells as early as 2 hours post doxycycline administration. Many differentially expressed genes found in the early data sets maintain their differentially expressed status with continued DUX4 expression. This corroborates the notion that DUX4 initiates a clear progressive cascade of events, and does not stochastically and/or randomly affects genes and pathways. This is further corroborated by our finding that transcription factors and co-factors are overrepresented in the list of differentially expressed genes and account for approximately ~33% of the differentially upregulated genes, whereas they only comprise 11-13.5% of the human genome. This indeed suggests that DUX4 activates a coherent network of transcriptional regulators that together initiate a new cellular program

that ultimately leads to cell death. Lastly the presence of transcription factors could also be deduced from the detection of their transcriptional footprint, even when the expression of the individual factors themselves were not always detected. This is a common shortcoming of single cell sequencing, where the detection of low abundant transcripts can be missed.

Table 2: Summary of differentially expressed genes found in induced DIE cells at different induction times.

Induction state	2h	3h	4h	6h
Upregulated genes	43	77	133	248
Downregulated genes	8	36	40	68
Differentially expressed genes total	51	113	173	361
Transcription- and co-factors	12	28	44	81
Kinases	3	3	6	10

Gene ontology

To identify which biological processes are affected by the temporal changes in gene expression in the induced DIE cells, the differentially upregulated and downregulated genes were analyzed using the PANTHER (Protein ANalysis THrough Evolutionary Relationships) algorithm. PANTHER is an online tool that classifies proteins (and their corresponding genes) based on their family or subfamily, their molecular function, and their involvement in any biological processes and pathways, to facilitate high throughput analysis of datasets²⁸⁻³⁰. The biological processes that were identified are shown in Fig. 3A and Table S7. Gene ontology (GO) terms were assigned a general “umbrella” term. Table S7 shows the full list of GO terms.

As shown, DUX4-induction initially triggered the activation of an early developmental program and processes involved in the cell cycle and proliferation. Three hours after induction, genes involved in developmental processes are less prominent, with the majority of processes now being involved in cell cycle and RNA processing. At 6h of induction, the first apoptotic processes were identified. GO terms identified with the downregulated genes appeared more incoherent than the GO terms detected with the upregulated gene sets. This could be due to the nature of DUX4 being more of a transcriptional activator, rather than a repressor¹⁰. GO terms found with the downregulated gene sets might thus reflect the loss of cell identity, consistent with the idea that DUX4 initiates an early embryonic transcriptional program. Nonetheless, processes involved in programmed cell death were also found upon analyzing the downregulated genes after 6h of DUX4 induction (Fig. 3B). Furthermore, at 3h post DUX4-induction, downregulated genes demonstrate changes in cellular respiration and energy production. These processes contribute to oxidative stress, a common occurrence in FSHD-affected cells that is likely involved in DUX4-induced apoptosis³¹⁻³³.

The temporal identification of altered biological processes revealed a sequential path that is activated upon DUX4 induction. This path starts by activating developmental processes, and subsequently many other processes involved in RNA processing, protein production and regulation, cellular respiration, kinase activity, eventually leading to the induction of apoptosis

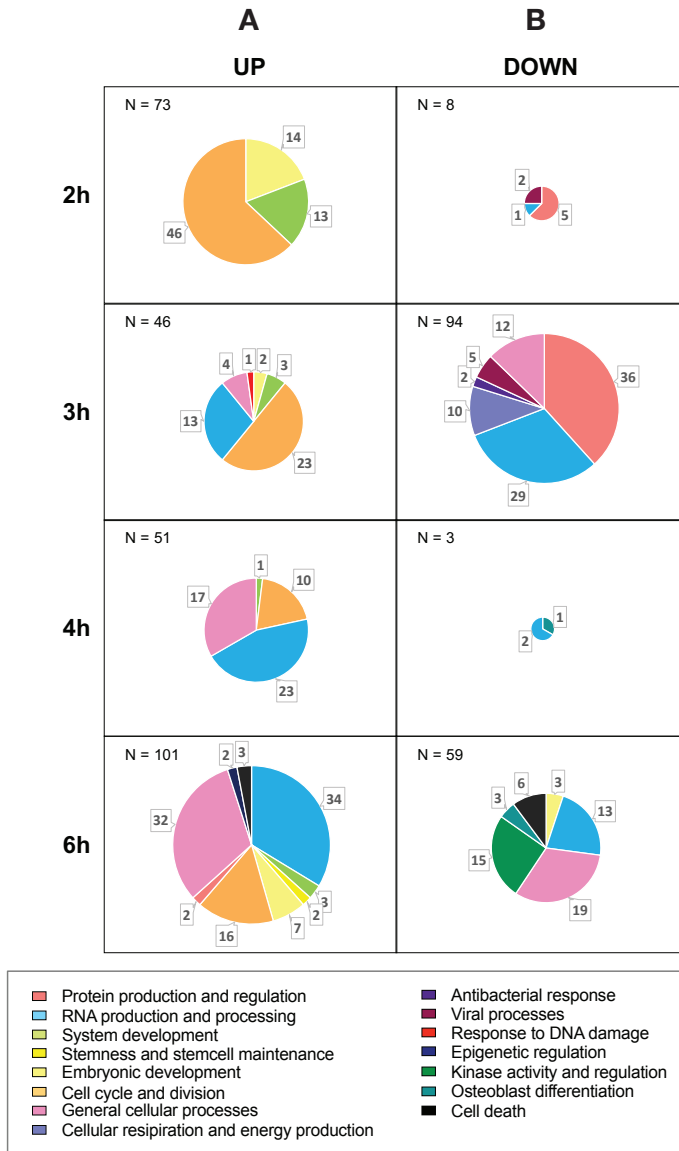


Figure 3. Gene ontology reveals DUX4-induced paths. Gene ontology performed on gene lists of differentially **A)** upregulated and **B)** downregulated genes from the 4 induction states. Only biological processes with an FDR < 0.05, and a raw p value of 10^{-3} are included. Biological processes are color-coded based on their general “umbrella” term. The total number of detected biological processes is indicated in the top left corner of each diagram, with the number of biological processes per umbrella term annotated in the pie chart. See table S7 for the full list of biological processes per induction state.

StemID

The observation that DUX4 initially activates an early embryonic transcriptional program was interesting and suggests that DUX4 temporarily converts cells toward a developmentally immature state. This notion is corroborated by StemID, an algorithm that uses transcriptome entropy to identify stem cells within a cell population³⁴. Determining transcriptome diversity in single cells is done by using Shannon's entropy³⁵, which measures disorder in high-dimensional systems. The entropy value of a given cell type indicates the degree of transcriptomic promiscuity. As pluripotent stem cells have the option of differentiation in any cell type, a wide number of signaling pathways need to remain active, which is reflected as high transcriptome entropy. As these cells become more committed to a specific cell fate, the number of active pathways decrease to a few specific pathways needed to maintain their cell identity, which in turn leads to a decrease in transcriptome entropy^{34,36,37}. When a lineage trajectory is projected onto the t-SNE map, it becomes clear by the color indication of the vertices that transcriptome entropy peaks in cluster 4 (Fig. 4A). The barplot in Fig. 4B clearly shows the increase in transcriptome entropy, until it peaks in cluster 4, which has an induction state of around 2h. Transcriptome entropy then slowly decreased with increasing induction times. This is thus corroborating gene ontology results that showed the induction of a more embryonic developmental state in ~2h induced DIE cells (clusters 3&4).

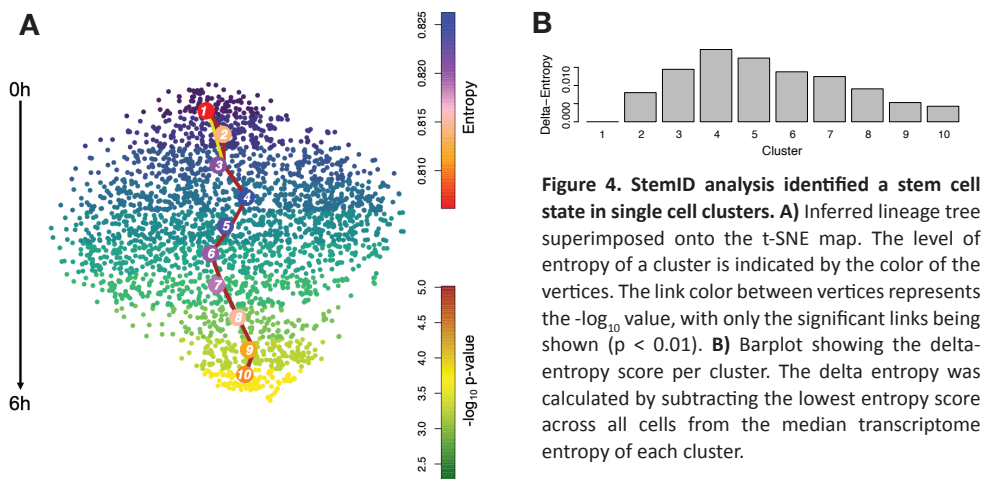


Figure 4. StemID analysis identified a stem cell state in single cell clusters. A) Inferred lineage tree superimposed onto the t-SNE map. The level of entropy of a cluster is indicated by the color of the vertices. The link color between vertices represents the $-\log_{10}$ value, with only the significant links being shown ($p < 0.01$). **B)** Barplot showing the delta-entropy score per cluster. The delta entropy was calculated by subtracting the lowest entropy score across all cells from the median transcriptome entropy of each cluster.

Pseudotime analysis with FateID

StemID revealed a transient trajectory of stemness in induced DIE clusters, which peaked at cluster 4 and was subsequently followed by a gradual decrease of stemness at later time points, when further transcriptome changes reflect profound changes in metabolism and RNA processing. To follow up on this observation, pseudotime analysis was employed to further define temporal stages of transcriptional states, using FateID³⁸. By doing so, we were able to identify stage-specific co-expression patterns across this vertical trajectory based on previous t-SNE clustering (Fig. 1D). Expression patterns of known DUX4 target genes show a gradual increase (LEUTX, ZSCAN4, ZNF217, and PRAMFE1), or decrease in expression (ID1 and ID3) as DUX4 expression persisted (Fig. 5A), which is in line with earlier observations seen above (Fig. 1A) and previous observations in other studies^{9,11,19,24}. These genes were

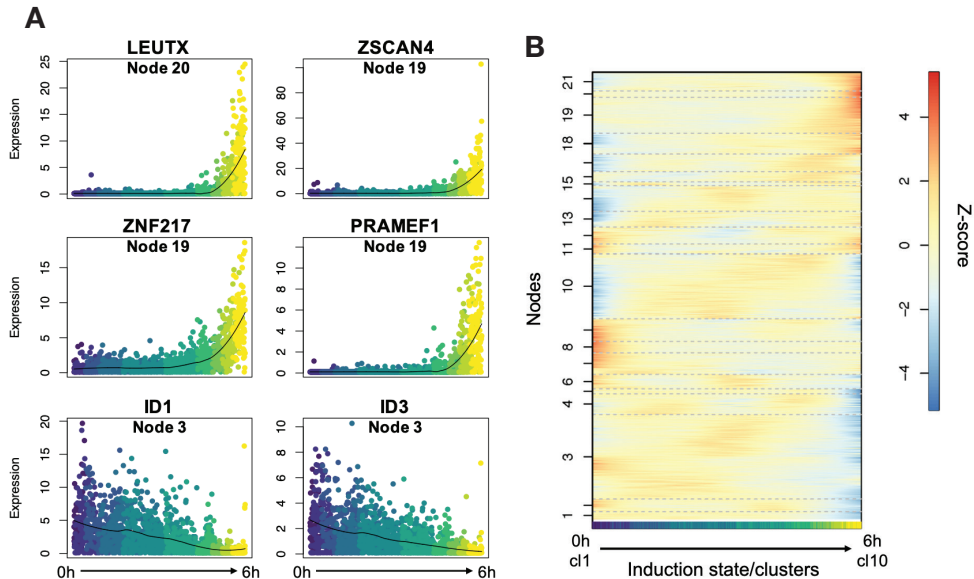


Figure 5. Gene expression patterns in the induced DIE cell trajectory by FateID pseudotime analysis. A) Expression patterns of 6 known DUX4 marker genes following the DIE cell induced trajectory (0h-6h induction). The nodes in which these genes are contained are annotated below the gene name. **B)** Self-organizing heatmap of z-score transformed pseudotime expression profiles across the DIE cell induced trajectory (0h-6h induction), based on the t-SNE map. Cells are represented on the x-axis, and the genes are organized in nodes that are represented on the y-axis. Genes with a similar expression pattern are clustered in nodes, with a color indication representing gene expression, based on their transformed z-score.

present in gene nodes that showed a gradual increase or decrease in gene expression (e.g. nodes 18-21 or 1-3 respectively) (Fig. 5B). Moreover, dynamic gene expression patterns were identified in other gene nodes, such as oscillating expression patterns during the 6 hours of DUX4 induction (e.g. 4, 6, 10 and 17) (Fig. 5B and S2). This suggests the activation of a very dynamic underlying process, upon DUX4 induction, in which some genes are induced and inhibited multiple times in a relatively short time frame.

Analysis of the differentially expressed genes in the oscillating nodes did not yield a clear answer as to why these particular sets of genes vary in their expression during DUX4 induction. Of the nodes that demonstrated clear oscillating patterns (4, 6, 10, 12, 17), node 4 and 6 did not contain differentially expressed genes, node 12 contained one differentially downregulated gene (COX7A2), node 10 contained 27 differentially upregulated genes, and node 17 contained 17 differentially upregulated genes (Table 3). Using the STRING database^{39,40}, we were able to determine that the differentially expressed genes from node 10 are primarily involved in developmental process, system development, and cell cycle and division, with many genes interconnecting and involved in all three processes. STRING is a database of known and predicted protein-protein interactions (both direct and indirect), that allowed us to visualize the types of associations between genes (Fig. 6A). The biological processes identified with STRING are similar to those identified using the PANTHER algorithm (Fig. 3A). The expression pattern in node 10 therefore fits previous

GO analyses, demonstrating a temporal increase in the number of developmental genes and processes, peaking at around 2-3h (Fig. 5B and 6B). Differentially upregulated genes in node 17 did not show to be part of any significantly affected biological processes, nor a clear coherent core network could be seen between the genes as was found in node 10.

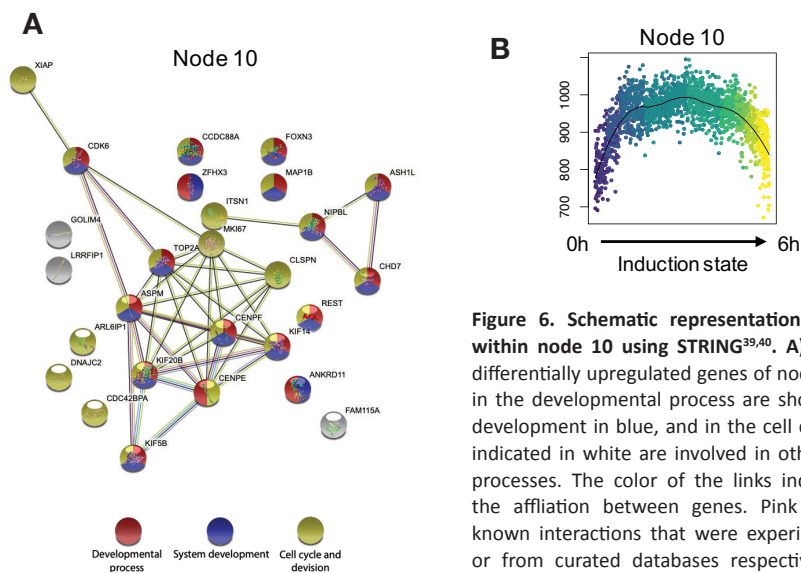


Figure 6. Schematic representation of gene affiliations within node 10 using STRING^{39,40}. **A)** Affiliations between differentially upregulated genes of node 10. Genes involved in the developmental process are shown in red, in system development in blue, and in the cell cycle in yellow. Genes indicated in white are involved in other general molecular processes. The color of the links indicates the nature of the affiliation between genes. Pink and cyan represent known interactions that were experimentally determined, or from curated databases respectively. Green, red and blue interactions represent predicted interactions based on gene neighborhood, gene fusions and gene occurrence

respectively. Yellow, black and purple interactions are based on textmining, co-expression, and protein homology respectively. Images were adapted from STRING-derived interaction networks. **B)** Expression pattern graph of node 10 genes, following the DIE cell induced time trajectory (0h-6h induction).

Discussion

Although ectopic activation of the transcription factor DUX4 has been identified as the main culprit of FSHD this past 10 years²⁻⁴, the exact mechanism by which DUX4 expression initiates muscle fiber degeneration remains elusive. A thorough understanding of the temporal molecular changes that are brought about by DUX4 is essential for the identification of potential targets to modulate DUX4 cytotoxicity. To gain knowledge of the molecular mechanism of FSHD, many researchers have studied the transcriptomics of FSHD models or FSHD-affected primary cells^{9,11,19,24}. Yet due to the broad range of gene regulation and the low and stochastic expression of the disease-causing gene (DUX4)^{7,16-18}, finding key players in the DUX4 induced cytotoxic cascade has been challenging. We therefore applied a novel transgenic cell model in which DUX4 expression can robustly be induced, allowing high-resolution temporal analysis of early transcriptional events following DUX4 induction at single-cell level. Our SCS data reveals how DUX4 induction lead to the activation of a transient early embryonic and stemness state, by activating a network of developmental factors. Early intervention in the DUX4-induced cytotoxic cascade will most likely have a greater impact on slowing down disease development and progression, than intervening at a later stage. By doing so we aimed to identify perhaps subtle and early changes, that can ideally be tracked further along the cytotoxic cascade, with potential implications in the

progressive cytotoxic cascade.

Our analysis of the single-cell transcriptional changes following DUX4 induction suggests that DUX4 initiates a non-random consecutive chain of events that is exacerbated as time progresses. In addition, analysis of gene expression profiles using Enrichr revealed that transcriptional signatures of transcription factors, most of which were not detected during the induction periods, are already prevalent 2 hours post-induction and indeed became more profound as induction time progresses. As such, we have uncovered a number of transcription factors that might play a role in triggering subsequent molecular changes that ultimately contribute to DUX4-mediated cytotoxicity.

At a high resolution, we were able to show the early events that occur only a few hours after DUX4 induction. Gene ontology analysis of the differentially upregulated genes at different time points after DUX4 induction corroborates our conclusion that DUX4 induces a non-random, consecutive sequence of events. Our results suggest an initial activation of developmental processes that lead cells to an increased stemness state only 2 hours after DUX4 induction. Next, additional biological processes such as RNA processing, and protein production and regulation were activated, eventually leading to the activation of apoptotic processes 6 hours post DUX4 induction.

The analysis performed in this study revealed which genes started diverging in their expression during the first few hours of DUX4 induction. At these early timepoints, some transcriptional changes were subtle, but as these changes in expression remained or even intensified with increased induction times, we believe them to be significant. More attention should be focused towards some of these subtle changes in expression at these early post-induction timepoints, that could normally be missed due to too stringent filtering. DUX4 itself proves that the smallest changes in expression can cause major consequences. This factor was not identified in our transcriptomic analysis, concurrent with its low transcriptional and abundance levels in muscle tissue of FSHD patients^{3,7,17,18}. More attention should thus be directed towards genes that leave behind a detectable “footprint” in the transcriptome of affected cells, again with DUX4 as prime example, as classification of most cells that are DUX4-affected is based on the detection of DUX4 marker genes, and not the detection of DUX4 itself. These elusive genes could have great implications in the pathophysiology of FSHD, and could therefore hold promise for its treatment.

Methods

Cell culturing and seeding

DIE cells were cultured in growth medium consisting of IMDM basal medium with 10% Tet system-approved FBS (Clontech) and 55 μ M 2-mercaptoethanol, supplemented with 5 μ g/ml Puromycin and 6 μ g/ml Blasticidin.

Sample preparation and SORT-seq

DIE cells were grown in 48-wells plates, until a ~90% confluency was reached. Cells were exposed to 1 μ g/ml doxycycline for 2, 3, 4, and 6 hours. Doxycycline-exposed and untreated DIE cells were rinsed with DPBS after which 0.25% Trypsin-EDTA (Thermo Scientific) was added. Trypsin was immediately removed after it had covered the complete surface. Cell

3

were incubated for 1 minute at 5% CO₂ and 37°C, after which the trypsin was deactivated by adding IMDM media supplemented with 10% Tet-system approved FBS and DAPI nuclear stain. Trypsinized DIE cells were resuspended in the media and then strained using Cell-strainer capped tubes (Falcon). Cells were stored on ice until FACS sorting. Viable DAPI negative cells were sorted into 384 hard shell plates (Biorad) with 5 µl of vapor-lock (QIAGEN) containing 100-200 nl RT primers, dNTPs, and synthetic mRNA Spike-Ins, using the FACSJazz (BD biosciences). The plates were immediately spun down and stored at -80°C. Cells were processed as described in Muraro et al.⁴¹, using the CEL-seq2-bases scRNA-seq. Samples were sequenced using Illumina Nextseq 500, 2x75 kit, high output. Two biological replicates per samples were sent for sequencing. Initial normalization and mapping were done as described by Muraro et al.⁴¹.

Data analysis

Illumina sequencing-generated paired-end reads were aligned, mapped, and normalized as previously described⁴¹. For single cell analysis, cells with a minimum of 6000 transcripts were considered, and data normalization was performed by downsampling transcript counts to 6000 for all cells (Fig. S1). Initial analysis revealed a batch effect between the two uninduced biological replicates. One sample showed signs of additional metabolic stress. The top 187 diverging genes ($\text{padj} < 10^{*-7}$) between the two uninduced biological replicates were removed from all data to account for any source of metabolic stress. Dimensionality reduction of cells was done using RaceID²³, after which the clusters were manually determined by dividing the y-axis of the tSNE map in 10 clusters of equal size. Differential expression of genes between cell clusters were identified as described by Muraro et al.⁴¹, based on a previous publication of Anders and Huber⁴². Pseudotime analysis was performed using StemID³⁴ and FateID³⁸.

Data Resources

RNA sequencing data is available in the GEO data base, accession number: GSE156154.

References

1. Deenen, J. C. W. *et al.* Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* **83**, 1056–1059 (2014).
2. Kowaljow, V. *et al.* The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul. Disord.* **17**, 611–623 (2007).
3. Dixit, M. *et al.* DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18157–18162 (2007).
4. Snider, L. *et al.* RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: New candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum. Mol. Genet.* **18**, 2414–2430 (2009).
5. Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **49**, 935–940 (2017).
6. Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
7. Snider, L. *et al.* Facioscapulohumeral dystrophy: Incomplete suppression of a retrotransposed gene. *PLoS Genet.* **6**, 1–14 (2010).
8. Das, S. & Chadwick, B. P. Influence of repressive histone and DNA methylation upon D4Z4 transcription in non-myogenic cells. *PLoS One* **11**, 1–26 (2016).
9. Geng, L. N. *et al.* DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy. *Dev. Cell* **22**, 38–51 (2012).

10. Knopp, P. *et al.* DUX4 induces a transcriptome more characteristic of a less-differentiated cell state and inhibits myogenesis. *J. Cell Sci.* **129**, 3816–3831 (2016).
11. Rickard, A. M., Petek, L. M. & Miller, D. G. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum. Mol. Genet.* **24**, 5901–5914 (2015).
12. Jagannathan, S., Ogata, Y., Gafken, P. R., Tapscott, S. J. & Bradley, R. K. Quantitative proteomics reveals key roles for post-transcriptional gene regulation in the molecular pathology of facioscapulohumeral muscular dystrophy. *Elife* **8**, 1–16 (2019).
13. Homma, S., Beermann, M., Lou, Boyce, F. M. & Miller, J. B. Expression of FSHD-related DUX4-FL alters proteostasis and induces TDP-43 aggregation. *Ann. Clin. Transl. Neurol.* **2**, 151–166 (2015).
14. Winokur, S. T. *et al.* Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. *Neuromuscul. Disord.* **13**, 322–333 (2003).
15. Bosnakovski, D. *et al.* Transcriptional and cytopathological hallmarks of FSHD in chronic DUX4-expressing mice. *J. Clin. Invest.* **130**, 2465–2477 (2020).
16. Bosnakovski, D. *et al.* Muscle pathology from stochastic low level DUX4 expression in an FSHD mouse model. *Nat. Commun.* **8**, 1–9 (2017).
17. Tassin, A. *et al.* DUX4 expression in FSHD muscle cells: How could such a rare protein cause a myopathy? *J. Cell. Mol. Med.* **17**, 76–89 (2013).
18. Jones, T. I. *et al.* Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: Evidence for disease modifiers and a quantitative model of pathogenesis. *Hum. Mol. Genet.* **21**, 4419–4430 (2012).
19. Van Den Heuvel, A. *et al.* Single-cell RNA sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. *Hum. Mol. Genet.* **28**, 1064–1075 (2019).
20. Yao, Z. *et al.* DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle. *Hum. Mol. Genet.* **23**, 5342–5352 (2014).
21. Ashoti, A. *et al.* A genome-wide CRISPR/Cas phenotypic screen for modulators of DUX4 cytotoxicity reveals screen complications. *bioRxiv* 2020.07.27.223420 (2020) doi:10.1101/2020.07.27.223420.
22. van der Maaten, L. & Hinton, G. H. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
23. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
24. Jagannathan, S. *et al.* Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum. Mol. Genet.* **25**, ddw271 (2016).
25. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, (2013).
26. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
27. Choi, S. H. *et al.* DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **44**, 5161–5173 (2016).
28. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
29. Carbon, S. *et al.* The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
30. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
31. Tsumagari, K. *et al.* Gene expression during normal and FSHD myogenesis. *BMC Med. Genomics* **4**, (2011).
32. Banerji, C. R. S. *et al.* β -catenin is central to DUX4-driven network rewiring in facioscapulohumeral muscular dystrophy. *J. R. Soc. Interface* **12**, (2015).
33. Lek, A. *et al.* Applying genome-wide CRISPR-Cas9 screens for therapeutic discovery in facioscapulohumeral muscular dystrophy. *Sci. Transl. Med.* **12**, 9–11 (2020).
34. Grün, D. *et al.* De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell* **19**, 266–277 (2016).
35. Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **27**, 623–656

- (1948).
36. Macarthur, B. D. & Lemischka, I. R. Statistical mechanics of pluripotency. *Cell* **154**, 484–489 (2013).
 37. Banerji, C. R. S. *et al.* Cellular network entropy as the energy potential in Waddington’s differentiation landscape. *Sci. Rep.* **3**, 25–27 (2013).
 38. Herman, J. S., Sagar & Grün, D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nat. Methods* **15**, 379–386 (2018).
 39. Snel, B., Lehmann, G., Bork, P. & Huynen, M. A. String: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res.* **28**, 3442–3444 (2000).
 40. Szklarczyk, D. *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
 41. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst.* **3**, 385–394.e3 (2016).
 42. Anders, S. *et al.* Differential expression analysis for sequence count data via mixtures of negative binomials. *Adv. Environ. Biol.* **7**, 2803–2809 (2010).

Supplementary data

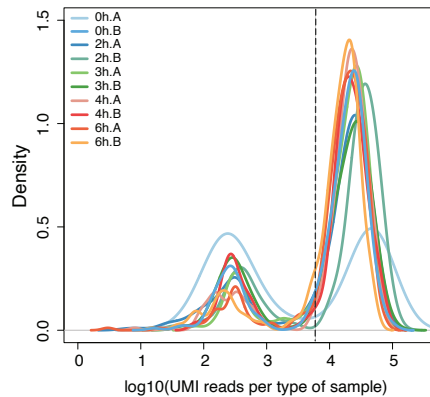


Figure S1. Density plot representing the total read count of all samples. Samples are color coded. An A or B annotation represents to which biological replicate the sample belongs. The intermittent line shows the cutoff of the number of UMI reads (6000) used to determine which cells to include for the analysis. All cells with a normalized transcript count above 6000 have been included in the RaceID analysis.

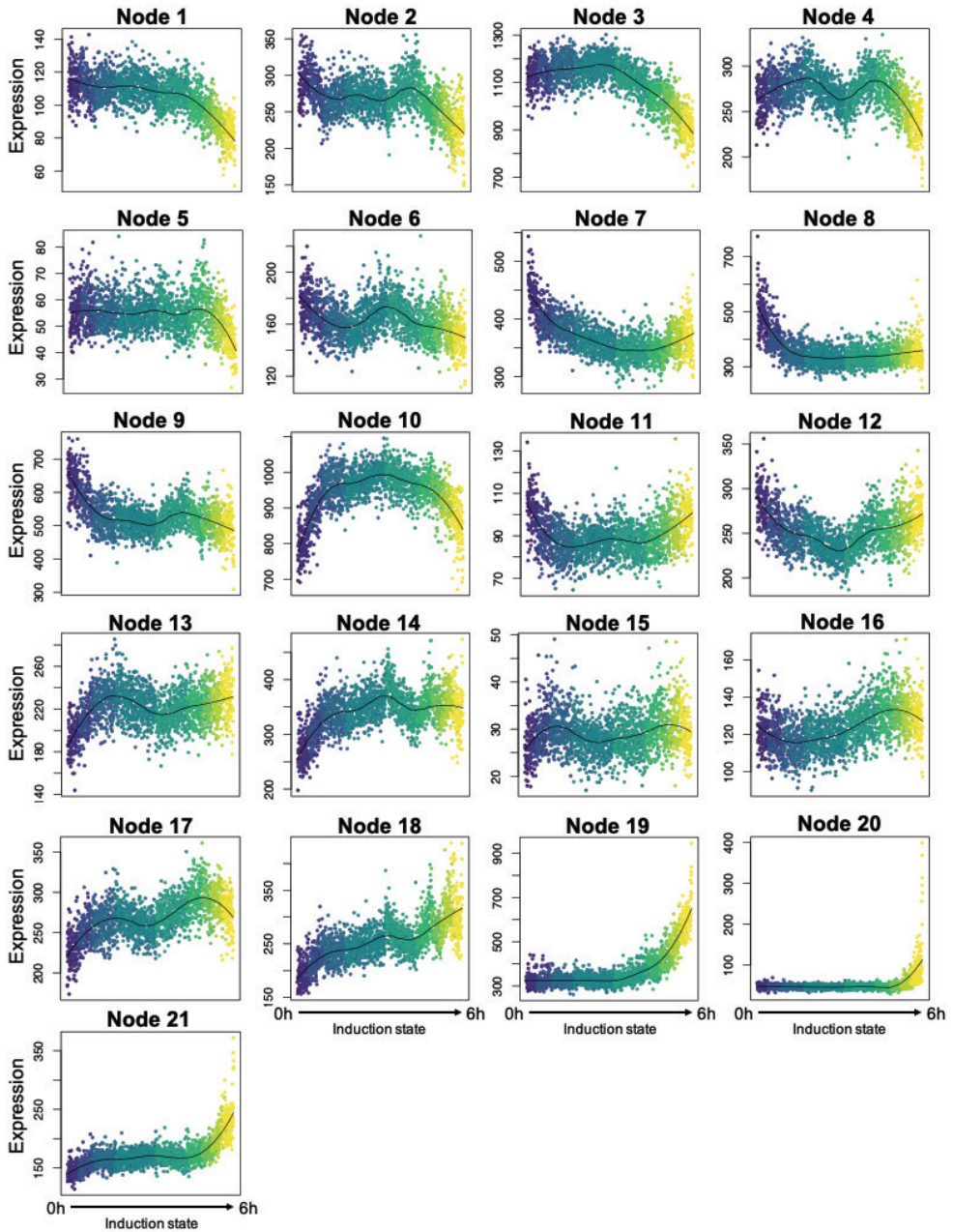


Figure S2. Gene expression patterns of gene nodes from pseudotime analysis. Dynamic gene expression patterns of all nodes of the self-organizing heat map of figure 5A. Each point represents a cell. The color of the point and its location on the x-axis represents its induction state from uninduced (0h) to 6h induced. Normalized expression is plotted on the y-axis. The black line indicates a local regression.

Table S1. Differentially expressed genes between uninduced (0h, clusters 1&2) and 2h induced DIE cells (clusters 3&4).

* Adjusted p value < 10⁻⁶, absolute log₂FC > 0.5

Gene	base Mean	base MeanA	base MeanB	fold Change	log ₂ FC	pval	padj	shared in states	Node
CHD2	0.68	0.52	0.83	1.61	0.68	4.4E-09	3.6E-07	All 4	14
UQCR11	1.09	1.29	0.89	0.69	-0.54	1.8E-09	1.6E-07	All 4	8
MDN1	0.33	0.22	0.45	2.08	1.06	1.1E-09	9.9E-08	All 4	14
SREK1	1.03	0.83	1.24	1.49	0.57	1.0E-09	9.2E-08	All 4	17
PGM5P2	0.70	0.52	0.87	1.66	0.74	1.6E-10	1.7E-08	All 4	18
BPTF	1.26	1.02	1.50	1.46	0.55	6.5E-11	7.3E-09	All 4	14
AKAP9	0.72	0.53	0.91	1.71	0.77	7.9E-12	1.0E-09	All 4	14
PRR11	1.06	0.82	1.30	1.60	0.67	4.6E-13	6.5E-11	All 4	13
NKTR	1.36	1.07	1.64	1.53	0.61	1.2E-13	1.7E-11	All 4	14
SRRM1	2.02	1.66	2.38	1.43	0.52	5.9E-15	9.9E-13	All 4	14
BDP1	0.77	0.53	1.00	1.87	0.90	2.5E-16	4.6E-14	All 4	14
GOLGA4	1.21	0.90	1.52	1.69	0.76	4.9E-18	9.8E-16	All 4	14
PRPF38B	2.03	1.62	2.43	1.50	0.59	1.5E-18	3.1E-16	All 4	18
TOP1	1.59	1.23	1.95	1.59	0.67	1.3E-18	2.9E-16	All 4	14
MAB21L3	0.91	0.63	1.19	1.88	0.91	2.2E-19	5.3E-17	All 4	18
USMG5	1.91	2.35	1.47	0.63	-0.68	4.7E-24	1.5E-21	All 4	8
BRD4	1.55	1.13	1.97	1.75	0.80	2.1E-25	7.3E-23	All 4	14
UGDH-AS1	1.38	0.98	1.78	1.82	0.87	4.7E-26	1.7E-23	All 4	18
ANKRD11	2.05	1.55	2.55	1.64	0.72	1.1E-26	3.9E-24	All 4	10
LOC100131257	2.17	1.63	2.70	1.65	0.72	5.8E-29	2.4E-26	All 4	18
CENPE	1.09	0.71	1.48	2.07	1.05	3.3E-29	1.4E-26	All 4	10
ASPM	1.41	0.91	1.91	2.09	1.07	6.3E-38	3.9E-35	All 4	10
SMC4	6.10	4.99	7.21	1.44	0.53	9.8E-44	8.4E-41	All 4	14
RPL37A	8.14	9.57	6.70	0.70	-0.52	2.3E-57	2.5E-54	All 4	8
CENPF	3.13	2.14	4.13	1.94	0.95	1.0E-67	2.0E-64	All 4	10
RPS29	12.43	14.98	9.87	0.66	-0.60	1.8E-117	1.1E-113	All 4	8
KCNQ10T1	7.48	5.20	9.75	1.87	0.91	6.7E-145	8.0E-141	All 4	18
TOP2A	1.74	1.25	2.24	1.79	0.84	1.6E-30	7.2E-28	2h 6h	10
MKI67	2.57	1.96	3.19	1.63	0.71	1.8E-32	9.3E-30	2h 6h	10
FOXN3	1.18	0.97	1.40	1.44	0.53	1.2E-09	1.1E-07	2h 4h	10
HIST1H2BK	0.78	0.95	0.61	0.64	-0.64	1.2E-09	1.1E-07	2h 3h 6h	19
KIF14	0.95	0.69	1.20	1.73	0.79	1.5E-15	2.6E-13	2h 3h 6h	10



Table S1 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
GOLIM4	0.43	0.30	0.56	1.86	0.90	1.2E-09	1.1E-07	2h 3h 4h	10
ANKRD12	0.81	0.63	0.99	1.58	0.66	6.8E-10	6.4E-08	2h 3h 4h	13
CDC42BPA	0.75	0.58	0.93	1.61	0.69	3.5E-10	3.5E-08	2h 3h 4h	10
ROCK1	0.52	0.37	0.67	1.80	0.85	2.4E-10	2.5E-08	2h 3h 4h	14
PIP5K1A	0.32	0.44	0.21	0.48	-1.06	1.4E-10	1.5E-08	2h 3h 4h	8
CCDC88A	0.65	0.47	0.82	1.73	0.79	3.6E-11	4.2E-09	2h 3h 4h	10
ITSN1	1.11	0.88	1.35	1.53	0.61	1.4E-11	1.7E-09	2h 3h 4h	10
NOP10	1.82	2.13	1.50	0.70	-0.51	1.3E-13	1.9E-11	2h 3h 4h	8
DNAJC2	1.17	0.89	1.44	1.62	0.70	3.4E-15	5.8E-13	2h 3h 4h	10
KTN1	2.33	1.90	2.76	1.45	0.54	3.9E-18	8.0E-16	2h 3h 4h	14
ZFHX3	1.78	1.38	2.17	1.57	0.65	1.3E-19	3.2E-17	2h 3h 4h	10
RPL39	8.54	10.09	6.99	0.69	-0.53	2.1E-63	2.5E-60	2h 3h 4h	8
MAP1B	7.22	5.48	8.96	1.63	0.71	1.8E-88	5.4E-85	2h 3h 4h	10
TAF3	0.39	0.27	0.50	1.87	0.91	5.9E-09	4.8E-07	2h 3h	14
NIPBL	0.83	0.66	1.01	1.53	0.62	3.3E-09	2.8E-07	2h 3h	10
CHD7	1.06	0.85	1.26	1.49	0.57	5.6E-10	5.3E-08	2h 3h	10
CDK6	0.63	0.48	0.78	1.62	0.70	8.8E-09	6.9E-07	2h	10
KIF20B	1.33	1.09	1.56	1.43	0.51	4.3E-10	4.2E-08	2h	10
KIF5B	2.41	1.93	2.88	1.49	0.58	2.4E-21	7.0E-19	2h	10

Table S2. Differentially expressed genes between uninduced (0h, clusters 1&2) and 3h induced DIE cells (clusters 5&6).

* Adjusted p value < 10⁻⁶, absolute log2FC > 0.5

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
RPL37A	7.67	9.25	6.10	0.66	-0.60	1.5E-75	1.6E-72	All 4	8
UQCR11	1.02	1.25	0.79	0.63	-0.66	1.6E-13	1.7E-11	All 4	8
RPS29	11.76	14.47	9.04	0.63	-0.68	1.0E-145	6.0E-142	All 4	8
USMG5	1.80	2.27	1.33	0.58	-0.78	8.9E-31	2.6E-28	All 4	8
ANKRD11	2.16	1.50	2.81	1.87	0.90	5.6E-44	2.7E-41	All 4	10
ASPM	1.24	0.88	1.60	1.81	0.86	7.8E-24	1.5E-21	All 4	10
CENPF	2.83	2.07	3.60	1.74	0.80	5.3E-46	2.7E-43	All 4	10
CENPE	0.92	0.69	1.14	1.65	0.72	2.3E-13	2.2E-11	All 4	10
PRR11	1.00	0.79	1.21	1.53	0.61	5.6E-11	4.3E-09	All 4	13

Table S2 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
MDN1	0.33	0.21	0.45	2.13	1.09	1.1E-10	7.8E-09	All 4	14
BDP1	0.76	0.52	0.99	1.91	0.93	2.5E-17	3.6E-15	All 4	14
BRD4	1.56	1.09	2.03	1.86	0.89	8.9E-32	2.7E-29	All 4	14
AKAP9	0.73	0.52	0.94	1.81	0.85	2.1E-14	2.5E-12	All 4	14
NKTR	1.46	1.04	1.87	1.80	0.85	3.9E-27	9.4E-25	All 4	14
CHD2	0.70	0.51	0.90	1.77	0.83	3.5E-13	3.4E-11	All 4	14
GOLGA4	1.21	0.87	1.55	1.77	0.83	8.4E-22	1.5E-19	All 4	14
TOP1	1.58	1.19	1.97	1.66	0.73	3.6E-22	6.5E-20	All 4	14
BPTF	1.29	0.99	1.59	1.61	0.68	1.5E-16	2.1E-14	All 4	14
SRRM1	2.08	1.61	2.55	1.58	0.66	1.8E-24	3.6E-22	All 4	14
SMC4	5.84	4.82	6.86	1.42	0.51	6.8E-40	3.0E-37	All 4	14
SREK1	1.05	0.81	1.28	1.59	0.67	3.5E-13	3.4E-11	All 4	17
KCNQ10T1	8.73	5.03	12.44	2.47	1.31	0.0E+00	0.0E+00	All 4	18
UGDH-AS1	1.57	0.95	2.19	2.32	1.21	3.4E-54	2.0E-51	All 4	18
PGM5P2	0.80	0.51	1.10	2.17	1.12	6.7E-25	1.5E-22	All 4	18
MAB21L3	0.96	0.61	1.31	2.15	1.10	8.1E-29	2.3E-26	All 4	18
LOC100131257	2.36	1.58	3.14	1.98	0.99	3.0E-56	2.1E-53	All 4	18
PRPF38B	2.07	1.57	2.58	1.65	0.72	2.8E-28	7.2E-26	All 4	18
POLR2L	2.15	2.53	1.76	0.70	-0.52	2.3E-17	3.4E-15	3h 6h	8
ID1	3.17	4.19	2.14	0.51	-0.97	6.7E-80	8.0E-77	3h 4h 6h	3
ID3	1.54	2.09	0.98	0.47	-1.09	2.0E-49	1.2E-46	3h 4h 6h	3
FKBP10	0.41	0.52	0.29	0.55	-0.86	2.2E-09	1.3E-07	3h 4h 6h	7
COX7C	4.86	5.69	4.02	0.71	-0.50	1.1E-34	4.0E-32	3h 4h 6h	8
PRDX4	1.43	1.70	1.15	0.68	-0.56	1.2E-13	1.2E-11	3h 4h 6h	8
APP	2.32	2.86	1.78	0.63	-0.68	3.3E-30	9.4E-28	3h 4h 6h	8
MAGOH	0.71	0.87	0.55	0.64	-0.66	8.4E-10	5.4E-08	3h 4h 6h	9
ATRX	2.10	1.66	2.55	1.54	0.62	8.6E-22	1.5E-19	3h 4h 6h	14
CCAR1	3.46	2.75	4.18	1.52	0.61	1.7E-33	5.5E-31	3h 4h 6h	14
DMWD	1.23	1.01	1.44	1.43	0.51	1.3E-09	8.1E-08	3h 4h 6h	14
SLC4A7	0.59	0.45	0.74	1.64	0.71	6.4E-09	3.5E-07	3h 4h 6h	17
MLL5	0.67	0.52	0.83	1.60	0.68	4.5E-09	2.5E-07	3h 4h 6h	17
ZNF471	0.58	0.38	0.77	2.02	1.01	1.4E-15	1.8E-13	3h 4h 6h	18
F5	0.42	0.28	0.55	1.95	0.96	1.5E-10	1.0E-08	3h 4h 6h	18
TMEM212	0.46	0.32	0.60	1.87	0.90	1.8E-10	1.3E-08	3h 4h 6h	18
CCDC144B	0.46	0.33	0.59	1.80	0.85	2.0E-09	1.2E-07	3h 4h 6h	18

Table S2 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
TSIX	0.44	0.32	0.56	1.77	0.82	8.6E-09	4.6E-07	3h 4h 6h	18
BOD1L1	0.77	0.57	0.96	1.68	0.75	4.9E-12	4.3E-10	3h 4h 6h	18
GOLGB1	0.95	0.71	1.19	1.67	0.74	1.8E-14	2.3E-12	3h 4h 6h	18
LUC7L3	2.10	1.65	2.55	1.55	0.63	2.5E-22	4.6E-20	3h 4h 6h	18
RBM25	3.15	2.49	3.82	1.54	0.62	6.1E-32	1.9E-29	3h 4h 6h	18
MPHOSPH8	0.83	0.66	1.00	1.53	0.61	3.4E-09	1.9E-07	3h 4h 6h	18
GADD45A	1.72	1.37	2.07	1.51	0.59	8.3E-17	1.2E-14	3h 4h 6h	18
PNISR	1.94	1.58	2.31	1.47	0.55	1.7E-16	2.3E-14	3h 4h 6h	18
SLTM	1.25	1.03	1.47	1.42	0.51	1.1E-09	6.8E-08	3h 4h 6h	18
RFPL4B	0.46	0.26	0.65	2.49	1.31	2.2E-19	3.5E-17	3h 4h 6h	19
ZNF296	0.86	0.60	1.12	1.87	0.90	2.2E-18	3.3E-16	3h 4h 6h	19
RBBP6	3.70	2.59	4.81	1.86	0.89	1.2E-72	1.0E-69	3h 4h 6h	19
SRSF8	2.42	1.77	3.08	1.74	0.80	3.7E-39	1.6E-36	3h 4h 6h	19
ZMAT3	0.89	0.69	1.09	1.58	0.66	3.4E-11	2.6E-09	3h 4h 6h	19
ZNF217	0.81	0.64	0.97	1.52	0.60	6.1E-09	3.3E-07	3h 4h 6h	19
PNN	3.62	2.95	4.30	1.46	0.54	1.0E-28	2.8E-26	3h 4h 6h	21
MT2A	1.26	1.48	1.03	0.70	-0.52	1.2E-10	8.7E-09	3h 4h	7
SNRPD2	2.08	2.47	1.68	0.68	-0.56	3.6E-19	5.7E-17	3h 4h	7
TOMM7	1.48	1.79	1.17	0.65	-0.61	1.3E-16	1.8E-14	3h 4h	7
TMEM258	1.26	1.50	1.03	0.68	-0.55	1.1E-11	9.5E-10	3h 4h	8
ATP5O	2.52	3.02	2.01	0.67	-0.59	1.1E-24	2.3E-22	3h 4h	8
ATP5J2	3.47	4.21	2.74	0.65	-0.62	6.0E-37	2.4E-34	3h 4h	8
CLSPN	1.51	1.21	1.82	1.50	0.59	1.4E-14	1.8E-12	3h 4h	10
ZC3H13	0.52	0.39	0.65	1.66	0.74	2.0E-08	1.0E-06	3h 4h	14
CHD9	1.00	0.80	1.20	1.50	0.59	4.7E-10	3.2E-08	3h 4h	14
REV3L	0.99	0.80	1.19	1.49	0.58	5.9E-10	3.9E-08	3h 4h	14
THOC2	1.52	1.23	1.80	1.46	0.55	4.0E-13	3.8E-11	3h 4h	14
APH1A	0.29	0.39	0.19	0.49	-1.03	1.5E-09	9.0E-08	3h 4h	NA
PSMB3	1.10	1.30	0.90	0.70	-0.52	1.4E-09	8.2E-08	3h	7
NDUFA11	1.34	1.58	1.10	0.70	-0.53	2.1E-11	1.7E-09	3h	7
LAMTOR5	0.94	1.13	0.76	0.68	-0.56	1.7E-09	9.9E-08	3h	7
UQCR10	1.16	1.40	0.92	0.66	-0.60	1.1E-12	9.9E-11	3h	7
RPA3	0.51	0.63	0.39	0.61	-0.71	1.7E-08	8.7E-07	3h	7
UBE2T	1.10	1.29	0.91	0.71	-0.51	3.9E-09	2.2E-07	3h	8
RPL31	5.01	5.90	4.13	0.70	-0.51	4.7E-37	2.0E-34	3h	8



Table S2 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
RPL37	14.22	16.72	11.72	0.70	-0.51	6.6E-100	2.0E-96	3h	8
RPL28	7.98	9.43	6.54	0.69	-0.53	4.6E-61	3.7E-58	3h	8
UBL5	2.88	3.41	2.35	0.69	-0.54	7.8E-24	1.5E-21	3h	8
SNRPG	2.82	3.35	2.28	0.68	-0.56	8.4E-25	1.8E-22	3h	8
RPS28	6.63	7.93	5.33	0.67	-0.57	7.8E-60	5.8E-57	3h	8
ASH1L	0.75	0.56	0.94	1.67	0.74	9.8E-12	8.3E-10	3h	10
XIAP	0.82	0.64	1.00	1.55	0.63	9.1E-10	5.8E-08	3h	10
FAM115A	0.88	0.71	1.05	1.48	0.56	1.9E-08	9.5E-07	3h	10
REST	0.93	0.75	1.11	1.48	0.56	5.5E-09	3.0E-07	3h	10
COX7A2	2.82	3.36	2.29	0.68	-0.55	7.9E-25	1.7E-22	3h	12
TNRC6B	0.74	0.56	0.93	1.67	0.74	2.3E-11	1.8E-09	3h	14
TPR	0.82	0.65	0.99	1.51	0.60	7.6E-09	4.1E-07	3h	14
SETD2	0.87	0.70	1.03	1.48	0.57	1.8E-08	9.2E-07	3h	14
ESF1	0.92	0.74	1.10	1.48	0.56	7.7E-09	4.1E-07	3h	14
LRRC58	1.01	0.82	1.20	1.46	0.54	5.0E-09	2.8E-07	3h	14
WNK1	1.28	1.05	1.50	1.43	0.51	4.5E-10	3.1E-08	3h	14
KIF14	0.86	0.67	1.04	1.54	0.62	1.0E-09	6.7E-08	2h 3h 6h	10
HIST1H2BK	0.74	0.92	0.57	0.63	-0.68	9.8E-11	7.2E-09	2h 3h 6h	19
NOP10	1.70	2.06	1.33	0.65	-0.63	9.6E-20	1.6E-17	2h 3h 4h	8
RPL39	8.02	9.75	6.29	0.65	-0.63	4.8E-87	7.1E-84	2h 3h 4h	8
PIP5K1A	0.31	0.42	0.20	0.46	-1.11	1.9E-11	1.5E-09	2h 3h 4h	8
GOLIM4	0.44	0.29	0.58	1.99	0.99	1.1E-11	9.5E-10	2h 3h 4h	10
CCDC88A	0.65	0.46	0.84	1.83	0.87	2.1E-13	2.1E-11	2h 3h 4h	10
DNAJC2	1.15	0.86	1.44	1.67	0.74	2.7E-17	3.9E-15	2h 3h 4h	10
MAP1B	7.04	5.30	8.79	1.66	0.73	4.0E-94	8.0E-91	2h 3h 4h	10
ITSN1	1.13	0.86	1.41	1.65	0.72	6.9E-16	9.2E-14	2h 3h 4h	10
CDC42BPA	0.73	0.56	0.91	1.62	0.70	2.6E-10	1.8E-08	2h 3h 4h	10
ZFHX3	1.73	1.34	2.13	1.59	0.67	6.8E-21	1.1E-18	2h 3h 4h	10
ANKRD12	0.81	0.61	1.00	1.65	0.73	7.2E-12	6.2E-10	2h 3h 4h	13
ROCK1	0.54	0.36	0.72	1.98	0.99	4.2E-14	4.8E-12	2h 3h 4h	14
KTN1	2.34	1.84	2.85	1.55	0.63	5.4E-25	1.2E-22	2h 3h 4h	14
CHD7	1.04	0.82	1.26	1.54	0.62	1.3E-11	1.1E-09	2h 3h	10
NIPBL	0.80	0.64	0.97	1.52	0.60	1.2E-08	6.4E-07	2h 3h	10
TAF3	0.38	0.26	0.49	1.88	0.91	5.2E-09	2.9E-07	2h 3h	14

Table S3. Differentially expressed genes between uninduced (0h, clusters 1&2) and 4h induced DIE cells (clusters 7&8).

* Adjusted p value < 10**⁻⁶, absolute log₂FC > 0.5

Gene	base Mean	base MeanA	base MeanB	fold Change	log ₂ FC	pval	padj	shared in states	Node
RPL37A	7.63	9.21	6.06	0.66	-0.61	8.0E-58	4.8E-55	All 4	8
RPS29	11.81	14.41	9.21	0.64	-0.65	4.3E-96	5.2E-93	All 4	8
USMG5	1.84	2.27	1.42	0.63	-0.67	3.5E-18	4.8E-16	All 4	8
UQCR11	1.01	1.24	0.77	0.62	-0.70	1.5E-11	1.1E-09	All 4	8
ASPM	1.25	0.88	1.61	1.83	0.87	1.2E-19	1.8E-17	All 4	10
ANKRD11	2.03	1.50	2.57	1.72	0.78	9.2E-26	2.0E-23	All 4	10
CENPE	0.89	0.69	1.09	1.58	0.66	3.3E-09	1.7E-07	All 4	10
CENPF	2.61	2.06	3.16	1.53	0.62	3.3E-21	5.7E-19	All 4	10
PRR11	1.00	0.79	1.21	1.53	0.62	5.1E-09	2.6E-07	All 4	13
MDN1	0.34	0.21	0.47	2.25	1.17	2.9E-10	1.8E-08	All 4	14
BDP1	0.78	0.52	1.04	2.00	1.00	3.9E-16	4.3E-14	All 4	14
AKAP9	0.78	0.52	1.03	2.00	1.00	2.6E-16	3.0E-14	All 4	14
NKTR	1.47	1.04	1.91	1.84	0.88	1.4E-23	2.7E-21	All 4	14
BRD4	1.54	1.09	1.98	1.82	0.87	1.1E-23	2.1E-21	All 4	14
CHD2	0.69	0.50	0.88	1.75	0.81	4.4E-10	2.6E-08	All 4	14
GOLGA4	1.18	0.87	1.50	1.72	0.78	1.2E-15	1.3E-13	All 4	14
BPTF	1.34	0.99	1.69	1.72	0.78	2.4E-17	3.1E-15	All 4	14
SRRM1	2.08	1.60	2.56	1.60	0.68	2.4E-20	4.0E-18	All 4	14
TOP1	1.54	1.19	1.89	1.59	0.67	3.2E-15	3.3E-13	All 4	14
SMC4	5.84	4.80	6.89	1.43	0.52	3.9E-33	1.2E-30	All 4	14
SREK1	1.10	0.80	1.39	1.72	0.79	9.8E-15	9.6E-13	All 4	17
KCNQ10T1	10.12	5.01	15.22	3.04	1.60	0.0E+00	0.0E+00	All 4	18
UGDH-AS1	1.83	0.94	2.73	2.89	1.53	1.1E-76	1.1E-73	All 4	18
PGM5P2	0.92	0.51	1.34	2.64	1.40	9.3E-34	3.0E-31	All 4	18
MAB21L3	1.11	0.61	1.60	2.63	1.39	5.6E-40	2.2E-37	All 4	18
LOC100131257	2.73	1.58	3.88	2.47	1.30	3.3E-85	3.6E-82	All 4	18
PRPF38B	2.24	1.56	2.92	1.87	0.90	1.3E-36	4.6E-34	All 4	18
SLC7A5	0.87	1.08	0.66	0.61	-0.71	2.6E-10	1.6E-08	4h 6h	1
FTL	9.12	10.70	7.54	0.71	-0.50	1.3E-48	5.7E-46	4h 6h	2
UBE2S	2.92	3.50	2.35	0.67	-0.58	3.7E-21	6.4E-19	4h 6h	3
DYNLL1	6.98	8.57	5.39	0.63	-0.67	1.9E-64	1.3E-61	4h 6h	3
SHISA3	1.55	1.92	1.18	0.61	-0.71	5.3E-17	6.6E-15	4h 6h	3

Table S3 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
NEDD9	0.88	1.09	0.67	0.61	-0.71	2.5E-10	1.5E-08	4h 6h	3
MIDN	0.97	1.23	0.72	0.58	-0.78	3.9E-13	3.2E-11	4h 6h	3
CYR61	1.18	1.50	0.86	0.57	-0.81	7.1E-17	8.7E-15	4h 6h	3
PCDH18	0.57	0.76	0.39	0.51	-0.97	3.0E-12	2.4E-10	4h 6h	3
PIM1	0.38	0.50	0.25	0.51	-0.98	2.2E-08	9.9E-07	4h 6h	3
TGIF1	0.68	0.92	0.44	0.48	-1.05	3.1E-16	3.5E-14	4h 6h	3
NOG	0.31	0.43	0.20	0.47	-1.08	1.9E-08	8.7E-07	4h 6h	3
NUAK2	0.36	0.50	0.23	0.46	-1.12	4.7E-10	2.7E-08	4h 6h	3
HNRNPA0	1.61	1.94	1.28	0.66	-0.61	1.8E-13	1.5E-11	4h 6h	8
RPL23A	10.89	12.91	8.87	0.69	-0.54	7.1E-65	5.0E-62	4h 6h	9
ISOC2	0.43	0.56	0.30	0.53	-0.92	1.8E-08	8.6E-07	4h 6h	9
RIF1	1.52	1.21	1.83	1.51	0.59	4.4E-12	3.4E-10	4h 6h	13
GUSBP3	0.53	0.37	0.69	1.87	0.91	7.1E-10	4.0E-08	4h 6h	14
CUX1	0.97	0.77	1.18	1.53	0.61	1.2E-08	5.7E-07	4h 6h	14
BBX	1.48	1.19	1.76	1.48	0.56	7.7E-11	5.2E-09	4h 6h	14
PPIG	1.67	1.36	1.99	1.47	0.55	1.2E-11	9.0E-10	4h 6h	14
SYNE2	0.75	0.50	0.99	1.98	0.99	3.9E-15	3.9E-13	4h 6h	17
POLQ	0.85	0.64	1.06	1.66	0.74	1.5E-10	9.7E-09	4h 6h	17
SMC3	1.57	1.27	1.87	1.47	0.56	3.6E-11	2.5E-09	4h 6h	17
TNRC6A	1.61	1.32	1.91	1.45	0.54	1.1E-10	7.3E-09	4h 6h	17
ODF2L	0.55	0.38	0.72	1.91	0.93	1.0E-10	6.7E-09	4h 6h	18
GABPB1-AS1	0.51	0.36	0.66	1.87	0.90	1.7E-09	9.3E-08	4h 6h	18
PHACTR2	0.54	0.38	0.70	1.87	0.90	6.5E-10	3.6E-08	4h 6h	18
NET1	1.09	0.83	1.35	1.63	0.71	3.6E-12	2.8E-10	4h 6h	18
ZMYND8	0.99	0.76	1.21	1.60	0.67	2.9E-10	1.8E-08	4h 6h	18
SLC25A36	1.43	1.16	1.70	1.47	0.56	2.4E-10	1.4E-08	4h 6h	18
SRSF11	4.36	3.54	5.18	1.46	0.55	7.0E-28	1.8E-25	4h 6h	18
DPPA4	4.12	3.39	4.85	1.43	0.52	1.8E-23	3.4E-21	4h 6h	18
ZRANB2	2.49	2.05	2.93	1.43	0.52	9.3E-15	9.2E-13	4h 6h	18
ZSCAN4	1.16	0.26	2.06	7.90	2.98	3.1E-131	6.2E-128	4h 6h	19
TRIM51	0.74	0.25	1.23	4.96	2.31	3.1E-60	2.0E-57	4h 6h	19
PRAMEF1	0.27	0.11	0.44	3.80	1.93	9.3E-18	1.3E-15	4h 6h	19
PRRG4	0.37	0.17	0.58	3.36	1.75	6.1E-21	1.1E-18	4h 6h	19
PRAMEF12	0.27	0.13	0.41	3.14	1.65	3.0E-14	2.8E-12	4h 6h	19

Table S3 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
SLC34A2	0.23	0.12	0.35	2.89	1.53	9.2E-11	6.1E-09	4h 6h	19
KDM4E	0.20	0.10	0.29	2.75	1.46	7.4E-09	3.7E-07	4h 6h	19
GTF2F1	2.55	1.36	3.73	2.74	1.45	7.9E-97	1.1E-93	4h 6h	19
ARID5B	0.57	0.31	0.83	2.70	1.44	4.2E-22	7.7E-20	4h 6h	19
PNP	2.84	1.57	4.11	2.61	1.39	5.2E-99	7.8E-96	4h 6h	19
SPTY2D1	0.55	0.31	0.79	2.53	1.34	1.9E-19	2.9E-17	4h 6h	19
TFIP11	0.31	0.18	0.44	2.47	1.31	7.0E-11	4.8E-09	4h 6h	19
ESRG	0.45	0.27	0.63	2.36	1.24	2.7E-14	2.5E-12	4h 6h	19
HOXB2	1.29	0.78	1.80	2.29	1.20	9.8E-36	3.4E-33	4h 6h	19
ZNF622	0.71	0.44	0.97	2.20	1.14	1.1E-18	1.5E-16	4h 6h	19
PDGFRA	0.47	0.30	0.64	2.16	1.11	4.7E-12	3.6E-10	4h 6h	19
CDH10	0.76	0.50	1.01	2.01	1.00	4.8E-16	5.2E-14	4h 6h	19
DBR1	0.58	0.40	0.76	1.92	0.94	2.6E-11	1.9E-09	4h 6h	19
NFAT5	0.78	0.54	1.03	1.90	0.93	2.3E-14	2.2E-12	4h 6h	19
ARHGGEF26	1.07	0.74	1.39	1.88	0.91	2.2E-18	3.2E-16	4h 6h	19
EXOSC10	2.11	1.47	2.75	1.86	0.90	2.8E-34	9.2E-32	4h 6h	19
ZNF644	0.78	0.56	1.00	1.77	0.82	1.1E-11	8.2E-10	4h 6h	19
NXF1	1.38	1.04	1.73	1.67	0.74	1.6E-16	1.9E-14	4h 6h	19
SHC1	1.34	1.04	1.65	1.60	0.67	1.4E-13	1.2E-11	4h 6h	19
MRPL49	1.82	1.41	2.23	1.58	0.66	6.5E-17	8.0E-15	4h 6h	19
PLK4	0.93	0.72	1.13	1.57	0.65	3.3E-09	1.7E-07	4h 6h	19
CCNL2	1.29	1.05	1.53	1.45	0.54	4.9E-09	2.5E-07	4h 6h	19
CIRBP	3.89	3.18	4.60	1.45	0.53	1.9E-23	3.5E-21	4h 6h	19
RFPL4A	0.38	0.15	0.62	4.24	2.08	1.1E-27	2.8E-25	4h 6h	20
LEUTX	0.30	0.14	0.45	3.35	1.74	1.4E-16	1.7E-14	4h 6h	20
RICTOR	0.64	0.39	0.88	2.25	1.17	1.3E-17	1.7E-15	4h 6h	21
ITGB8	0.31	0.20	0.42	2.16	1.11	1.9E-08	8.7E-07	4h 6h	21
ZNF827	0.48	0.31	0.64	2.06	1.04	3.2E-11	2.2E-09	4h 6h	21
ZSWIM6	0.48	0.34	0.62	1.84	0.88	2.1E-08	9.5E-07	4h 6h	21
BTAF1	0.66	0.49	0.83	1.70	0.76	7.0E-09	3.5E-07	4h 6h	21
KIAA1551	0.75	0.56	0.94	1.69	0.76	1.1E-09	6.2E-08	4h 6h	21
YTHDC1	1.29	0.98	1.60	1.63	0.71	2.7E-14	2.5E-12	4h 6h	21
LRRRC8B	0.89	0.68	1.09	1.62	0.70	5.7E-10	3.2E-08	4h 6h	21
PUM1	2.15	1.76	2.54	1.44	0.53	1.3E-13	1.2E-11	4h 6h	21
CTGF	0.46	0.59	0.32	0.54	-0.90	7.0E-09	3.5E-07	4h	3



Table S3 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
CKS2	1.58	1.87	1.28	0.69	-0.55	5.8E-11	4.0E-09	4h	7
ID2	0.41	0.58	0.24	0.42	-1.24	1.3E-13	1.2E-11	4h	7
LRRFIP1	1.62	1.27	1.96	1.55	0.63	3.3E-14	3.0E-12	4h	10
LIMS1	0.75	0.56	0.94	1.67	0.74	1.6E-09	8.7E-08	4h	14
LOC642236	0.36	0.22	0.50	2.25	1.17	2.2E-10	1.4E-08	4h	17
PCLO	0.40	0.27	0.54	2.02	1.02	3.7E-09	1.9E-07	4h	17
BAZ2B	0.39	0.27	0.52	1.95	0.96	2.0E-08	9.1E-07	4h	17
ASCC3	0.81	0.62	0.99	1.60	0.68	9.2E-09	4.5E-07	4h	17
HELLS	0.81	0.62	0.99	1.58	0.66	1.8E-08	8.5E-07	4h	17
PAXBP1	1.22	0.95	1.48	1.56	0.64	3.1E-11	2.2E-09	4h	17
CDR2	0.99	0.78	1.20	1.53	0.61	8.7E-09	4.2E-07	4h	17
GLI3	0.97	0.77	1.17	1.52	0.61	1.2E-08	5.7E-07	4h	17
LRRN3	1.25	1.00	1.50	1.50	0.58	5.4E-10	3.1E-08	4h	17
ZFHX4	0.87	0.53	1.20	2.27	1.18	9.6E-24	1.9E-21	4h	21
ID3	1.26	2.09	0.44	0.21	-2.26	4.0E-104	6.9E-101	3h 4h 6h	3
ID1	2.42	4.18	0.67	0.16	-2.65	8.8E-249	3.5E-245	3h 4h 6h	3
FKBP10	0.38	0.52	0.24	0.47	-1.10	2.0E-10	1.2E-08	3h 4h 6h	7
COX7C	4.82	5.67	3.96	0.70	-0.52	1.1E-27	2.7E-25	3h 4h 6h	8
PRDX4	1.43	1.69	1.17	0.69	-0.53	1.0E-09	5.6E-08	3h 4h 6h	8
APP	2.24	2.84	1.64	0.58	-0.80	5.0E-30	1.4E-27	3h 4h 6h	8
MAGOH	0.68	0.87	0.49	0.57	-0.82	1.8E-10	1.1E-08	3h 4h 6h	9
ATRX	2.19	1.65	2.72	1.65	0.72	8.9E-24	1.8E-21	3h 4h 6h	14
CCAR1	3.55	2.74	4.37	1.60	0.68	2.1E-33	6.6E-31	3h 4h 6h	14
DMWD	1.26	1.01	1.50	1.49	0.58	7.1E-10	4.0E-08	3h 4h 6h	14
MLL5	0.83	0.52	1.15	2.23	1.15	3.2E-22	6.0E-20	3h 4h 6h	17
SLC4A7	0.70	0.45	0.96	2.13	1.09	3.5E-17	4.4E-15	3h 4h 6h	17
F5	0.50	0.28	0.72	2.52	1.34	8.4E-18	1.1E-15	3h 4h 6h	18
ZNF471	0.65	0.38	0.92	2.43	1.28	6.7E-21	1.1E-18	3h 4h 6h	18
TMEM212	0.52	0.32	0.72	2.24	1.17	1.1E-14	1.0E-12	3h 4h 6h	18
CCDC144B	0.52	0.33	0.71	2.20	1.14	6.9E-14	6.3E-12	3h 4h 6h	18
TSIX	0.49	0.32	0.67	2.11	1.08	3.0E-12	2.4E-10	3h 4h 6h	18
GOLGB1	1.03	0.71	1.36	1.92	0.94	2.7E-19	4.1E-17	3h 4h 6h	18
MPHOSPH8	0.91	0.65	1.16	1.77	0.83	1.1E-13	9.6E-12	3h 4h 6h	18
BOD1L1	0.79	0.57	1.01	1.76	0.82	8.0E-12	6.1E-10	3h 4h 6h	18

Table S3 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
RBM25	3.42	2.48	4.36	1.76	0.82	1.5E-45	6.2E-43	3h 4h 6h	18
LUC7L3	2.19	1.65	2.73	1.66	0.73	2.6E-24	5.5E-22	3h 4h 6h	18
PNISR	2.04	1.57	2.52	1.60	0.68	4.2E-20	6.9E-18	3h 4h 6h	18
GADD45A	1.74	1.37	2.10	1.54	0.62	7.3E-15	7.3E-13	3h 4h 6h	18
SLTM	1.28	1.03	1.54	1.50	0.59	3.9E-10	2.3E-08	3h 4h 6h	18
RFPL4B	1.70	0.26	3.14	11.94	3.58	3.6E-234	8.5E-231	3h 4h 6h	19
RBBP6	7.77	2.58	12.96	5.02	2.33	0.0E+00	0.0E+00	3h 4h 6h	19
SRSF8	4.16	1.77	6.55	3.71	1.89	5.4E-243	1.6E-239	3h 4h 6h	19
ZNF217	1.43	0.64	2.22	3.47	1.80	6.5E-78	6.5E-75	3h 4h 6h	19
ZNF296	1.29	0.60	1.99	3.33	1.74	4.0E-67	3.0E-64	3h 4h 6h	19
ZMAT3	1.01	0.69	1.33	1.94	0.95	4.0E-19	6.0E-17	3h 4h 6h	19
PNN	4.26	2.94	5.57	1.90	0.92	3.3E-71	2.6E-68	3h 4h 6h	21
MT2A	1.25	1.47	1.03	0.70	-0.52	2.2E-08	9.8E-07	3h 4h	7
SNRPD2	2.08	2.47	1.69	0.69	-0.54	8.0E-14	7.2E-12	3h 4h	7
TOMM7	1.49	1.79	1.19	0.67	-0.58	1.2E-11	8.8E-10	3h 4h	7
ATP5O	2.53	3.01	2.04	0.68	-0.56	1.9E-17	2.4E-15	3h 4h	8
TMEM258	1.25	1.49	1.01	0.68	-0.57	1.4E-09	7.5E-08	3h 4h	8
ATP5J2	3.49	4.19	2.79	0.67	-0.58	1.7E-25	3.6E-23	3h 4h	8
CLSPN	1.58	1.20	1.95	1.62	0.69	2.1E-16	2.5E-14	3h 4h	10
ZC3H13	0.55	0.39	0.70	1.80	0.85	4.4E-09	2.3E-07	3h 4h	14
CHD9	1.04	0.80	1.28	1.60	0.68	8.2E-11	5.5E-09	3h 4h	14
REV3L	1.00	0.79	1.20	1.52	0.60	1.2E-08	5.7E-07	3h 4h	14
THOC2	1.51	1.23	1.79	1.46	0.55	1.9E-10	1.1E-08	3h 4h	14
APH1A	0.28	0.39	0.18	0.46	-1.12	1.6E-08	7.7E-07	3h 4h	NA
FOXN3	1.17	0.93	1.40	1.50	0.59	2.0E-09	1.1E-07	2h 4h	10
NOP10	1.74	2.06	1.42	0.69	-0.54	1.1E-11	8.2E-10	2h 3h 4h	8
RPL39	8.10	9.71	6.50	0.67	-0.58	2.4E-56	1.4E-53	2h 3h 4h	8
PIP5K1A	0.30	0.42	0.18	0.43	-1.21	4.6E-10	2.7E-08	2h 3h 4h	8
GOLIM4	0.44	0.29	0.59	2.02	1.01	7.8E-10	4.4E-08	2h 3h 4h	10
CCDC88A	0.64	0.46	0.83	1.81	0.86	1.4E-10	8.8E-09	2h 3h 4h	10
ITSN1	1.16	0.85	1.46	1.72	0.78	2.8E-15	2.9E-13	2h 3h 4h	10
CDC42BPA	0.73	0.56	0.91	1.63	0.71	1.0E-08	4.9E-07	2h 3h 4h	10
ZFHX3	1.70	1.33	2.06	1.55	0.63	7.3E-15	7.3E-13	2h 3h 4h	10
MAP1B	6.64	5.28	8.01	1.52	0.60	3.2E-49	1.5E-46	2h 3h 4h	10



Table S3 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
DNAJC2	1.08	0.86	1.30	1.52	0.60	3.2E-09	1.7E-07	2h 3h 4h	10
ANKRD12	0.85	0.61	1.09	1.81	0.85	1.8E-13	1.5E-11	2h 3h 4h	13
ROCK1	0.54	0.36	0.72	1.99	0.99	1.5E-11	1.1E-09	2h 3h 4h	14
KTN1	2.33	1.83	2.84	1.55	0.63	8.4E-20	1.4E-17	2h 3h 4h	14

Table S4. Differentially expressed genes between uninduced (0h, clusters 1&2) and 6h induced DIE cells (clusters 9&10).

* Adjusted p value < 10⁻⁶, absolute log2FC > 0.5

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
UGDH-AS1	2.27	0.95	3.60	3.81	1.93	4.4E-119	2.5E-116	All 4	18
KCNQ1OT1	10.90	5.01	16.79	3.35	1.74	0	0	All 4	18
LOC100131257	3.28	1.58	4.99	3.16	1.66	1.7E-133	1.2E-130	All 4	18
MAB21L3	1.21	0.61	1.82	2.98	1.58	2.4E-46	6.6E-44	All 4	18
PGM5P2	0.93	0.51	1.36	2.69	1.43	2.5E-30	3.7E-28	All 4	18
AKAP9	0.89	0.52	1.26	2.43	1.28	4.6E-24	5.5E-22	All 4	14
MDN1	0.35	0.21	0.49	2.30	1.20	2.1E-09	7.6E-08	All 4	14
ASPM	1.42	0.88	1.97	2.23	1.16	1.3E-31	2.0E-29	All 4	10
PRPF38B	2.47	1.56	3.37	2.16	1.11	2.3E-49	6.4E-47	All 4	18
BDP1	0.79	0.52	1.06	2.04	1.03	9.8E-15	6.6E-13	All 4	14
CHD2	0.75	0.50	1.00	1.98	0.99	3.7E-13	2.2E-11	All 4	14
SREK1	1.16	0.81	1.51	1.88	0.91	2.9E-17	2.4E-15	All 4	17
BRD4	1.52	1.09	1.95	1.79	0.84	4.2E-19	3.9E-17	All 4	14
CENPE	0.96	0.69	1.23	1.78	0.83	2.2E-12	1.2E-10	All 4	10
BPTF	1.36	0.99	1.73	1.75	0.81	4.4E-16	3.3E-14	All 4	14
GOLGA4	1.18	0.87	1.49	1.71	0.77	4.4E-13	2.6E-11	All 4	14
PRR11	1.05	0.79	1.32	1.67	0.74	8.7E-11	3.8E-09	All 4	13
NKTR	1.36	1.04	1.69	1.63	0.70	1.0E-12	5.6E-11	All 4	14
SRRM1	2.10	1.60	2.60	1.62	0.70	2.1E-18	1.9E-16	All 4	14
CENPF	2.67	2.06	3.27	1.59	0.67	3.2E-21	3.3E-19	All 4	10
TOP1	1.52	1.19	1.86	1.56	0.64	5.1E-12	2.7E-10	All 4	14
ANKRD11	1.88	1.50	2.27	1.51	0.60	1.2E-12	6.8E-11	All 4	10
SMC4	5.84	4.81	6.88	1.43	0.52	1.2E-27	1.6E-25	All 4	14
RPL37A	7.68	9.22	6.14	0.67	-0.59	1.8E-44	4.7E-42	All 4	8

Table S4 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
USMG5	1.87	2.27	1.48	0.65	-0.62	1.9E-13	1.2E-11	All 4	8
RPS29	11.82	14.43	9.21	0.64	-0.65	3.9E-68	1.7E-65	All 4	8
UQCR11	0.99	1.25	0.74	0.59	-0.76	1.2E-10	5.2E-09	All 4	8
CCNA1	1.88	0.22	3.54	16.37	4.03	3.9E-256	4.2E-253	6h	20
LOC441081	0.89	0.13	1.66	13.20	3.72	4.6E-113	2.3E-110	6h	20
RFPL1	0.53	0.11	0.95	8.90	3.15	3.7E-57	1.2E-54	6h	20
LINC00633	0.65	0.15	1.15	7.82	2.97	6.1E-64	2.3E-61	6h	19
KHDC1L	0.45	0.11	0.79	7.38	2.88	4.8E-44	1.2E-41	6h	20
ALPPL2	0.44	0.11	0.78	7.26	2.86	5.6E-43	1.3E-40	6h	20
RFPL2	0.39	0.10	0.68	6.67	2.74	1.1E-35	2.2E-33	6h	20
SIAH1	1.40	0.37	2.44	6.67	2.74	5.3E-124	3.3E-121	6h	20
PLXNB3	0.53	0.15	0.91	6.26	2.65	1.2E-45	3.1E-43	6h	20
ART3	0.32	0.11	0.53	4.95	2.31	5.9E-24	6.9E-22	6h	19
SNAI1	0.38	0.14	0.63	4.66	2.22	3.5E-26	4.6E-24	6h	19
TRIM48	0.27	0.10	0.43	4.18	2.06	7.0E-17	5.6E-15	6h	20
RHOBTB1	0.49	0.20	0.78	3.89	1.96	2.0E-27	2.7E-25	6h	19
SAMD8	0.60	0.25	0.96	3.84	1.94	3.1E-33	5.1E-31	6h	19
DUSP18	0.35	0.15	0.55	3.80	1.93	1.3E-19	1.3E-17	6h	19
C1orf63	0.77	0.32	1.22	3.79	1.92	3.1E-41	7.0E-39	6h	19
EOMES	0.25	0.11	0.39	3.58	1.84	1.5E-13	9.7E-12	6h	19
MKRN9P	0.23	0.10	0.35	3.49	1.80	9.1E-12	4.6E-10	6h	19
MFSD11	0.49	0.22	0.77	3.44	1.78	5.1E-24	6.0E-22	6h	19
RIT2	0.24	0.11	0.37	3.40	1.77	2.6E-12	1.4E-10	6h	19
PHOX2B	0.24	0.11	0.36	3.14	1.65	3.6E-11	1.7E-09	6h	NA
TRIM49B	0.22	0.11	0.33	3.12	1.64	4.2E-10	1.6E-08	6h	19
GRAMD1C	0.27	0.14	0.41	3.02	1.59	1.1E-11	5.6E-10	6h	19
PANX2	0.69	0.34	1.03	3.02	1.59	1.9E-27	2.6E-25	6h	19
C3orf80	0.33	0.17	0.49	2.96	1.57	2.6E-13	1.6E-11	6h	19
ZNHIT6	1.33	0.67	1.98	2.96	1.57	7.0E-50	2.0E-47	6h	19
BHLHE22	0.26	0.14	0.39	2.93	1.55	9.4E-11	4.1E-09	6h	19
PRSS23	0.38	0.20	0.57	2.89	1.53	4.5E-15	3.2E-13	6h	19
DPPA3	0.21	0.11	0.31	2.87	1.52	2.0E-08	6.3E-07	6h	19
NKIRAS1	0.42	0.22	0.63	2.87	1.52	5.8E-16	4.3E-14	6h	19
LOC100188947	1.95	1.03	2.88	2.80	1.49	3.2E-66	1.3E-63	6h	19

Table S4 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
ADPGK	1.82	0.97	2.67	2.76	1.47	1.7E-60	5.9E-58	6h	19
ZFHX4	0.99	0.53	1.44	2.72	1.45	9.0E-33	1.4E-30	6h	21
C9orf66	0.21	0.11	0.31	2.70	1.43	3.3E-08	9.8E-07	6h	19
C1D	0.74	0.40	1.08	2.67	1.41	3.8E-24	4.5E-22	6h	19
NDEL1	0.42	0.23	0.61	2.64	1.40	6.8E-14	4.4E-12	6h	19
LRRK1	0.39	0.22	0.57	2.62	1.39	4.2E-13	2.5E-11	6h	21
ANK3	0.61	0.34	0.88	2.58	1.37	3.7E-19	3.5E-17	6h	19
DNM3	0.29	0.16	0.42	2.57	1.36	1.2E-09	4.4E-08	6h	19
OSR2	0.27	0.15	0.38	2.54	1.34	7.5E-09	2.5E-07	6h	19
AVP1	0.35	0.20	0.49	2.48	1.31	1.7E-10	7.0E-09	6h	19
CLK1	0.56	0.33	0.80	2.47	1.30	2.5E-16	1.9E-14	6h	19
MAST1	0.34	0.20	0.48	2.41	1.27	4.3E-10	1.7E-08	6h	19
STK17B	0.31	0.18	0.43	2.39	1.26	6.6E-09	2.2E-07	6h	21
TAF4B	0.40	0.24	0.57	2.38	1.25	4.0E-11	1.9E-09	6h	21
PELI2	0.64	0.38	0.90	2.35	1.23	1.2E-16	9.4E-15	6h	21
EPHA4	0.31	0.19	0.44	2.34	1.23	6.6E-09	2.2E-07	6h	19
HEXIM1	0.64	0.38	0.89	2.31	1.21	4.9E-16	3.6E-14	6h	20
CNNM4	0.48	0.29	0.67	2.30	1.20	3.4E-12	1.8E-10	6h	19
PRELP	0.40	0.24	0.55	2.26	1.18	3.9E-10	1.5E-08	6h	19
ZNF574	0.87	0.53	1.20	2.25	1.17	2.5E-20	2.5E-18	6h	19
ACAP2	0.51	0.32	0.71	2.24	1.17	1.3E-12	7.4E-11	6h	19
RGS2	0.65	0.40	0.89	2.24	1.17	1.6E-15	1.1E-13	6h	19
KIAA1217	0.44	0.27	0.61	2.24	1.16	7.1E-11	3.2E-09	6h	21
DNAJC25	0.51	0.32	0.71	2.23	1.16	3.0E-12	1.6E-10	6h	21
TPMT	0.45	0.28	0.62	2.22	1.15	6.7E-11	3.0E-09	6h	19
NCOA7	0.36	0.23	0.50	2.21	1.15	7.5E-09	2.5E-07	6h	19
SUPT6H	1.21	0.76	1.67	2.21	1.15	8.6E-27	1.1E-24	6h	19
C21orf91	0.59	0.37	0.81	2.19	1.13	2.4E-13	1.5E-11	6h	19
ALG13	0.87	0.55	1.20	2.19	1.13	3.3E-19	3.1E-17	6h	19
EPN2	0.35	0.22	0.48	2.19	1.13	1.9E-08	5.9E-07	6h	19
MGC21881	0.45	0.28	0.61	2.19	1.13	1.8E-10	7.5E-09	6h	19
USP3	0.88	0.56	1.21	2.16	1.11	6.8E-19	6.2E-17	6h	19
RP1-177G6.2	0.44	0.28	0.60	2.16	1.11	3.6E-10	1.4E-08	6h	18
C5orf44	0.66	0.42	0.90	2.16	1.11	2.8E-14	1.8E-12	6h	19
ZNF480	0.85	0.54	1.16	2.15	1.10	7.6E-18	6.4E-16	6h	21

Table S4 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
OXR1	0.47	0.30	0.63	2.13	1.09	3.5E-10	1.4E-08	6h	19
G2E3	0.76	0.49	1.04	2.12	1.09	6.9E-16	5.0E-14	6h	21
LOC100131067	0.41	0.27	0.56	2.10	1.07	1.0E-08	3.3E-07	6h	21
SHQ1	0.47	0.31	0.64	2.09	1.06	6.2E-10	2.4E-08	6h	19
BRCA2	0.37	0.24	0.51	2.08	1.05	2.9E-08	8.9E-07	6h	14
TIPARP	0.97	0.63	1.31	2.07	1.05	1.1E-18	9.5E-17	6h	19
TFAP2C	0.42	0.28	0.57	2.04	1.03	1.1E-08	3.4E-07	6h	19
STIL	0.57	0.38	0.76	2.02	1.02	9.3E-11	4.0E-09	6h	19
BIRC2	0.49	0.32	0.65	2.02	1.02	2.4E-09	8.6E-08	6h	19
ZNF91	0.46	0.30	0.61	2.00	1.00	9.1E-09	3.0E-07	6h	18
FBXO33	0.58	0.38	0.77	2.00	1.00	1.2E-10	4.9E-09	6h	19
SCAPER	0.48	0.32	0.64	2.00	1.00	4.4E-09	1.5E-07	6h	18
EPM2AIP1	0.83	0.56	1.11	1.99	0.99	9.6E-15	6.6E-13	6h	19
LUZP1	0.51	0.34	0.67	1.98	0.99	1.9E-09	7.0E-08	6h	18
UFL1	0.64	0.43	0.85	1.98	0.98	1.7E-11	8.3E-10	6h	21
TOPORS	0.71	0.48	0.94	1.95	0.96	5.2E-12	2.7E-10	6h	21
TMEM185A	0.52	0.36	0.69	1.94	0.95	3.7E-09	1.3E-07	6h	19
CCNL1	0.78	0.53	1.02	1.92	0.94	9.6E-13	5.4E-11	6h	21
TCEB3	1.07	0.73	1.40	1.92	0.94	7.3E-17	5.7E-15	6h	20
MLL3	0.50	0.34	0.66	1.91	0.93	1.7E-08	5.3E-07	6h	18
PNPLA8	0.55	0.38	0.72	1.90	0.93	4.6E-09	1.6E-07	6h	19
NOTCH2	0.57	0.39	0.74	1.90	0.92	2.9E-09	1.0E-07	6h	21
ELOF1	1.54	1.07	2.02	1.89	0.92	1.3E-22	1.4E-20	6h	19
ZNF281	0.66	0.46	0.87	1.88	0.91	1.7E-10	7.0E-09	6h	21
CASP6	0.67	0.47	0.88	1.87	0.90	2.5E-10	1.0E-08	6h	19
RPP14	1.24	0.87	1.62	1.86	0.90	8.2E-18	6.8E-16	6h	19
CLCN3	0.83	0.58	1.08	1.86	0.90	3.7E-12	1.9E-10	6h	19
TRAPPC6B	0.53	0.37	0.69	1.86	0.89	3.2E-08	9.8E-07	6h	21
RBM26	0.70	0.49	0.91	1.85	0.89	1.6E-10	6.5E-09	6h	21
MELK	1.31	0.93	1.69	1.83	0.87	9.6E-18	8.0E-16	6h	19
METTL8	0.66	0.47	0.86	1.82	0.87	1.7E-09	6.1E-08	6h	21
RNF213	0.56	0.40	0.72	1.82	0.86	2.4E-08	7.5E-07	6h	21
TNFRSF10D	1.07	0.76	1.38	1.81	0.86	3.3E-14	2.1E-12	6h	18
USP33	1.02	0.72	1.31	1.81	0.86	9.1E-14	5.8E-12	6h	18
KDM5B	0.60	0.43	0.78	1.80	0.85	9.5E-09	3.1E-07	6h	21

Table S4 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
PTP4A1	3.78	2.70	4.87	1.80	0.85	3.8E-46	1.0E-43	6h	19
WAPAL	0.78	0.56	1.01	1.79	0.84	2.0E-10	8.3E-09	6h	21
KIAA0020	0.74	0.53	0.95	1.79	0.84	5.2E-10	2.0E-08	6h	18
SACM1L	0.68	0.49	0.87	1.78	0.83	5.0E-09	1.7E-07	6h	19
PSME4	0.73	0.53	0.93	1.77	0.82	1.8E-09	6.4E-08	6h	18
STX16	0.92	0.67	1.17	1.76	0.82	1.2E-11	6.0E-10	6h	21
ZNF609	1.02	0.74	1.30	1.76	0.82	1.5E-12	8.3E-11	6h	18
ZHX1	0.68	0.49	0.86	1.76	0.81	6.4E-09	2.2E-07	6h	19
TBPL1	0.74	0.54	0.95	1.76	0.81	1.8E-09	6.7E-08	6h	19
MT1X	0.65	0.47	0.83	1.76	0.81	2.0E-08	6.3E-07	6h	20
ALDH9A1	0.94	0.68	1.19	1.74	0.80	3.1E-11	1.5E-09	6h	19
RSRC2	1.83	1.34	2.33	1.74	0.80	1.7E-20	1.7E-18	6h	19
DEPDC1	0.71	0.52	0.90	1.72	0.78	2.0E-08	6.1E-07	6h	21
DIS3	0.79	0.59	1.00	1.71	0.78	2.2E-09	7.8E-08	6h	19
KDM5A	0.72	0.54	0.91	1.69	0.76	2.6E-08	8.0E-07	6h	21
HMGXB4	0.91	0.68	1.15	1.69	0.76	5.4E-10	2.1E-08	6h	21
BRD8	0.82	0.61	1.03	1.69	0.76	5.0E-09	1.7E-07	6h	21
ZNF292	0.82	0.61	1.02	1.67	0.74	8.5E-09	2.8E-07	6h	14
SH3KBP1	0.75	0.56	0.94	1.67	0.74	2.9E-08	8.7E-07	6h	21
MTF2	1.21	0.91	1.50	1.65	0.73	5.3E-12	2.7E-10	6h	18
CTR9	0.87	0.66	1.08	1.63	0.71	1.5E-08	4.8E-07	6h	19
AAR2	0.91	0.69	1.13	1.63	0.71	4.4E-09	1.5E-07	6h	21
SMARCAD1	1.13	0.86	1.39	1.63	0.70	1.1E-10	4.6E-09	6h	19
LOC152217	1.25	0.96	1.53	1.60	0.68	5.3E-11	2.4E-09	6h	18
ENAH	4.17	3.21	5.13	1.60	0.68	3.0E-33	5.1E-31	6h	18
TERF2IP	1.34	1.03	1.65	1.60	0.68	1.3E-11	6.5E-10	6h	19
TFE3	0.99	0.77	1.22	1.60	0.67	6.1E-09	2.1E-07	6h	21
MCC	1.19	0.92	1.46	1.58	0.66	3.5E-10	1.4E-08	6h	17
RBM5	1.23	0.95	1.50	1.57	0.65	4.2E-10	1.7E-08	6h	21
ARL6IP1	3.46	2.70	4.21	1.56	0.65	2.2E-25	2.8E-23	6h	10
TUG1	1.34	1.06	1.63	1.54	0.62	5.5E-10	2.1E-08	6h	18
KIF11	1.43	1.13	1.74	1.54	0.62	1.2E-10	5.0E-09	6h	21
SOX2	1.31	1.03	1.58	1.53	0.61	9.6E-10	3.7E-08	6h	19
ANXA5	8.57	6.88	10.26	1.49	0.58	5.1E-46	1.4E-43	6h	18
MARCH6	1.54	1.24	1.84	1.49	0.58	6.5E-10	2.5E-08	6h	19

Table S4 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
SEC61A1	1.50	1.21	1.79	1.48	0.57	2.0E-09	7.1E-08	6h	19
TOMM70A	1.99	1.63	2.35	1.45	0.53	6.7E-11	3.0E-09	6h	21
OGT	2.55	2.10	3.00	1.43	0.52	7.5E-13	4.3E-11	6h	16
FBX1	8.48	9.95	7.02	0.71	-0.50	6.1E-35	1.1E-32	6h	9
CCNB1	1.89	2.22	1.56	0.70	-0.51	1.6E-09	6.0E-08	6h	3
TKT	2.10	2.47	1.73	0.70	-0.51	1.8E-10	7.5E-09	6h	1
NACA	2.13	2.50	1.76	0.70	-0.51	1.0E-10	4.4E-09	6h	9
NQO1	2.18	2.58	1.77	0.69	-0.54	5.2E-12	2.7E-10	6h	1
MRPS34	1.63	1.95	1.32	0.67	-0.57	3.5E-10	1.4E-08	6h	3
HSPH1	1.93	2.30	1.55	0.67	-0.57	7.2E-12	3.6E-10	6h	3
CHCHD10	1.41	1.69	1.13	0.67	-0.58	2.6E-09	9.2E-08	6h	3
DNAJA1	1.81	2.20	1.43	0.65	-0.63	3.6E-13	2.2E-11	6h	3
SLBP	1.50	1.83	1.18	0.65	-0.63	3.2E-11	1.5E-09	6h	2
SNRPE	2.65	3.22	2.08	0.65	-0.63	9.6E-19	8.7E-17	6h	9
PGAM1	1.12	1.37	0.88	0.64	-0.64	5.8E-09	2.0E-07	6h	2
TRIM24	1.04	1.26	0.81	0.64	-0.64	1.7E-08	5.4E-07	6h	3
MRPL34	1.17	1.45	0.90	0.62	-0.68	1.7E-10	7.0E-09	6h	1
SOCS2	1.17	1.46	0.87	0.60	-0.75	3.6E-12	1.9E-10	6h	1
TRIP10	1.30	1.63	0.97	0.60	-0.75	2.6E-13	1.6E-11	6h	1
SAPCD2	0.81	1.02	0.60	0.59	-0.77	3.5E-09	1.2E-07	6h	3
OTX2	1.44	1.84	1.04	0.57	-0.82	2.9E-17	2.4E-15	6h	3
CBX4	0.61	0.78	0.43	0.55	-0.86	1.1E-08	3.7E-07	6h	3
MEX3A	1.13	1.46	0.80	0.54	-0.88	1.8E-15	1.3E-13	6h	3
EFNB1	0.85	1.11	0.59	0.53	-0.92	8.1E-13	4.6E-11	6h	3
SH3BP4	0.83	1.09	0.57	0.52	-0.94	5.6E-13	3.2E-11	6h	3
PPP1R18	0.58	0.76	0.40	0.52	-0.94	1.9E-09	6.9E-08	6h	1
MSMO1	0.71	0.94	0.48	0.52	-0.95	9.4E-12	4.7E-10	6h	1
CDC42EP1	0.46	0.61	0.31	0.51	-0.96	2.7E-08	8.4E-07	6h	3
BCOR	0.55	0.73	0.37	0.50	-0.99	7.5E-10	2.9E-08	6h	3
CBX2	0.57	0.78	0.37	0.47	-1.10	1.6E-12	8.5E-11	6h	1
DUSP14	0.89	1.22	0.57	0.46	-1.11	1.6E-18	1.4E-16	6h	1
IRF2BP2	0.49	0.67	0.31	0.46	-1.12	8.9E-11	3.9E-09	6h	3
TRAF4	0.34	0.47	0.21	0.45	-1.14	3.0E-08	9.0E-07	6h	1
SOX21	0.46	0.63	0.28	0.45	-1.15	1.5E-10	6.1E-09	6h	3
AJUBA	0.45	0.63	0.28	0.45	-1.16	5.5E-11	2.5E-09	6h	3



Table S4 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
XBP1	0.44	0.61	0.27	0.44	-1.18	8.8E-11	3.9E-09	6h	1
GDF15	0.45	0.63	0.27	0.43	-1.21	2.0E-11	9.8E-10	6h	3
BTG2	0.42	0.59	0.25	0.42	-1.24	2.6E-11	1.3E-09	6h	3
PHLDA1	0.32	0.46	0.18	0.40	-1.33	6.6E-10	2.5E-08	6h	3
NR2F2	1.04	1.52	0.56	0.37	-1.45	7.7E-34	1.3E-31	6h	3
ARL4C	0.70	1.03	0.37	0.36	-1.48	7.2E-24	8.3E-22	6h	3
NAB2	0.26	0.39	0.13	0.32	-1.64	1.3E-11	6.6E-10	6h	3
IRF2BPL	0.30	0.46	0.15	0.32	-1.66	4.2E-13	2.5E-11	6h	3
ZSCAN4	5.37	0.26	10.48	40.09	5.33	0	0	4h 6h	19
LEUTX	2.31	0.14	4.48	33.08	5.05	0	0	4h 6h	20
RFPL4A	1.97	0.15	3.79	26.02	4.70	4.2E-304	5.0E-301	4h 6h	20
PRAMEF1	1.47	0.11	2.83	24.70	4.63	2.6E-224	2.4E-221	4h 6h	19
PRAMEF12	1.01	0.13	1.89	14.44	3.85	2.1E-133	1.4E-130	4h 6h	19
SLC34A2	0.89	0.12	1.65	13.83	3.79	2.9E-114	1.5E-111	4h 6h	19
TRIM51	1.68	0.25	3.11	12.51	3.65	6.7E-208	5.7E-205	4h 6h	19
TFIP11	0.72	0.18	1.26	7.11	2.83	4.3E-67	1.8E-64	4h 6h	19
KDM4E	0.41	0.10	0.72	6.92	2.79	2.1E-38	4.1E-36	4h 6h	19
GTF2F1	4.95	1.37	8.54	6.25	2.64	0	0	4h 6h	19
PRRG4	0.55	0.17	0.93	5.43	2.44	5.8E-43	1.4E-40	4h 6h	19
PNP	4.95	1.57	8.32	5.29	2.40	0	0	4h 6h	19
SPTY2D1	0.83	0.31	1.34	4.28	2.10	1.2E-50	3.7E-48	4h 6h	19
ESRG	0.69	0.27	1.11	4.14	2.05	3.7E-41	8.1E-39	4h 6h	19
ARID5B	0.76	0.31	1.21	3.94	1.98	2.4E-42	5.5E-40	4h 6h	19
ZNF622	1.09	0.44	1.74	3.92	1.97	1.3E-59	4.6E-57	4h 6h	19
DBR1	0.96	0.40	1.52	3.82	1.93	4.2E-51	1.3E-48	4h 6h	19
NXF1	2.35	1.04	3.66	3.53	1.82	8.2E-112	3.9E-109	4h 6h	19
HOXB2	1.71	0.79	2.63	3.35	1.74	2.8E-76	1.2E-73	4h 6h	19
EXOSC10	3.03	1.47	4.59	3.11	1.64	2.4E-120	1.5E-117	4h 6h	19
ITGB8	0.40	0.20	0.60	3.06	1.61	1.6E-16	1.2E-14	4h 6h	21
RICTOR	0.78	0.39	1.17	2.99	1.58	1.3E-30	1.9E-28	4h 6h	21
PDGFRA	0.58	0.30	0.86	2.90	1.54	9.9E-22	1.1E-19	4h 6h	19
CDH10	0.96	0.51	1.41	2.80	1.48	2.2E-33	3.8E-31	4h 6h	19
ZNF827	0.55	0.31	0.79	2.54	1.34	6.4E-17	5.1E-15	4h 6h	21
MRPL49	2.41	1.42	3.40	2.40	1.27	2.7E-61	9.8E-59	4h 6h	19
ZSWIM6	0.56	0.34	0.79	2.34	1.23	1.3E-14	8.8E-13	4h 6h	21

Table S4 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
ODF2L	0.62	0.38	0.87	2.29	1.20	1.0E-15	7.4E-14	4h 6h	18
SHC1	1.70	1.04	2.36	2.28	1.19	5.3E-39	1.1E-36	4h 6h	19
KIAA1551	0.91	0.56	1.27	2.28	1.19	1.0E-21	1.1E-19	4h 6h	21
YTHDC1	1.58	0.98	2.18	2.23	1.15	1.3E-34	2.2E-32	4h 6h	21
GABPB1-AS1	0.57	0.36	0.79	2.22	1.15	1.4E-13	8.6E-12	4h 6h	18
SYNE2	0.81	0.50	1.11	2.22	1.15	2.8E-18	2.4E-16	4h 6h	17
LRRC8B	1.08	0.68	1.49	2.20	1.14	1.3E-23	1.5E-21	4h 6h	21
BTA1F1	0.77	0.49	1.06	2.18	1.12	5.2E-17	4.2E-15	4h 6h	21
ARHGGEF26	1.17	0.74	1.61	2.17	1.12	1.1E-24	1.4E-22	4h 6h	19
NFAT5	0.85	0.54	1.16	2.14	1.10	6.6E-18	5.6E-16	4h 6h	19
ZNF644	0.86	0.57	1.16	2.05	1.03	3.6E-16	2.7E-14	4h 6h	19
GUSBP3	0.53	0.37	0.70	1.89	0.92	1.3E-08	4.1E-07	4h 6h	14
PUM1	2.53	1.76	3.29	1.87	0.90	8.2E-35	1.5E-32	4h 6h	21
PHACTR2	0.54	0.38	0.70	1.87	0.90	1.7E-08	5.4E-07	4h 6h	18
PLK4	1.03	0.72	1.33	1.85	0.89	1.3E-14	8.6E-13	4h 6h	19
ZMYND8	1.05	0.76	1.35	1.77	0.82	3.4E-13	2.0E-11	4h 6h	18
CCNL2	1.45	1.06	1.84	1.74	0.80	9.4E-17	7.4E-15	4h 6h	19
NET1	1.10	0.83	1.38	1.67	0.74	1.5E-11	7.4E-10	4h 6h	18
CIRBP	4.24	3.19	5.30	1.66	0.73	5.7E-39	1.2E-36	4h 6h	19
POLQ	0.85	0.64	1.06	1.65	0.73	1.2E-08	3.8E-07	4h 6h	17
ZRANB2	2.71	2.05	3.38	1.65	0.72	1.0E-24	1.3E-22	4h 6h	18
SRSF11	4.64	3.54	5.74	1.62	0.70	1.4E-38	2.8E-36	4h 6h	18
PPIG	1.76	1.36	2.16	1.59	0.67	1.2E-14	7.8E-13	4h 6h	14
CUX1	1.00	0.77	1.22	1.59	0.67	7.9E-09	2.6E-07	4h 6h	14
SLC25A36	1.47	1.16	1.79	1.54	0.62	5.8E-11	2.7E-09	4h 6h	18
BBX	1.51	1.19	1.82	1.53	0.61	7.3E-11	3.3E-09	4h 6h	14
RIF1	1.53	1.22	1.85	1.52	0.61	8.9E-11	3.9E-09	4h 6h	13
SMC3	1.58	1.27	1.90	1.49	0.58	2.8E-10	1.1E-08	4h 6h	17
DPPA4	4.17	3.40	4.95	1.46	0.54	2.9E-22	3.2E-20	4h 6h	18
TNRC6A	1.62	1.32	1.92	1.46	0.54	1.8E-09	6.4E-08	4h 6h	17
FTL	9.08	10.71	7.44	0.70	-0.53	1.3E-39	2.8E-37	4h 6h	2
RPL23A	10.78	12.92	8.63	0.67	-0.58	4.6E-53	1.5E-50	4h 6h	9
UBE2S	2.92	3.50	2.33	0.67	-0.59	4.1E-18	3.5E-16	4h 6h	3
HNRNPA0	1.59	1.95	1.23	0.63	-0.67	4.4E-13	2.6E-11	4h 6h	8
DYNLL1	6.59	8.58	4.60	0.54	-0.90	1.5E-86	7.0E-84	4h 6h	3

Table S4 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
SLC7A5	0.83	1.08	0.57	0.53	-0.92	1.6E-12	8.9E-11	4h 6h	1
PCDH18	0.57	0.76	0.37	0.48	-1.05	4.2E-11	2.0E-09	4h 6h	3
ISOC2	0.40	0.56	0.25	0.44	-1.18	2.5E-10	1.0E-08	4h 6h	9
SHISA3	1.37	1.92	0.83	0.43	-1.22	7.6E-33	1.2E-30	4h 6h	3
TGIF1	0.64	0.92	0.35	0.38	-1.39	1.3E-19	1.3E-17	4h 6h	3
PIM1	0.34	0.50	0.19	0.37	-1.43	6.1E-12	3.1E-10	4h 6h	3
NEDD9	0.75	1.09	0.40	0.37	-1.44	2.2E-24	2.7E-22	4h 6h	3
MIDN	0.83	1.23	0.43	0.35	-1.51	2.4E-29	3.4E-27	4h 6h	3
CYR61	1.02	1.51	0.53	0.35	-1.52	8.4E-36	1.6E-33	4h 6h	3
NUAK2	0.33	0.50	0.16	0.32	-1.63	8.6E-14	5.5E-12	4h 6h	3
NOG	0.27	0.43	0.12	0.27	-1.87	2.6E-14	1.7E-12	4h 6h	3
POLR2L	2.14	2.53	1.76	0.70	-0.52	4.7E-11	2.2E-09	3h 6h	8
RFPL4B	6.91	0.26	13.56	51.62	5.69	0	0	3h 4h 6h	19
ZNF217	3.16	0.64	5.68	8.89	3.15	0	0	3h 4h 6h	19
RBBP6	12.63	2.58	22.68	8.78	3.13	0	0	3h 4h 6h	19
SRSF8	6.06	1.77	10.35	5.86	2.55	0	0	3h 4h 6h	19
ZNF296	2.03	0.60	3.46	5.79	2.53	4.3E-161	3.2E-158	3h 4h 6h	19
F5	0.56	0.28	0.83	2.94	1.56	6.1E-22	6.6E-20	3h 4h 6h	18
PNN	5.76	2.94	8.57	2.92	1.54	2.5E-205	2.0E-202	3h 4h 6h	21
ZMAT3	1.30	0.69	1.91	2.78	1.47	4.2E-44	1.1E-41	3h 4h 6h	19
CCDC144B	0.60	0.33	0.87	2.67	1.42	1.3E-19	1.2E-17	3h 4h 6h	18
TSIX	0.58	0.32	0.84	2.64	1.40	6.6E-19	6.1E-17	3h 4h 6h	18
ZNF471	0.69	0.38	0.99	2.60	1.38	1.6E-21	1.7E-19	3h 4h 6h	18
GOLGB1	1.24	0.71	1.77	2.51	1.33	7.3E-35	1.3E-32	3h 4h 6h	18
TMEM212	0.54	0.32	0.76	2.37	1.24	9.6E-15	6.6E-13	3h 4h 6h	18
MLL5	0.85	0.52	1.19	2.31	1.21	3.3E-21	3.4E-19	3h 4h 6h	17
MPHOSPH8	1.03	0.65	1.41	2.15	1.11	2.8E-21	2.9E-19	3h 4h 6h	18
BOD1L1	0.90	0.57	1.22	2.13	1.09	2.6E-18	2.2E-16	3h 4h 6h	18
GADD45A	2.12	1.37	2.88	2.11	1.08	9.9E-41	2.2E-38	3h 4h 6h	18
RBM25	3.77	2.48	5.06	2.04	1.03	3.4E-65	1.3E-62	3h 4h 6h	18
SLC4A7	0.68	0.45	0.91	2.01	1.01	1.6E-12	8.6E-11	3h 4h 6h	17
PNISR	2.24	1.57	2.90	1.85	0.89	4.9E-30	7.2E-28	3h 4h 6h	18
LUC7L3	2.31	1.65	2.97	1.81	0.85	8.3E-29	1.2E-26	3h 4h 6h	18
SLTM	1.35	1.03	1.68	1.64	0.71	8.0E-13	4.6E-11	3h 4h 6h	18
ATRX	2.10	1.65	2.55	1.54	0.63	3.3E-15	2.3E-13	3h 4h 6h	14

Table S4 continued

Gene	base Mean	base MeanA	base MeanB	fold Change	log2 FC	pval	padj	shared in states	Node
DMWD	1.26	1.01	1.52	1.51	0.59	8.7E-09	2.9E-07	3h 4h 6h	14
CCAR1	3.42	2.74	4.09	1.50	0.58	1.0E-20	1.1E-18	3h 4h 6h	14
COX7C	4.83	5.68	3.99	0.70	-0.51	2.2E-22	2.4E-20	3h 4h 6h	8
PRDX4	1.43	1.70	1.17	0.69	-0.54	2.8E-08	8.5E-07	3h 4h 6h	8
APP	2.16	2.85	1.47	0.52	-0.96	6.7E-33	1.1E-30	3h 4h 6h	8
MAGOH	0.65	0.87	0.43	0.50	-1.01	5.6E-12	2.9E-10	3h 4h 6h	9
FKBP10	0.37	0.52	0.21	0.41	-1.30	7.5E-11	3.3E-09	3h 4h 6h	7
ID1	2.32	4.18	0.45	0.11	-3.21	2.2E-237	2.2E-234	3h 4h 6h	3
ID3	1.16	2.09	0.23	0.11	-3.21	8.9E-119	4.9E-116	3h 4h 6h	3
TOP2A	1.67	1.21	2.12	1.76	0.81	1.3E-19	1.3E-17	2h 6h	10
MKI67	2.31	1.89	2.74	1.45	0.54	1.4E-12	7.9E-11	2h 6h	10
KIF14	0.89	0.67	1.10	1.64	0.71	7.2E-09	2.4E-07	2h 3h 6h	10
HIST1H2BK	1.18	0.91	1.44	1.57	0.65	9.1E-10	3.5E-08	2h 3h 6h	19

Table S5. Differentially expressed transcription factors, cofactors and kinases between uninduced and induced DIE cells.

Gene	DE	Induction states	Factor
CDK6	UP	2h	Kinase
TAF3	UP	2h 3h	Cofactor
NIPBL	UP	2h 3h	Cofactor
CHD7	UP	2h 3h	Cofactor
ROCK1	UP	2h 3h 4h	Kinase
CDC42BPA	UP	2h 3h 4h	Kinase
CCDC88A	UP	2h 3h 4h	Transcription factor
DNAJC2	UP	2h 3h 4h	Transcription factor
ZFHX3	UP	2h 3h 4h	Transcription factor
CENPF	UP	All 4	Cofactor
BDP1	UP	All 4	Cofactor
BRD4	UP	All 4	Cofactor
TOP1	UP	All 4	Cofactor
BPTF	UP	All 4	Cofactor
FOXN3	UP	2h 4h	Transcription factor
ASH1L	UP	3h	Cofactor
XIAP	UP	3h	Cofactor
TPR	UP	3h	Cofactor

Table S5 continued

Gene	DE	Induction states	Factor
SETD2	UP	3h	Cofactor
ESF1	UP	3h	Cofactor
WNK1	UP	3h	Kinase
REST	UP	3h	Transcription factor
CHD9	UP	3h 4h	Cofactor
ATRX	UP	3h 4h 6h	Cofactor
MPHOSPH8	UP	3h 4h 6h	Cofactor
CCAR1	UP	3h 4h 6h	Cofactor
SLTM	UP	3h 4h 6h	Cofactor
ZNF471	UP	3h 4h 6h	Transcription factor
ZNF296	UP	3h 4h 6h	Transcription factor
ZNF217	UP	3h 4h 6h	Transcription factor
ASCC3	UP	4h	Cofactor
HELLS	UP	4h	Cofactor
ZFHX4	UP	4h	Transcription factor
BAZ2B	UP	4h	Transcription factor
PAXBP1	UP	4h	Transcription factor
LRRFIP1	UP	4h	Transcription factor
GLI3	UP	4h	Transcription factor
KDM4E	UP	4h 6h	Cofactor
GTF2F1	UP	4h 6h	Cofactor
BTAF1	UP	4h 6h	Cofactor
ZMYND8	UP	4h 6h	Cofactor
DPPA4	UP	4h 6h	Cofactor
PDGFRA	UP	4h 6h	Kinase
PLK4	UP	4h 6h	Kinase
ZSCAN4	UP	4h 6h	Transcription factor
LEUTX	UP	4h 6h	Transcription factor
ARID5B	UP	4h 6h	Transcription factor
HOXB2	UP	4h 6h	Transcription factor
ZNF622	UP	4h 6h	Transcription factor
ZNF827	UP	4h 6h	Transcription factor
NFAT5	UP	4h 6h	Transcription factor
ZNF644	UP	4h 6h	Transcription factor
CUX1	UP	4h 6h	Transcription factor

Table S5 continued

Gene	DE	Induction states	Factor
BBX	UP	4h 6h	Transcription factor
CCNA1	UP	6h	Cofactor
C1D	UP	6h	Cofactor
TAF4B	UP	6h	Cofactor
HEXIM1	UP	6h	Cofactor
NCOA7	UP	6h	Cofactor
SUPT6H	UP	6h	Cofactor
BRCA2	UP	6h	Cofactor
BIRC2	UP	6h	Cofactor
UFL1	UP	6h	Cofactor
TOPORS	UP	6h	Cofactor
NOTCH2	UP	6h	Cofactor
ELOF1	UP	6h	Cofactor
KDM5B	UP	6h	Cofactor
TBPL1	UP	6h	Cofactor
DEPDC1	UP	6h	Cofactor
KDM5A	UP	6h	Cofactor
BRD8	UP	6h	Cofactor
MTF2	UP	6h	Cofactor
CTR9	UP	6h	Cofactor
SMARCAD1	UP	6h	Cofactor
TERF2IP	UP	6h	Cofactor
OGT	UP	6h	Cofactor
LRRK1	UP	6h	Kinase
CLK1	UP	6h	Kinase
MAST1	UP	6h	Kinase
STK17B	UP	6h	Kinase
EPHA4	UP	6h	Kinase
MELK	UP	6h	Kinase
SNAI1	UP	6h	Transcription factor
EOMES	UP	6h	Transcription factor
PHOX2B	UP	6h	Transcription factor
BHLHE22	UP	6h	Transcription factor
OSR2	UP	6h	Transcription factor
ZNF574	UP	6h	Transcription factor

Table S5 continued

Gene	DE	Induction states	Factor
ZNF480	UP	6h	Transcription factor
TFAP2C	UP	6h	Transcription factor
ZNF91	UP	6h	Transcription factor
ZNF281	UP	6h	Transcription factor
ZNF609	UP	6h	Transcription factor
ZHX1	UP	6h	Transcription factor
HMGXB4	UP	6h	Transcription factor
ZNF292	UP	6h	Transcription factor
TFE3	UP	6h	Transcription factor
SOX2	UP	6h	Transcription factor
ID1	DOWN	3h 4h 6h	Transcription factor
ID3	DOWN	3h 4h 6h	Transcription factor
ID2	DOWN	4h	Transcription factor
PIM1	DOWN	4h 6h	Cofactor/Kinase
NUAK2	DOWN	4h 6h	Kinase
TGIF1	DOWN	4h 6h	Transcription factor
NACA	DOWN	6h	Cofactor
TRIM24	DOWN	6h	Cofactor
CBX4	DOWN	6h	Cofactor
BCOR	DOWN	6h	Cofactor
CBX2	DOWN	6h	Cofactor
AJUBA	DOWN	6h	Cofactor
NAB2	DOWN	6h	Cofactor
IRF2BPL	DOWN	6h	Cofactor
OTX2	DOWN	6h	Transcription factor
SOX21	DOWN	6h	Transcription factor
XBP1	DOWN	6h	Transcription factor
NR2F2	DOWN	6h	Transcription factor

Table S6. Enrichr detected expression profiles of transcription factor in induced DIE cells (Adjusted p value < 0.001)

- Category A: The expression/activation of the transcription factor can cause the upregulation of a set of genes found to be differentially expressed induced DIE cells.
- Category B: The inhibition/deactivation of the transcription factor can cause the upregulation of a set of genes found to be differentially expressed induced DIE cells.
- Category C: The expression/activation of the transcription factor can cause the downregulation of a set of genes found to be differentially expressed induced DIE cells.
- Category D: The inhibition/deactivation of the transcription factor can cause the downregulation of a set of genes found to be differentially expressed induced DIE cells.
- Other: No clear annotation of the expression/activation status of the transcription factors.

TF	Exp. Profile found at	Category
ADAR	3h	A
DUX4	4h 6h	A
HIF1A	2h 6h	A
LIN28	All 4	A
MYC	2h 3h 6h	A
PAX7	2h	A
RBM10	2h	A
SOX5	All 4	A
ZIC3	All 4	A
AFF4	2h 3h 4h	B
ASCL1	2h	B
ATF4	2h	B
BNC2	2h 3h	B
ELF3	2h	B
EZH2	2h	B
FOXP1	6h	B
HOXA7	2h	B
HSF1	3h 4h 6h	B
JUNB	All 4	B
KLF10	All 4	B
MEIS2	All 4	B
MYCN	All 4	B
NFKB1	6h	B
PITX2	All 4	B
PPARD	2h 3h	B

TF	Exp. Profile found at	Category
ATF3	4h	D
BNC2	3h 4h	D
CREB1	3h	D
DOT1L	3h 4h 6h	D
E2F1	3h	D
EHF	6h	D
ELF3	3h	D
ELK1	3h 4h 6h	D
EPAS1	3h 4h	D
ERG	6h	D
ESR1	3h 4h 6h	D
EZH2	3h 4h 6h	D
FOXA2	6h	D
FOXM1	3h 4h 6h	D
FOXP1	3h 4h 6h	D
HNF4A	3h 4h 6h	D
IRF4	3h 4h 6h	D
JUN	6h	D
JUNB	6h	D
JUND	6h	D
KLF10	3h 6h	D
KLF2	3h 4h 6h	D
MBD2	4h	D
MBNL1	6h	D
MECOM	6h	D

Table S6 continued

TF	Exp. Profile found at	Category
SETDB1	All 4	B
SOX11	2h	B
STAT3	2h	B
TP53	4h 6h	B
ZEB2	All 4	B
ZMAT4	2h 3h 4h	B
ZNF253	2h 3h 4h	B
ZNF503	All 4	B
ZNF750	2h 6h	B
ATF6	3h 6h	C
DLX4	3h 4h 6h	C
DUX4	4h 6h	C
E2F1	4h 6h	C
EHF	3h	C
FOXP1	6h	C
FOXP2	3h	C
FOXP3	3h 4h	C
GATA4	3h 4h 6h	C
GATA6	3h	C
HIF1A	4h 6h	C
HNF1A	3h 4h 6h	C
HNF1B	3h	C
HNF4G	3h	C
KLF4	3h 4h 6h	C
MYB	All 4	C
MYC	3h 4h	C
NANOG	4h	C
NME2	6h	C
NR4A2	3h 4h 6h	C
OVOL1	3h 4h 6h	C
OVOL2	3h 4h 6h	C
POU1F1	3h 4h 6h	C
RARA	6h	C
RBM10	3h 4h 6h	C
SOX17	3h 4h	C

TF	Exp. Profile found at	Category
MEIS2	3h	D
MITF	4h 6h	D
MYB	6h	D
MYC	3h 4h 6h	D
MYCN	4h 6h	D
NANOG	3h 4h 6h	D
NFKB1	3h 4h 6h	D
NFXL1	3h 4h 6h	D
NR2F2	All 4	D
OTX2	4h	D
PCGF2	3h 4h 6h	D
POU5F1	4h 6h	D
PPARD	3h	D
RARA	4h 6h	D
RELA	3h 4h	D
SALL4	3h	D
SETDB1	6h	D
SON	3h 4h 6h	D
SOX11	3h 4h 6h	D
SOX4	3h 4h 6h	D
SP1	6h	D
SP3	4h	D
STAT3	6h	D
SUZ12	3h 4h 6h	D
TBX3	3h 4h 6h	D
TCF21	6h	D
TCF4	3h 4h	D
TCF7L2	4h 6h	D
TP53	3h 4h	D
TP63	4h 6h	D
TSHZ3	4h 6h	D
YY1	3h 4h 6h	D
ZBTB48	4h 6h	D
ZNF395	4h	D
ZNF658	3h	D

Table S6 continued

TF	Exp. Profile found at	Category
SOX2	3h 4h	C
SOX7	3h 4h 6h	C
WT1	All 4	C
ZIC3	3h 4h 6h	C
ZNF217	4h 6h	C
AFF4	4h 6h	D
AR	3h 4h 6h	D
ARID2	4h 6h	D
ARX	3h 6h	D

TF	Exp. Profile found at	Category
ZXDC	6h	D
ESR1	3h	Other
FLI1	2h 3h 4h	Other
IKZF1	3h 4h	Other
MECP2	3h 4h	Other
THRA	3h	Other
THRB	3h	Other
TP63	6h	Other
TWIST2	3h 4h 6h	Other

Tables S7. Gene ontology results of differentially expressed genes in DIE cells

- Analysis Type: PANTHER Overrepresentation Test (Released 20200728)
- Annotation Version and Release Date: GO Ontology database DOI: 10.5281/zenodo.4033054 Released 2020-09-10
- Analyzed List: upload_1 (Homo sapiens)
- Reference List: Homo sapiens (all genes in database)
- Test Type: FISHER
- Correction: False Discovery Rate (FDR < 0.05)

2h UP							
General term	GO #	in Ref	in data	expected	fold Enr.	raw P-value	FDR
Embryonic development	GO:0045927	270	5	0.53	9.42	1.9E-04	4.5E-02
Embryonic development	GO:0060324	50	3	0.1	30.51	1.5E-04	3.9E-02
Embryonic development	GO:0009790	1003	9	1.97	4.56	1.2E-04	3.2E-02
Embryonic development	GO:0010720	568	7	1.12	6.27	1.2E-04	3.1E-02
Embryonic development	GO:0051130	1229	10	2.42	4.14	1.1E-04	2.9E-02
Embryonic development	GO:0040016	5	2	0.01	> 100	7.8E-05	2.4E-02
Embryonic development	GO:0048856	5489	23	10.79	2.13	5.7E-05	2.0E-02
Embryonic development	GO:0045595	1884	13	3.7	3.51	4.2E-05	1.8E-02
Embryonic development	GO:0048639	180	5	0.35	14.13	3.0E-05	1.4E-02
Embryonic development	GO:2000026	2107	14	4.14	3.38	3.0E-05	1.4E-02
Embryonic development	GO:0007275	5106	23	10.04	2.29	2.5E-05	1.2E-02
Embryonic development	GO:0051128	2436	16	4.79	3.34	7.1E-06	5.4E-03
Embryonic development	GO:0060322	820	11	1.61	6.82	3.9E-07	5.7E-04
Embryonic development	GO:0060284	978	12	1.92	6.24	2.7E-07	4.2E-04
CNS development	GO:0051960	959	12	1.89	6.36	2.2E-07	4.3E-04
CNS development	GO:0050767	847	11	1.67	6.6	5.4E-07	7.1E-04
CNS development	GO:0007420	775	10	1.52	6.56	2.1E-06	2.1E-03
CNS development	GO:0007399	2437	16	4.79	3.34	7.1E-06	5.2E-03
CNS development	GO:0048699	1599	13	3.14	4.13	7.4E-06	5.1E-03
CNS development	GO:0022008	1703	13	3.35	3.88	1.5E-05	8.6E-03
CNS development	GO:0007417	1025	10	2.02	4.96	2.4E-05	1.2E-02
CNS development	GO:0050769	490	7	0.96	7.27	4.6E-05	1.8E-02
CNS development	GO:0045664	680	8	1.34	5.98	4.9E-05	1.7E-02
CNS development	GO:0051962	558	7	1.1	6.38	1.0E-04	3.0E-02
CNS development	GO:0045773	44	3	0.09	34.67	1.1E-04	3.0E-02
CNS development	GO:0048731	4525	20	8.9	2.25	1.6E-04	4.1E-02
CNS development	GO:0021537	265	5	0.52	9.6	1.8E-04	4.4E-02

Table S7 continued

2h UP							
General term	GO #	in Ref	in data	expected	fold Enr.	raw P-value	FDR
Cell cycle and cell division	GO:0022402	1069	15	2.1	7.14	8.6E-10	1.4E-05
Cell cycle and cell division	GO:0051301	501	11	0.99	11.17	2.8E-09	2.2E-05
Cell cycle and cell division	GO:0007049	1390	16	2.73	5.85	3.4E-09	1.8E-05
Cell cycle and cell division	GO:0007346	640	11	1.26	8.74	3.3E-08	1.3E-04
Cell cycle and cell division	GO:0051726	1210	14	2.38	5.88	4.1E-08	1.3E-04
Cell cycle and cell division	GO:0051276	1062	13	2.09	6.23	7.6E-08	2.0E-04
Cell cycle and cell division	GO:1903047	695	11	1.37	8.05	7.6E-08	1.7E-04
Cell cycle and cell division	GO:0000278	772	11	1.52	7.25	2.2E-07	3.8E-04
Cell cycle and cell division	GO:0006996	3576	21	7.03	2.99	6.6E-07	8.1E-04
Cell cycle and cell division	GO:0007059	278	7	0.55	12.81	1.2E-06	1.4E-03
Cell cycle and cell division	GO:0010564	770	10	1.51	6.6	2.0E-06	2.1E-03
Cell cycle and cell division	GO:0071103	309	7	0.61	11.52	2.5E-06	2.3E-03
Cell cycle and cell division	GO:0007017	806	10	1.58	6.31	2.9E-06	2.6E-03
Cell cycle and cell division	GO:0098813	220	6	0.43	13.87	4.9E-06	4.1E-03
Cell cycle and cell division	GO:0033043	1332	12	2.62	4.58	6.6E-06	5.2E-03
Cell cycle and cell division	GO:0016043	5699	25	11.21	2.23	8.8E-06	5.8E-03
Cell cycle and cell division	GO:0000819	140	5	0.28	18.16	9.2E-06	5.9E-03
Cell cycle and cell division	GO:0051383	19	3	0.04	80.3	1.1E-05	6.4E-03
Cell cycle and cell division	GO:0071840	5919	25	11.64	2.15	1.5E-05	8.6E-03
Cell cycle and cell division	GO:1901990	420	7	0.83	8.48	1.8E-05	9.6E-03
Cell cycle and cell division	GO:0007088	168	5	0.33	15.14	2.2E-05	1.2E-02
Cell cycle and cell division	GO:0000280	302	6	0.59	10.1	2.9E-05	1.4E-02
Cell cycle and cell division	GO:1901987	458	7	0.9	7.77	3.0E-05	1.3E-02
Cell cycle and cell division	GO:0051783	195	5	0.38	13.04	4.3E-05	1.8E-02
Cell cycle and cell division	GO:0051651	96	4	0.19	21.19	4.4E-05	1.8E-02
Cell cycle and cell division	GO:0051642	32	3	0.06	47.68	4.4E-05	1.8E-02
Cell cycle and cell division	GO:0010389	198	5	0.39	12.84	4.7E-05	1.8E-02
Cell cycle and cell division	GO:0048285	332	6	0.65	9.19	4.8E-05	1.8E-02
Cell cycle and cell division	GO:0061842	33	3	0.06	46.23	4.8E-05	1.7E-02
Cell cycle and cell division	GO:0006325	701	8	1.38	5.8	6.0E-05	2.0E-02
Cell cycle and cell division	GO:0007018	346	6	0.68	8.82	6.0E-05	2.0E-02
Cell cycle and cell division	GO:0051983	108	4	0.21	18.84	6.8E-05	2.2E-02
Cell cycle and cell division	GO:1902749	215	5	0.42	11.83	6.8E-05	2.2E-02
Cell cycle and cell division	GO:0000070	111	4	0.22	18.33	7.5E-05	2.4E-02



Table S7 continued

2h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
Cell cycle and cell division	GO:0033044	367	6	0.72	8.31	8.3E-05	2.5E-02
Cell cycle and cell division	GO:0051661	6	2	0.01	> 100	1.0E-04	3.0E-02
Cell cycle and cell division	GO:0030261	44	3	0.09	34.67	1.1E-04	3.0E-02
Cell cycle and cell division	GO:0045786	579	7	1.14	6.15	1.3E-04	3.4E-02
Cell cycle and cell division	GO:0051640	605	7	1.19	5.88	1.7E-04	4.2E-02
Cell cycle and cell division	GO:0044848	270	5	0.53	9.42	1.9E-04	4.7E-02
Cell cycle and cell division	GO:0022403	270	5	0.53	9.42	1.9E-04	4.5E-02
Cell cycle and cell division	GO:0098763	270	5	0.53	9.42	1.9E-04	4.6E-02
Cell cycle and cell division	GO:0006928	1599	11	3.14	3.5	2.0E-04	4.5E-02
Cell cycle and cell division	GO:0051310	55	3	0.11	27.74	2.0E-04	4.5E-02
Cell cycle and cell division	GO:0006265	9	2	0.02	> 100	2.0E-04	4.5E-02
Cell cycle and cell division	GO:0007050	145	4	0.29	14.03	2.1E-04	4.5E-02

2h DOWN							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
Protein production and regulation	GO:0072657	528	4	0.2	19.75	2.7E-05	4.3E-02
Protein production and regulation	GO:0006413	145	3	0.06	53.92	1.9E-05	3.8E-02
Protein production and regulation	GO:0070972	142	3	0.05	55.06	1.8E-05	4.1E-02
Protein production and regulation	GO:0072599	115	3	0.04	67.99	9.7E-06	3.1E-02
Protein production and regulation	GO:0045047	111	3	0.04	70.44	8.7E-06	4.6E-02
RNA productio and processing	GO:0000184	121	3	0.05	64.62	1.1E-05	3.0E-02
Viral processes	GO:0019080	154	3	0.06	50.77	2.3E-05	4.0E-02
Viral processes	GO:0019083	115	3	0.04	67.99	9.7E-06	3.9E-02

3h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
Embryonic development	GO:0060322	820	11	2.87	3.83	1.4E-04	4.8E-02
Embryonic development	GO:0060284	978	13	3.42	3.8	3.4E-05	2.2E-02
CNS development	GO:0051960	959	13	3.36	3.87	2.8E-05	2.0E-02
CNS development	GO:0050767	847	12	2.97	4.05	3.9E-05	2.2E-02
CNS development	GO:0045664	680	10	2.38	4.2	1.4E-04	4.9E-02
Cell cycle and cell division	GO:0051276	1062	17	3.72	4.57	1.3E-07	1.0E-03
Cell cycle and cell division	GO:0022402	1069	17	3.74	4.54	1.4E-07	7.6E-04

Table S7 continued

3h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
Cell cycle and cell division	GO:0007049	1390	19	4.87	3.9	2.3E-07	9.2E-04
Cell cycle and cell division	GO:1903047	695	13	2.43	5.34	9.1E-07	2.9E-03
Cell cycle and cell division	GO:0006325	701	13	2.45	5.3	1.0E-06	2.6E-03
Cell cycle and cell division	GO:0006260	223	8	0.78	10.25	1.4E-06	2.8E-03
Cell cycle and cell division	GO:0051301	501	11	1.75	6.27	1.5E-06	2.7E-03
Cell cycle and cell division	GO:0000278	772	13	2.7	4.81	2.9E-06	4.1E-03
Cell cycle and cell division	GO:0007017	806	13	2.82	4.61	4.5E-06	5.1E-03
Cell cycle and cell division	GO:0006996	3576	29	12.52	2.32	5.5E-06	5.8E-03
Cell cycle and cell division	GO:0033047	73	5	0.26	19.56	7.8E-06	7.7E-03
Cell cycle and cell division	GO:0071103	309	8	1.08	7.39	1.5E-05	1.3E-02
Cell cycle and cell division	GO:0033045	85	5	0.3	16.8	1.6E-05	1.3E-02
Cell cycle and cell division	GO:0000226	568	10	1.99	5.03	3.1E-05	2.2E-02
Cell cycle and cell division	GO:0051983	108	5	0.38	13.22	4.7E-05	2.6E-02
Cell cycle and cell division	GO:0032508	110	5	0.39	12.98	5.1E-05	2.7E-02
Cell cycle and cell division	GO:0051383	19	3	0.07	45.1	6.0E-05	3.0E-02
Cell cycle and cell division	GO:0032392	118	5	0.41	12.1	7.1E-05	3.3E-02
Cell cycle and cell division	GO:0051726	1210	14	4.24	3.3	7.3E-05	3.3E-02
Cell cycle and cell division	GO:0010564	770	11	2.7	4.08	7.9E-05	3.5E-02
Cell cycle and cell division	GO:1905269	122	5	0.43	11.71	8.2E-05	3.5E-02
Cell cycle and cell division	GO:0007346	640	10	2.24	4.46	8.3E-05	3.5E-02
Cell cycle and cell division	GO:1901990	420	8	1.47	5.44	1.2E-04	4.5E-02
RNA production and processing	GO:0010467	2109	19	7.38	2.57	9.2E-05	3.7E-02
RNA production and processing	GO:1901360	3218	25	11.27	2.22	6.4E-05	3.1E-02
RNA production and processing	GO:0032239	19	3	0.07	45.1	6.0E-05	3.1E-02
RNA production and processing	GO:0046831	16	3	0.06	53.56	3.8E-05	2.2E-02
RNA production and processing	GO:0016070	1621	17	5.68	3	3.7E-05	2.3E-02
RNA production and processing	GO:0016071	701	11	2.45	4.48	3.4E-05	2.3E-02
RNA production and processing	GO:0006725	2986	25	10.45	2.39	1.7E-05	1.3E-02
RNA production and processing	GO:0046483	2936	25	10.28	2.43	1.2E-05	1.2E-02
RNA production and processing	GO:0006396	929	14	3.25	4.3	4.0E-06	4.9E-03
RNA production and processing	GO:0006139	2740	25	9.59	2.61	3.6E-06	4.8E-03
RNA production and processing	GO:0008380	410	10	1.44	6.97	1.9E-06	3.1E-03
RNA production and processing	GO:0006397	489	11	1.71	6.43	1.2E-06	2.8E-03
RNA production and processing	GO:0090304	2237	25	7.83	3.19	8.0E-08	1.3E-03



Table S7 continued

3h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
General cellular processes	GO:0043170	6337	40	22.19	1.8	1.7E-05	1.3E-02
General cellular processes	GO:0010558	1566	16	5.48	2.92	9.0E-05	3.7E-02
General cellular processes	GO:0019222	7060	41	24.72	1.66	1.0E-04	3.9E-02
General cellular processes	GO:0031327	1626	16	5.69	2.81	1.4E-04	4.8E-02
Response to DNA damage	GO:0006974	785	11	2.75	4	9.3E-05	3.6E-02

3h DOWN							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
Protein production and regulation	GO:0006614	96	7	0.16	43.44	3.5E-10	1.8E-06
Protein production and regulation	GO:0006613	101	7	0.17	41.29	4.9E-10	1.9E-06
Protein production and regulation	GO:0045047	111	7	0.19	37.57	9.1E-10	2.9E-06
Protein production and regulation	GO:0072599	115	7	0.19	36.26	1.1E-09	2.6E-06
Protein production and regulation	GO:0070972	142	7	0.24	29.37	4.6E-09	6.2E-06
Protein production and regulation	GO:0006413	145	7	0.24	28.76	5.3E-09	6.1E-06
Protein production and regulation	GO:0006612	184	7	0.31	22.66	2.6E-08	2.3E-05
Protein production and regulation	GO:0090150	294	8	0.49	16.21	2.9E-08	2.4E-05
Protein production and regulation	GO:0046907	1528	14	2.56	5.46	7.1E-08	4.9E-05
Protein production and regulation	GO:0006605	374	8	0.63	12.74	1.8E-07	1.1E-04
Protein production and regulation	GO:0072657	528	9	0.89	10.15	1.8E-07	1.1E-04
Protein production and regulation	GO:0051641	3007	18	5.05	3.57	2.8E-07	1.4E-04
Protein production and regulation	GO:0033365	776	10	1.3	7.68	4.2E-07	1.9E-04
Protein production and regulation	GO:0072594	453	8	0.76	10.52	7.4E-07	3.0E-04
Protein production and regulation	GO:0034613	1646	13	2.76	4.71	1.3E-06	5.0E-04
Protein production and regulation	GO:0070727	1655	13	2.78	4.68	1.4E-06	4.9E-04
Protein production and regulation	GO:0051649	2378	15	3.99	3.76	2.5E-06	8.4E-04
Protein production and regulation	GO:0006412	394	7	0.66	10.58	3.9E-06	1.3E-03
Protein production and regulation	GO:0043043	419	7	0.7	9.95	5.7E-06	1.8E-03
Protein production and regulation	GO:0034645	1639	12	2.75	4.36	8.6E-06	2.5E-03
Protein production and regulation	GO:0002181	75	4	0.13	31.77	9.1E-06	2.6E-03
Protein production and regulation	GO:0009059	1689	12	2.84	4.23	1.2E-05	3.2E-03
Protein production and regulation	GO:1901576	2802	15	4.7	3.19	1.9E-05	4.8E-03
Protein production and regulation	GO:0009058	2861	15	4.8	3.12	2.5E-05	6.0E-03
Protein production and regulation	GO:0006518	539	7	0.9	7.74	2.9E-05	6.8E-03



Table S7 continued

3h DOWN							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
Protein production and regulation	GO:0006886	992	9	1.67	5.4	2.9E-05	6.9E-03
Protein production and regulation	GO:0006810	4572	19	7.67	2.48	3.0E-05	7.0E-03
Protein production and regulation	GO:0043604	547	7	0.92	7.62	3.1E-05	7.1E-03
Protein production and regulation	GO:0008104	2200	13	3.69	3.52	3.2E-05	7.1E-03
Protein production and regulation	GO:0051234	4704	19	7.9	2.41	4.6E-05	9.8E-03
Protein production and regulation	GO:0034622	823	8	1.38	5.79	5.5E-05	1.1E-02
Protein production and regulation	GO:1901566	1407	10	2.36	4.23	7.7E-05	1.6E-02
Protein production and regulation	GO:0051179	5862	21	9.84	2.13	9.1E-05	1.8E-02
Protein production and regulation	GO:0033036	2564	13	4.3	3.02	1.6E-04	2.8E-02
Protein production and regulation	GO:0045184	1594	10	2.68	3.74	2.1E-04	3.7E-02
Protein production and regulation	GO:0065003	1304	9	2.19	4.11	2.4E-04	4.0E-02
RNA productio and processing	GO:0000375	306	5	0.51	9.73	1.6E-04	2.8E-02
RNA productio and processing	GO:0006397	489	6	0.82	7.31	1.6E-04	2.8E-02
RNA productio and processing	GO:0000398	303	5	0.51	9.83	1.5E-04	2.8E-02
RNA productio and processing	GO:0000377	303	5	0.51	9.83	1.5E-04	2.8E-02
RNA productio and processing	GO:0009057	1058	9	1.78	5.07	4.9E-05	1.0E-02
RNA productio and processing	GO:0071840	5919	22	9.94	2.21	3.1E-05	7.0E-03
RNA productio and processing	GO:0034660	525	7	0.88	7.94	2.4E-05	5.9E-03
RNA productio and processing	GO:0034470	433	7	0.73	9.63	7.1E-06	2.1E-03
RNA productio and processing	GO:0042254	336	7	0.56	12.41	1.4E-06	4.9E-04
RNA productio and processing	GO:1901361	491	8	0.82	9.71	1.3E-06	4.8E-04
RNA productio and processing	GO:0019439	458	8	0.77	10.41	8.0E-07	3.0E-04
RNA productio and processing	GO:1901360	3218	18	5.4	3.33	8.0E-07	3.1E-04
RNA productio and processing	GO:0046700	440	8	0.74	10.83	6.0E-07	2.6E-04
RNA productio and processing	GO:0010467	2109	15	3.54	4.24	5.5E-07	2.4E-04
RNA productio and processing	GO:0016072	270	7	0.45	15.45	3.3E-07	1.5E-04
RNA productio and processing	GO:0006725	2986	18	5.01	3.59	2.6E-07	1.3E-04
RNA productio and processing	GO:0006364	260	7	0.44	16.04	2.6E-07	1.4E-04
RNA productio and processing	GO:0034655	383	8	0.64	12.44	2.1E-07	1.2E-04
RNA productio and processing	GO:0046483	2936	18	4.93	3.65	2.0E-07	1.1E-04
RNA productio and processing	GO:0090304	2237	16	3.75	4.26	1.7E-07	1.1E-04
RNA productio and processing	GO:0016070	1621	14	2.72	5.15	1.5E-07	9.8E-05
RNA productio and processing	GO:0006139	2740	18	4.6	3.91	6.8E-08	4.9E-05
RNA productio and processing	GO:0022613	469	9	0.79	11.43	6.7E-08	5.0E-05



Table S7 continued

3h DOWN							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
RNA productio and processing	GO:0006401	251	8	0.42	18.99	8.8E-09	8.3E-06
RNA production and processing	GO:0006402	220	8	0.37	21.66	3.3E-09	4.7E-06
RNA productio and processing	GO:0006396	929	13	1.56	8.34	1.7E-09	2.7E-06
RNA productio and processing	GO:0000956	200	8	0.34	23.83	1.6E-09	3.1E-06
RNA productio and processing	GO:0016071	701	13	1.18	11.05	5.8E-11	4.6E-07
RNA productio and processing	GO:0000184	121	8	0.2	39.39	3.5E-11	5.5E-07
Cellular respirarion and energy production	GO:0006119	121	7	0.2	34.46	1.6E-09	2.8E-06
Cellular respirarion and energy production	GO:0046034	210	7	0.35	19.86	6.2E-08	5.0E-05
Cellular respirarion and energy production	GO:0006091	414	7	0.69	10.07	5.3E-06	1.7E-03
Cellular respirarion and energy production	GO:0016310	1304	11	2.19	5.03	6.1E-06	1.9E-03
Cellular respirarion and energy production	GO:0022900	178	5	0.3	16.73	1.3E-05	3.5E-03
Cellular respirarion and energy production	GO:0042775	89	4	0.15	26.77	1.7E-05	4.5E-03
Cellular respirarion and energy production	GO:0042773	90	4	0.15	26.48	1.8E-05	4.6E-03
Cellular respirarion and energy production	GO:0022904	110	4	0.18	21.66	3.9E-05	8.4E-03
Cellular respirarion and energy production	GO:1902600	138	4	0.23	17.27	9.1E-05	1.8E-02
Cellular respirarion and energy production	GO:0007005	467	6	0.78	7.65	1.2E-04	2.4E-02
Cellular respirarion and energy production	GO:0045333	160	4	0.27	14.89	1.6E-04	2.8E-02
Antibacterial response	GO:0019731	56	3	0.09	31.91	1.3E-04	2.5E-02
Antibacterial response	GO:0044419	2110	12	3.54	3.39	1.1E-04	2.1E-02
Viral processes	GO:0044403	912	9	1.53	5.88	1.5E-05	4.0E-03
Viral processes	GO:0016032	820	9	1.38	6.54	6.5E-06	2.0E-03
Viral processes	GO:0019080	154	7	0.26	27.08	8.0E-09	8.4E-06
Viral processes	GO:0019083	115	7	0.19	36.26	1.1E-09	3.0E-06
Viral processes	GO:0006807	7090	23	11.9	1.93	2.1E-04	3.7E-02
General cellular processes	GO:0044238	7570	24	12.71	1.89	1.4E-04	2.6E-02
General cellular processes	GO:0044248	1825	11	3.06	3.59	1.3E-04	2.6E-02
General cellular processes	GO:0044260	5145	20	8.64	2.32	4.3E-05	9.2E-03
General cellular processes	GO:0044265	924	9	1.55	5.8	1.7E-05	4.4E-03
General cellular processes	GO:0044249	2699	15	4.53	3.31	1.2E-05	3.4E-03
General cellular processes	GO:0034641	3401	18	5.71	3.15	1.8E-06	6.2E-04
General cellular processes	GO:0044271	1571	13	2.64	4.93	7.9E-07	3.2E-04
General cellular processes	GO:0044270	441	8	0.74	10.81	6.1E-07	2.5E-04
General cellular processes	GO:0044085	2656	17	4.46	3.81	2.9E-07	1.4E-04
General cellular processes	GO:0008152	8585	31	14.41	2.15	8.0E-09	7.9E-06

Table S7 continued

3h DOWN							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
General cellular processes	GO:0044237	7782	30	13.06	2.3	5.2E-09	6.3E-06

4h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
CNS development	GO:0045664	680	14	4.27	3.28	1.1E-04	3.8E-02
Cell cycle and cell division	GO:0051276	1062	21	6.67	3.15	3.5E-06	2.3E-03
Cell cycle and cell division	GO:0006260	223	9	1.4	6.42	1.6E-05	6.7E-03
Cell cycle and cell division	GO:0006338	175	8	1.1	7.28	2.0E-05	8.4E-03
Cell cycle and cell division	GO:0071103	309	10	1.94	5.15	3.3E-05	1.3E-02
Cell cycle and cell division	GO:0032508	110	6	0.69	8.68	8.9E-05	3.2E-02
Cell cycle and cell division	GO:0051130	1229	20	7.72	2.59	9.5E-05	3.3E-02
Cell cycle and cell division	GO:0051128	2436	31	15.3	2.03	1.1E-04	3.8E-02
Cell cycle and cell division	GO:0032392	118	6	0.74	8.09	1.3E-04	4.2E-02
Cell cycle and cell division	GO:0033044	367	10	2.31	4.34	1.3E-04	4.2E-02
Cell cycle and cell division	GO:1905269	122	6	0.77	7.83	1.5E-04	4.8E-02
RNA production and processing	GO:0008380	410	17	2.58	6.6	1.4E-09	2.3E-05
RNA production and processing	GO:0019219	4078	54	25.62	2.11	1.4E-08	1.1E-04
RNA production and processing	GO:0006397	489	17	3.07	5.53	1.8E-08	9.5E-05
RNA production and processing	GO:0051252	3809	51	23.93	2.13	4.0E-08	1.6E-04
RNA production and processing	GO:0090304	2237	36	14.05	2.56	9.4E-08	3.0E-04
RNA production and processing	GO:0045934	1572	29	9.88	2.94	1.5E-07	2.9E-04
RNA production and processing	GO:0010468	4913	58	30.87	1.88	2.4E-07	4.2E-04
RNA production and processing	GO:0006396	929	21	5.84	3.6	4.4E-07	5.4E-04
RNA production and processing	GO:0016071	701	18	4.4	4.09	5.5E-07	6.2E-04
RNA production and processing	GO:0006139	2740	38	17.21	2.21	1.7E-06	1.4E-03
RNA production and processing	GO:0006355	3462	44	21.75	2.02	2.6E-06	1.8E-03
RNA production and processing	GO:1903506	3531	44	22.18	1.98	3.6E-06	2.2E-03
RNA production and processing	GO:2001141	3536	44	22.22	1.98	3.7E-06	2.2E-03
RNA production and processing	GO:0000398	303	11	1.9	5.78	4.6E-06	2.6E-03
RNA production and processing	GO:0000377	303	11	1.9	5.78	4.6E-06	2.5E-03
RNA production and processing	GO:0000375	306	11	1.92	5.72	5.0E-06	2.7E-03
RNA production and processing	GO:0046483	2936	38	18.45	2.06	1.2E-05	5.9E-03
RNA production and processing	GO:1903507	1351	23	8.49	2.71	1.3E-05	6.3E-03



Table S7 continued

4h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
RNA production and processing	GO:1902679	1353	23	8.5	2.71	1.3E-05	6.2E-03
RNA production and processing	GO:0051253	1455	24	9.14	2.63	1.4E-05	6.2E-03
RNA production and processing	GO:0006725	2986	38	18.76	2.03	1.5E-05	6.5E-03
RNA production and processing	GO:1901360	3218	39	20.22	1.93	4.6E-05	1.8E-02
RNA production and processing	GO:0045892	1310	21	8.23	2.55	7.7E-05	2.8E-02
General cellular processes	GO:0034641	3401	40	21.37	1.87	7.0E-05	2.6E-02
General cellular processes	GO:0031324	2700	34	16.96	2	6.6E-05	2.5E-02
General cellular processes	GO:0051053	140	7	0.88	7.96	3.9E-05	1.5E-02
General cellular processes	GO:0051172	2497	34	15.69	2.17	1.1E-05	5.6E-03
General cellular processes	GO:0080090	6118	64	38.44	1.67	3.0E-06	2.0E-03
General cellular processes	GO:0051171	5920	63	37.19	1.69	2.2E-06	1.6E-03
General cellular processes	GO:0009889	4270	51	26.83	1.9	1.9E-06	1.5E-03
General cellular processes	GO:0019222	7060	72	44.36	1.62	1.0E-06	8.5E-04
General cellular processes	GO:0031326	4184	51	26.29	1.94	8.8E-07	7.7E-04
General cellular processes	GO:0031323	6329	67	39.76	1.68	8.2E-07	7.7E-04
General cellular processes	GO:2000113	1512	27	9.5	2.84	8.1E-07	8.0E-04
General cellular processes	GO:0060255	6510	69	40.9	1.69	5.7E-07	6.1E-04
General cellular processes	GO:0009890	1657	29	10.41	2.79	4.3E-07	5.7E-04
General cellular processes	GO:0010556	4032	51	25.33	2.01	3.2E-07	4.6E-04
General cellular processes	GO:0031327	1626	29	10.22	2.84	2.9E-07	4.7E-04
General cellular processes	GO:0010558	1566	29	9.84	2.95	1.4E-07	3.1E-04
General cellular processes	GO:2000112	3924	51	24.65	2.07	1.2E-07	3.1E-04

4h DOWN							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
Osteoblast differentiation	GO:0045668	48	4	0.09	43.44	2.9E-06	4.6E-02
RNA production and processing	GO:0000184	121	5	0.23	21.54	4.1E-06	2.2E-02
Osteoblast differentiation	GO:0045667	121	5	0.23	21.54	4.1E-06	3.3E-02

6h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
RNA production and processing	GO:0010629	2065	42	23.47	1.79	2.7E-04	4.4E-02
RNA production and processing	GO:1903311	351	13	3.99	3.26	2.6E-04	4.3E-02

Table S7 continued

6h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
RNA production and processing	GO:0050684	154	9	1.75	5.14	9.8E-05	1.9E-02
RNA production and processing	GO:0010467	2109	44	23.97	1.84	8.5E-05	1.7E-02
RNA production and processing	GO:0000375	306	13	3.48	3.74	7.0E-05	1.4E-02
RNA production and processing	GO:0045892	1310	32	14.89	2.15	7.0E-05	1.4E-02
RNA production and processing	GO:0000398	303	13	3.44	3.77	6.4E-05	1.4E-02
RNA production and processing	GO:0000377	303	13	3.44	3.77	6.4E-05	1.4E-02
RNA production and processing	GO:0045893	1601	37	18.2	2.03	4.2E-05	1.1E-02
RNA production and processing	GO:0016071	701	22	7.97	2.76	2.4E-05	6.4E-03
RNA production and processing	GO:1902679	1353	34	15.38	2.21	1.7E-05	4.9E-03
RNA production and processing	GO:1903507	1351	34	15.36	2.21	1.6E-05	4.8E-03
RNA production and processing	GO:0051253	1455	36	16.54	2.18	1.6E-05	4.9E-03
RNA production and processing	GO:0045935	1953	44	22.2	1.98	1.6E-05	4.9E-03
RNA production and processing	GO:0016070	1621	39	18.42	2.12	1.4E-05	4.5E-03
RNA production and processing	GO:1901360	3218	63	36.58	1.72	1.2E-05	3.9E-03
RNA production and processing	GO:1902680	1687	40	19.18	2.09	1.2E-05	3.9E-03
RNA production and processing	GO:1903508	1686	40	19.16	2.09	1.2E-05	4.0E-03
RNA production and processing	GO:0006357	2628	55	29.87	1.84	7.8E-06	2.7E-03
RNA production and processing	GO:0006397	489	19	5.56	3.42	5.2E-06	2.0E-03
RNA production and processing	GO:0006725	2986	61	33.94	1.8	4.1E-06	1.6E-03
RNA production and processing	GO:0051254	1780	43	20.23	2.13	3.4E-06	1.4E-03
RNA production and processing	GO:0046483	2936	61	33.37	1.83	2.3E-06	1.0E-03
RNA production and processing	GO:0008380	410	18	4.66	3.86	1.8E-06	9.1E-04
RNA production and processing	GO:0006396	929	30	10.56	2.84	4.1E-07	2.6E-04
RNA production and processing	GO:0045934	1572	42	17.87	2.35	3.8E-07	2.5E-04
RNA production and processing	GO:0006139	2740	61	31.14	1.96	2.2E-07	1.7E-04
RNA production and processing	GO:0090304	2237	57	25.43	2.24	6.1E-09	5.4E-06
RNA production and processing	GO:0006355	3462	78	39.35	1.98	1.2E-09	1.6E-06
RNA production and processing	GO:2001141	3536	79	40.19	1.97	1.1E-09	1.6E-06
RNA production and processing	GO:1903506	3531	79	40.13	1.97	1.0E-09	1.7E-06
RNA production and processing	GO:0010468	4913	100	55.84	1.79	4.1E-10	8.1E-07
RNA production and processing	GO:0051252	3809	89	43.29	2.06	4.7E-12	3.7E-08
RNA production and processing	GO:0019219	4078	94	46.35	2.03	1.5E-12	2.4E-08
CNS development	GO:0003008	2079	7	23.63	0.3	6.6E-05	1.4E-02
CNS development	GO:0050877	1413	3	16.06	0.19	1.2E-04	2.2E-02

Table S7 continued

6h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
CNS development	GO:0007600	987	1	11.22	0.09	2.7E-04	4.3E-02
Stemness and stemcell maintenance	GO:0098727	143	9	1.63	5.54	5.7E-05	1.3E-02
Stemness and stemcell maintenance	GO:0019827	141	9	1.6	5.62	5.2E-05	1.2E-02
Embryonic development	GO:0001701	360	15	4.09	3.67	2.4E-05	6.3E-03
Embryonic development	GO:0048608	433	16	4.92	3.25	5.3E-05	1.2E-02
Embryonic development	GO:0061458	437	16	4.97	3.22	5.8E-05	1.3E-02
Embryonic development	GO:0009792	643	20	7.31	2.74	6.5E-05	1.4E-02
Embryonic development	GO:0009790	1003	26	11.4	2.28	1.1E-04	2.0E-02
Embryonic development	GO:0043009	623	19	7.08	2.68	1.3E-04	2.3E-02
Embryonic development	GO:0010171	46	5	0.52	9.56	2.7E-04	4.3E-02
Cell cycle and cell division	GO:0071103	309	12	3.51	3.42	3.0E-04	4.7E-02
Cell cycle and cell division	GO:0031023	98	7	1.11	6.28	1.9E-04	3.2E-02
Cell cycle and cell division	GO:0033043	1332	31	15.14	2.05	1.6E-04	2.9E-02
Cell cycle and cell division	GO:0022402	1069	27	12.15	2.22	1.5E-04	2.8E-02
Cell cycle and cell division	GO:0007098	86	7	0.98	7.16	8.6E-05	1.7E-02
Cell cycle and cell division	GO:0006338	175	10	1.99	5.03	4.8E-05	1.2E-02
Cell cycle and cell division	GO:1905269	122	9	1.39	6.49	1.8E-05	5.1E-03
Cell cycle and cell division	GO:0016043	5699	97	64.78	1.5	7.1E-06	2.6E-03
Cell cycle and cell division	GO:2001251	138	10	1.57	6.38	7.0E-06	2.6E-03
Cell cycle and cell division	GO:0071840	5919	101	67.28	1.5	3.2E-06	1.4E-03
Cell cycle and cell division	GO:1902275	213	13	2.42	5.37	1.8E-06	9.2E-04
Cell cycle and cell division	GO:0007049	1390	38	15.8	2.41	6.6E-07	3.9E-04
Cell cycle and cell division	GO:0006996	3576	73	40.65	1.8	3.7E-07	2.6E-04
Cell cycle and cell division	GO:0033044	367	19	4.17	4.55	8.5E-08	7.1E-05
Cell cycle and cell division	GO:0006325	701	29	7.97	3.64	3.9E-09	3.9E-06
Cell cycle and cell division	GO:0051276	1062	40	12.07	3.31	4.7E-11	2.5E-07
Protein production and regulation	GO:0010557	1930	42	21.94	1.91	6.6E-05	1.4E-02
Protein production and regulation	GO:0010604	3630	67	41.26	1.62	4.6E-05	1.1E-02
General cellular processes	GO:0009890	1657	41	18.83	2.18	3.0E-06	1.3E-03
General cellular processes	GO:0031327	1626	41	18.48	2.22	2.3E-06	1.1E-03
General cellular processes	GO:2000113	1512	39	17.19	2.27	2.0E-06	9.5E-04
General cellular processes	GO:0010558	1566	41	17.8	2.3	7.4E-07	4.2E-04
General cellular processes	GO:0010605	2951	62	33.54	1.85	1.4E-06	7.5E-04
General cellular processes	GO:0009892	3205	64	36.43	1.76	5.0E-06	1.9E-03

Table S7 continued

6h UP							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
General cellular processes	GO:0009889	4270	88	48.53	1.81	5.6E-09	5.3E-06
General cellular processes	GO:0031326	4184	88	47.56	1.85	1.6E-09	1.9E-06
General cellular processes	GO:2000112	3924	86	44.6	1.93	3.4E-10	7.7E-07
General cellular processes	GO:0010556	4032	88	45.83	1.92	2.8E-10	7.3E-07
General cellular processes	GO:0019222	7060	126	80.25	1.57	1.9E-09	2.0E-06
General cellular processes	GO:0034641	3401	65	38.66	1.68	1.9E-05	5.3E-03
General cellular processes	GO:0043170	6337	109	72.03	1.51	5.6E-07	3.4E-04
General cellular processes	GO:0051172	2497	57	28.38	2.01	3.2E-07	2.3E-04
General cellular processes	GO:0006807	7090	114	80.59	1.41	9.2E-06	3.2E-03
General cellular processes	GO:0009893	3937	69	44.75	1.54	1.6E-04	2.9E-02
General cellular processes	GO:0051173	3267	61	37.13	1.64	7.1E-05	1.4E-02
General cellular processes	GO:0051171	5920	113	67.29	1.68	4.6E-10	8.0E-07
General cellular processes	GO:0031324	2700	60	30.69	1.96	3.1E-07	2.3E-04
General cellular processes	GO:0048523	4981	89	56.62	1.57	3.7E-06	1.5E-03
General cellular processes	GO:0051053	140	9	1.59	5.66	4.9E-05	1.2E-02
General cellular processes	GO:0031323	6329	119	71.94	1.65	2.5E-10	7.8E-07
General cellular processes	GO:0060255	6510	123	74	1.66	7.6E-11	3.0E-07
General cellular processes	GO:0080090	6118	114	69.54	1.64	1.7E-09	1.9E-06
General cellular processes	GO:0009891	2066	42	23.48	1.79	2.7E-04	4.4E-02
General cellular processes	GO:0031328	2031	42	23.09	1.82	1.6E-04	2.8E-02
General cellular processes	GO:0065007	12629	179	143.55	1.25	1.7E-06	9.1E-04
General cellular processes	GO:0048519	5643	95	64.14	1.48	1.8E-05	5.1E-03
General cellular processes	GO:0031325	3454	63	39.26	1.6	1.0E-04	2.0E-02
General cellular processes	GO:0048522	5742	96	65.27	1.47	2.0E-05	5.5E-03
General cellular processes	GO:0050789	11955	167	135.88	1.23	3.9E-05	9.8E-03
General cellular processes	GO:0050794	11390	162	129.46	1.25	2.4E-05	6.5E-03
Epigenetic regulation	GO:0031062	43	5	0.49	10.23	2.0E-04	3.4E-02
Epigenetic regulation	GO:0031060	72	7	0.82	8.55	3.0E-05	7.8E-03
Cell death	GO:0006915	918	25	10.43	2.4	8.3E-05	1.7E-02
Cell death	GO:0008219	1087	27	12.36	2.19	1.8E-04	3.1E-02
Cell death	GO:0012501	1049	26	11.92	2.18	2.6E-04	4.3E-02



Table S7 continued

6h DOWN							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
Embryonic development	GO:0048638	342	7	1.13	6.19	1.50E-04	4.2E-02
Embryonic development	GO:0042127	1677	17	5.55	3.06	2.62E-05	1.2E-02
Embryonic development	GO:0040008	685	11	2.27	4.85	1.61E-05	9.1E-03
RNA production and processing	GO:2001141	3536	25	11.7	2.14	1.37E-04	3.9E-02
RNA production and processing	GO:1903506	3531	25	11.68	2.14	1.35E-04	3.9E-02
RNA production and processing	GO:0006355	3462	25	11.46	2.18	1.19E-04	3.7E-02
RNA production and processing	GO:0010468	4913	31	16.26	1.91	9.57E-05	3.1E-02
RNA production and processing	GO:0051252	3809	27	12.6	2.14	5.56E-05	2.0E-02
RNA production and processing	GO:0019219	4078	28	13.49	2.07	5.49E-05	2.0E-02
RNA production and processing	GO:1902679	1353	17	4.48	3.8	1.56E-06	1.6E-03
RNA production and processing	GO:1903507	1351	17	4.47	3.8	1.53E-06	1.6E-03
RNA production and processing	GO:0045892	1310	17	4.34	3.92	1.00E-06	1.3E-03
RNA production and processing	GO:0051253	1455	18	4.81	3.74	8.86E-07	1.3E-03
RNA production and processing	GO:0000122	970	15	3.21	4.67	5.80E-07	9.2E-04
RNA production and processing	GO:0045934	1572	19	5.2	3.65	5.76E-07	1.0E-03
RNA production and processing	GO:0010629	2065	23	6.83	3.37	1.08E-07	1.7E-03
General cellular processes	GO:0009892	3205	29	10.61	2.73	1.09E-07	8.7E-04
General cellular processes	GO:0010605	2951	27	9.77	2.76	3.06E-07	8.1E-04
General cellular processes	GO:0048519	5643	39	18.67	2.09	3.67E-07	8.3E-04
General cellular processes	GO:0019222	7060	44	23.36	1.88	5.64E-07	1.1E-03
General cellular processes	GO:0031324	2700	24	8.93	2.69	3.17E-06	2.5E-03
General cellular processes	GO:0048523	4981	34	16.48	2.06	5.85E-06	4.4E-03
General cellular processes	GO:2000113	1512	17	5	3.4	6.85E-06	5.0E-03
General cellular processes	GO:0031323	6329	39	20.94	1.86	8.78E-06	6.0E-03
General cellular processes	GO:0010558	1566	17	5.18	3.28	1.08E-05	6.9E-03
General cellular processes	GO:0051172	2497	22	8.26	2.66	1.11E-05	6.8E-03
General cellular processes	GO:0031327	1626	17	5.38	3.16	1.76E-05	9.7E-03
General cellular processes	GO:0010604	3630	27	12.01	2.25	1.90E-05	9.8E-03
General cellular processes	GO:0060255	6510	39	21.54	1.81	1.98E-05	9.9E-03
General cellular processes	GO:0009890	1657	17	5.48	3.1	2.25E-05	1.0E-02
General cellular processes	GO:0080090	6118	37	20.25	1.83	2.79E-05	1.2E-02
General cellular processes	GO:0009893	3937	28	13.03	2.15	3.43E-05	1.4E-02
General cellular processes	GO:0065009	3078	23	10.19	2.26	1.07E-04	3.4E-02
General cellular processes	GO:0051171	5920	35	19.59	1.79	1.32E-04	3.9E-02

Table S7 continued

6h DOWN							
General term	GO #	in Ref	in data	exp-ected	fold Enr.	raw P-value	FDR
General cellular processes	GO:0010556	4032	27	13.34	2.02	1.67E-04	4.6E-02
Kinase activity and regulation	GO:0010563	577	9	1.91	4.71	1.27E-04	3.8E-02
Kinase activity and regulation	GO:0045936	576	9	1.91	4.72	1.25E-04	3.8E-02
Kinase activity and regulation	GO:0050790	2397	20	7.93	2.52	6.98E-05	2.3E-02
Kinase activity and regulation	GO:0033674	616	10	2.04	4.91	3.71E-05	1.4E-02
Kinase activity and regulation	GO:0031399	1891	18	6.26	2.88	3.29E-05	1.3E-02
Kinase activity and regulation	GO:0042326	456	9	1.51	5.96	2.13E-05	10.0E-03
Kinase activity and regulation	GO:0001932	1469	16	4.86	3.29	2.01E-05	9.7E-03
Kinase activity and regulation	GO:0051347	696	11	2.3	4.78	1.86E-05	9.9E-03
Kinase activity and regulation	GO:0045860	539	10	1.78	5.61	1.21E-05	7.1E-03
Kinase activity and regulation	GO:0045859	805	13	2.66	4.88	2.31E-06	2.2E-03
Kinase activity and regulation	GO:0051174	1842	20	6.1	3.28	1.43E-06	1.6E-03
Kinase activity and regulation	GO:0019220	1841	20	6.09	3.28	1.41E-06	1.7E-03
Kinase activity and regulation	GO:0043549	915	15	3.03	4.95	2.79E-07	8.9E-04
Kinase activity and regulation	GO:0042325	1642	20	5.43	3.68	2.35E-07	9.3E-04
Kinase activity and regulation	GO:0051338	1032	16	3.42	4.69	2.19E-07	1.2E-03
Osteoblast differentiation	GO:0045667	121	5	0.4	12.49	6.04E-05	2.1E-02
Osteoblast differentiation	GO:0030278	205	6	0.68	8.84	6.92E-05	2.3E-02
Osteoblast differentiation	GO:0030279	80	4	0.26	15.11	1.71E-04	4.6E-02
Cell death	GO:0042981	1566	18	5.18	3.47	2.51E-06	2.2E-03
Cell death	GO:0043067	1588	18	5.25	3.43	3.06E-06	2.6E-03
Cell death	GO:0010941	1722	18	5.7	3.16	9.37E-06	6.2E-03
Cell death	GO:0060548	1027	13	3.4	3.83	3.03E-05	1.3E-02
Cell death	GO:0043066	915	12	3.03	3.96	4.57E-05	1.7E-02
Cell death	GO:0043069	933	12	3.09	3.89	5.50E-05	2.0E-02

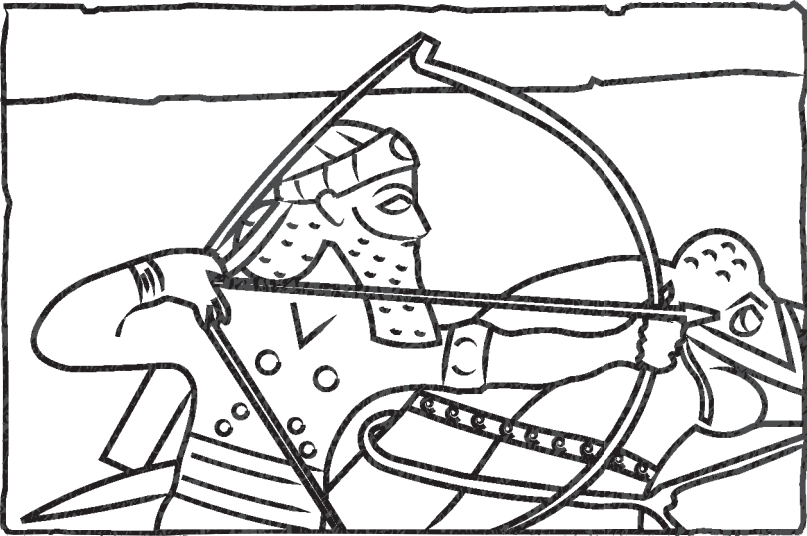


Illustration based on a stone carving on display at the British museum

Chapter 4

A genome-wide CRISPR/Cas phenotypic screen for modulators of DUX4 cytotoxicity reveals screen complications

Ator Ashoti^{1*}, Francesco Limone^{2,5*}, Melissa van Kranenburg¹, Anna Alemany¹, Mirna Baak¹, Judith Vivié^{1,3}, Federica Piccioni⁷, Menno Creyghton^{1,6}, Kevin Eggen^{2,5**}, Niels Geijsen^{1,4**}

¹Hubrecht Institute, Developmental Biology and Stem Cell Research, Utrecht, Netherlands

²Department of Stem Cell and Regenerative Biology, Harvard University
Cambridge, MA, USA

³Single Cell Discoveries, Utrecht, Netherlands

⁴Leiden University Medical Center, Leiden, The Netherlands

⁵Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard,
Cambridge, MA, USA

⁶Erasmus University Medical Centre, Rotterdam, The Netherlands

⁷Broad institute, Cambridge, MA, USA

* Equal contribution

**Correspondence: n.geijsen@lumc.nl, eggen@mcb.harvard.edu

This chapter is available on bioRxiv

Ashoti, A. et al. A genome-wide CRISPR/Cas phenotypic screen for modulators of DUX4 cytotoxicity reveals screen complications. bioRxiv (2020).
doi:10.1101/2020.07.27.223420

Abstract

Facioscapulohumeral muscular dystrophy (FHD), a fundamentally complex muscle disorder that thus far remains untreatable. As the name implies, FHD starts in the muscles of the face and shoulder girdle. The main perturber of the disease is the pioneer transcription factor DUX4, which is misexpressed in affected tissues due to a failure in epigenetic repressive mechanisms. In pursuit of unraveling the underlying mechanism of FHD and finding potential therapeutic targets or treatment options, we performed an exhaustive genome-wide CRISPR/Cas9 phenotypic rescue screen to identify modulators of DUX4 cytotoxicity. We found no key effectors other than DUX4 itself, suggesting treatment efforts in FHD should be directed towards its direct modulation.

The screen did however reveal some rare and unexpected Cas9-induced genomic events, that may provide important considerations for planning future CRISPR/Cas9 knock-out screens.

Introduction

Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant degenerative muscle disease. It's one of the most prevalent neuromuscular disorder¹, characterized by progressive and asymmetric muscle weakness which generally starts in facial muscles, and then slowly progresses to muscles of the shoulders, upper limbs and eventually the lower extremities². Age of onset is highly variable, but calculations based on a 122 case study demonstrates that the mean age of onset is in the early twenties (21-23)³. The primary cause of the disease is the misexpression of the double homeobox 4 (DUX4) transcription factor, due to failure in epigenetic silencing³⁻⁶. DUX4 is normally expressed early in development in the cleavage stage embryo^{7,8}, in the adult testis⁶ and in the thymus⁹. De-repression of DUX4 in muscle activates a large cascade of events, triggering the activation of many pathways^{8,10-19}, with target genes being involved in biological processes such as RNA splicing and processing (DBR1^{10,20-22}, CWC15^{10,20,22}, PNN^{10,21}, CLP1^{10,21,22}, TFIP11^{10,20-22}), spermatogenesis (CCNA1^{10,20-22}, ZNF296^{10,20-22}, TESK2^{10,20,21}), early embryonic development (ZSCAN4^{10,20-22}, LEUTX²⁰⁻²², STIL^{10,20,21}), protein processing and degradation (SIAH1^{10,20-22}, RHOBTB1^{10,20,21}, TRIM36^{10,20,21}), and cell motility and migration (CXCR4^{10,20,21}, ROCK1^{10,21}, SNAI1^{10,20-22}).

We hypothesized that of one or more factors downstream of DUX4 expression are responsible for the rapid apoptotic response that follows DUX4 induction. Knowing if there are key downstream targets of DUX4 can have important clinical applications as they could direct intelligent therapy design. We tested this hypothesis by performing a genome wide CRISPR/Cas9 knockout screen.

CRISPR/Cas9, which is now a highly popular and widely used genome editing technique, was initially discovered as the adaptive immune system of bacteria, to protect against viral infection^{23,24}. Although not the first genome editing method, CRISPR/Cas9 has proven to be much more user friendly due to its easy manipulability, and being more cost-, labor- and time-efficient compared to its predecessors: transcription activator-like effector nucleases (TALENs)²⁵⁻²⁸ and ZINC-fingers nucleases (ZFNs)²⁹⁻³³. Its ability to knock-out any gene by creating a double stranded break³⁴⁻³⁶ in such an easy manner, makes this technique very suitable for genome-wide loss of function studies. The advantages and ease-of-use of the CRISPR/Cas9 technology inspired us to perform a genome wide screen on a FSHD *in-vitro* model, to find potential modulators that contribute or aggravate the FSHD pathophysiology. Successful performance of a FSHD genome-wide screen will critically depend on the cell system being used. The cells should be highly proliferative, easily transfected and display a robust DUX4-induced phenotype. These parameters make human primary myoblasts a less suitable basis for a genome-wide CRISPR Knockout screen. Fortunately, DUX4 is a so-called pioneer factor^{37,38}, capable of regulating its target genes independent of their chromatin-state. The network of genes activated by pioneer factors is therefore less affected by cellular identity. Indeed, Jones and colleagues have demonstrated that DUX4 activates the same downstream target genes in B-lymphocytes as previously identified in skeletal muscle myoblasts^{39,40}. Using an adherent leukemic cell line that is frequently used for genome-wide screening purposes (KBM7^{41,42}), we performed an exhaustive CRISPR knockout screen to identify factors that could mitigate DUX4-induced cytotoxicity. We had inserted a doxycycline-inducible DUX4 transgene into the adherent KBM7 cells^{41,42} to generate DUX4 inducible expression (DIE) cells. Using the Brunello CRISPR/Cas9 library⁴³, we screened for modulators of DUX4 cytotoxicity. Our results suggest that no single gene knockout is capable of rescuing DUX4-triggered apoptosis in our transgene model system.

This study does however, provide some interesting insight into critical parameters that need to be considered when executing a genome-wide CRISPR screen.

Results

Genome-wide CRISPR Screen reveals large chromosomal truncations

Using our DIE cell system, we sought out to identify modulators of DUX4 cytotoxicity by performing a genome wide CRISPR/Cas9 knockout screen. The Brunello human CRISPR knockout pooled library was used for this purpose⁴³. This library contains 77.441 gRNAs targeting all protein coding genes, with an average of 4 gRNAs per gene as well as 1000 non-targeting control gRNAs. To optimize the signal-to-noise ratio of the experimental system, we titrated the timing and dose of the doxycycline-mediated DUX4 induction and selected two conditions, low (250ng/ml) and high (1000ng/ml) doxycycline with an exposure time of 24h (Fig. 1A). At these concentrations 95 to 99 % of the cells die, respectively. Figure 1B and S1 outline the setup of the screen. In addition to the high and low doxycycline concentrations, cells were harvested at two timepoints after doxycycline exposure to allow recovery, early (24h) and late (72h), ultimately resulting in 4 separate 4 screens; low doxycycline/early harvest, low doxycycline/late harvest, high doxycycline/early harvest, and high doxycycline/late harvest.

Upon doxycycline administration and induction of DUX4 expression, cells from the surviving populations were harvested, genomic DNA was extracted and the gRNA sequence was amplified and sent for sequencing. Sequencing results of the treated samples revealed a large number of significantly enriched hits (Fig. 1C and S2). This included DUX4 itself and some other hits performing as well as the DUX4 gRNAs. However, upon closer examination it became clear that the majority of these enriched guides were located on the q arm of chromosome 5, suggesting an FSHD unrelated experimental artefact. Since the rtTA transgene responsible for DUX4 induction is located on the 5q arm, it is likely that when Cas9 is being targeted to the q-arm of chromosome 5 it leads to the removal the rtTA transgene, potentially through generation of a large deletion, chromosomal truncation or chromosomal rearrangement. It appears that as the rtTA integration site is located at the end of chromosome 5q, each target upstream of this site (towards the centromere) can cause a Cas9-mediated truncation, thereby removing the rtTA. (Fig. 1D, for phenograms of all 4 screens see Fig. S3). The correlation between the significance of a hit and its position along chromosome 5 highlights the strong association of these unexpected chromosomal rearrangements and the integration of rtTA at the end of chromosome 5, where the most significant hits reside in all four screens (Supplementary Figure S4).

Some of these 5q locating guides (Fig. 1E) were tested individually in DIE cells containing a constitutively expressing Cas9 in its genome (DIE-Cas9), and without selecting for the rtTA and DUX4 transgene. No increased survival was detected compared to the background surviving cells that are seen in the control situation (Fig. 1F). This suggests that the Cas9-induced truncation of a chromosomal arm and subsequent removal of rtTA activity is a rare event that was only identified due to the high sensitivity of our screen.

Data shown here was analyzed one-sidedly, and only truly represent enrichment. When analyzing the screen data double sided, one can again notice a clear enrichment of gRNA sequences, however no real depletion is seen (Fig. S5).

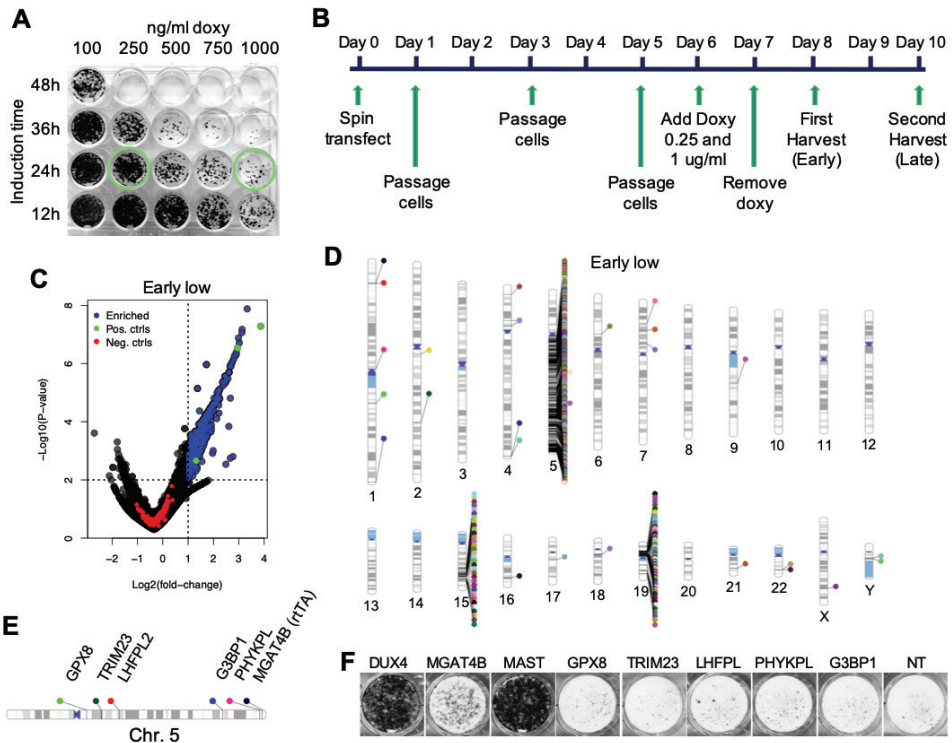


Figure 1. CRISPR Screen set up and discovery of a Cas9 artefact. (A) Viability staining of DIE cells treated with a doxycycline titration curve to determine which concentration and exposure time to use to induce sufficient cell death rates in DIE cells. Green circles indicate which conditions were used for the genome wide CRISPR/Cas9 screen. (B) The CRISPR/Cas9 screen timeline from the moment of library transfection (Day 0) to the final harvest of surviving DIE cells (Day 10). (C) Volcano plot showing the enrichment of sets of guides of the low doxycycline/early harvest screen. For a two-sided analysis see supplementary figure S5. Data shown here shows the average $\log_2(\text{foldchange})$ and $-\log_{10}(\text{p-value})$ of each guide set (set: 4 guides per gene). The $\log_2(\text{foldchange})$ is plotted on the x-axis and the $-\log_{10}(\text{p-value})$ on the y-axis. Blue points represent guide sets that are significantly enriched ($P\text{-value} \leq 0.01$, $\text{LFC} \geq 1$), green point are the positive controls (DUX4, MAST1, MGAT4B), red points represent the Non-Target/negative control guides. (D) Chromosomal ideogram indicating the location of enriched hits in the human genome, of the low doxycycline/early harvest screen. (E) Schematic representation of the location of a small number of false positive hits on chromosome 5. (F) Viability staining demonstrating surviving DIE-Cas9 cells, containing knock-outs of the same genes mentioned in (E), after 250ng/ml doxycycline exposure. Media did not contain any selection markers. NT: Non-Target controls.

Filtered Genome wide CRISPR screen results reveal no single targetable gene

Since potential hits were likely obscured by the large number of false-positive hits that resulted from Cas9-mediated elimination of either the DUX4 or the rtTA transgenes, we filtered the screen results to remove all hits located on the q-arm of chromosome 5, or the p-arm of chromosome 19 (Figure 2A). After analyzing individual guides for their apparent effectiveness in the genome wide screen (instead of the group average), a list of potential hits emerged ($p\text{-value} \leq 0.05$, $\text{Log}_2(\text{foldchange}) \geq 1$) for each of the 4 screens. Figure 2B shows the number of potential hits that met these criteria for each screen and how many of these hits are shared between them (See Table S1 to S4 for the lists of potential hits). We further focused on hits that emerged in at least 3 out of the 4 screens. Hits were validated

by performing individual knock outs in the DIE-Cas9 cells, now also containing an inducible eGFP in its genome (DIE -ieGFP-Cas9). The tetracycline response element (TRE) controlling eGFP expression is identical to the TRE controlling DUX4 expression. If there is a true target that can mitigate the apoptotic phenotype without interfering with the inducible system, these positively targeted cells should not only survive but also emit an eGFP signal upon doxycycline admission (Fig. S6). Results show that MED25 increased cell survival when knocked-out (Fig. 2C). MED25 is a subunit of Mediator, a large complex that functions as a bridge between transcription factors and the transcriptional machinery. This includes RNA polymerase II, needed for the transcription of all protein coding genes in eukaryotes (reviewed by Soutourina⁴⁴). The rescue seen in doxycycline induced DIE cells after MED25 knock-out diminishes upon higher doxycycline exposure, suggesting that loss of MED25 provides a partial rescue. Other genes belonging to the same mediator complex, that initially didn't meet our criteria, were reevaluated by lowering the parameters ($P \leq 0.05$, foldchange of ≥ 1.5), identifying a number of other subunits. When individual knock-outs of these genes were performed, two more subunits of the Mediator-complex showed partial rescue (Fig. 2D). Finally, the individually tested KO cells were analyzed by flow cytometry, for the detection of eGFP. FACS analysis reveals that Mediator-complex components have a general effect on the inducible transcription of DUX4, since the knock-out of Mediator genes did not induce eGFP expression in surviving DIE cells (Fig. 2E). This suggest that their survival was due to a generally reduced ability of rtTA to mediate transgene activation.

4

In a recent study by Shadle and colleagues, a siRNA screen was performed targeting the “druggable” genome to identify pathways of DUX4 toxicity. The study revealed the MYC-mediated apoptotic pathway and the viral dsRNA-mediated innate immune response pathway to be involved in DUX4 induced apoptosis⁴⁵. We examined our data for enrichment of gRNA sequences that target the genes identified in the Shadle study, but did not observe significant enrichment in our CRISPR screen data of these sequences. Figure 3A shows data plots that display the enrichment ($\text{Log}_2(\text{foldchange})$) and significance ($-\text{Log}_{10}(P\text{-value})$) of DUX4 and 3 other genes that were initially considered hits. However, subsequent single knock-outs validations demonstrated them to either have a generally effect on transcription (MED25), or upon their knockout did not exhibit any additional survival in induced DIE cells (RPS25 and CISD). Genes were only considered if a minimum of one gRNA showed significant enrichment in at least 3 out of 4 screens. Genes involved in the pathways identified by Shadle et al. did not meet these criteria (Fig. 3B). Furthermore, knocking out these genes in the DIE cells did not show an increased survival compared to background noise (Fig. 3C, top panel), as is noticeable in some of the false positives identified during this CRISPR screen (Fig. 3C, lower panel). It should be mentioned that the two screens have major technical differences, such as the screening method, the complete or partial loss of function of genes, the scale of the screens (druggable genome vs whole genome) and the different cellular backgrounds, which most likely all attributed to the little correlation seen between the two studies.

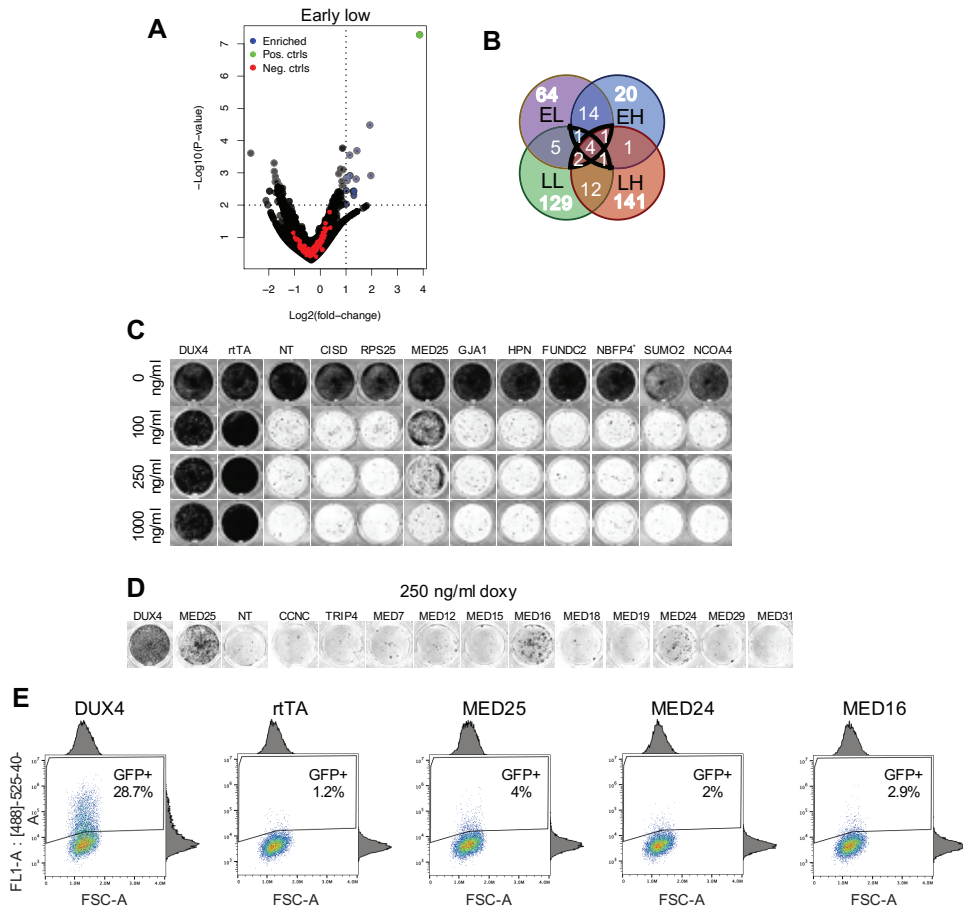


Figure 2. Filtered CRISPR screen data and validation of potential hits. (A) Adjusted volcano plot of low doxycycline/early harvest screen data showing the enrichment of sets of guides targeting genes not located on chromosome 5q or chromosome 19p. Blue points represent guide sets that are significantly enriched ($P\text{-value} \leq 0.01$), $\log_2(\text{foldchange}) \geq 1$, the green point is the positive control (DUX4), red points represent the Non-Target control guides. **(B)** Venn diagram showing the overlap of 5q-filtered hits between the four screens (EL: Early harvest/Low doxy, LL: Late harvest/Low doxy, EH: Early harvest/High doxy, LH: Late harvest/High doxy). **(C)** Viability staining showing surviving DIE cells containing single knockouts of potentials hits identified in the CRISPR screen, after exposure to 3 different concentrations of doxycycline. **(D)** Viability staining showing the surviving DIE-ieGFP-Cas9 cells containing single knockouts of mediator complex subunits, after exposure to 250ng/ml doxycycline. **(E)** FACS data showing GFP positive cells in surviving populations of DIE-ieGFP-Cas9 cells containing single knock-outs. DIE-ieGFP-Cas9 cells comprise of 42% of eGFP positive cells after DUX4 knock out. rtTA, MED25, MED24 and MED16 knock-outs show little eGFP expressing cells, comprising between 1.2-4% of eGFP expressing cells.

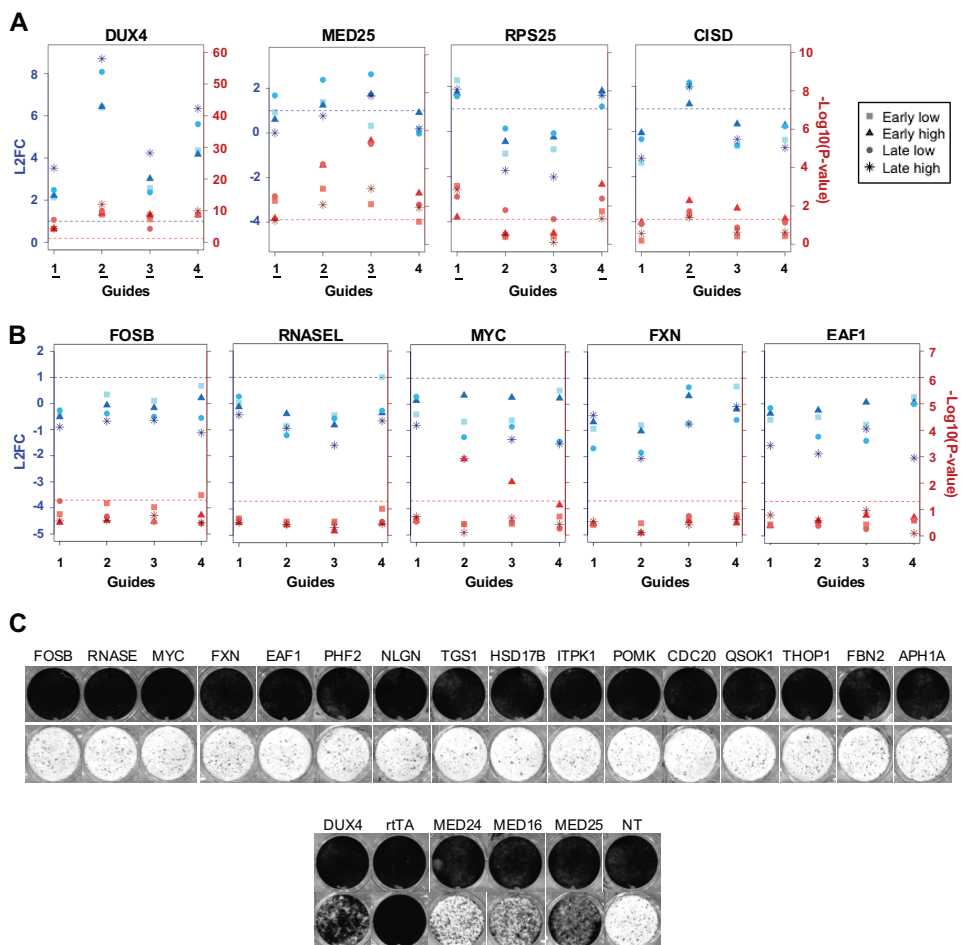


Figure 3. Validation of genes involved in the MYC-mediated apoptotic pathway and the viral dsRNA-mediated innate immune response pathway. (A) Data plots showing the significance and enrichment of gRNAs targeting DUX4, MED25, RPS25 and CISD, in all 4 screens. The Log₂(fold-change) (L2FC) of each individual guide is plotted on the left y-axis indicated in blue, and the $-\text{Log}_{10}(\text{P-value})$ is plotted on the right y-axis, in red. When guides fall above the blue and red intermittent ablines, they are considered significant ($\text{Log}_2(\text{foldchange}) > 1$, $-\text{Log}_{10}(\text{P-value}) > 1.3$). The gRNAs that are significantly enriched in all 4 screens are underlined. All 4 gRNAs targeting DUX4 are significantly enriched. 3 out of 4 gRNA's targeting MED25 are significantly enriched (guides 1, 2 and 3). Guides 1 and 4 targeting PRS25 are significantly enriched, and CISD has one guide that is significantly enriched in all 4 screens. **(B)** Data plots showing the enrichment of gRNA targeting FOSB, RNASEL, MYC, FXN and EAF1. None of the 4 guides show significant enrichment in any of the 4 screens. **(C)** Viability staining showing surviving DIE-Cas9 cells containing single knockouts of genes involved in the MYC-mediated apoptotic pathway and the dsRNA-mediated immune response pathway (Top panel). Controls can be found in the bottom panel and are as followed, positive controls: DUX4, rTA, MED24, MED16 and MED25; Negative control: NT

Discussion

At present there are no effective pharmacological treatment options that can improve muscle strength or slow down disease progression in FSHD patients⁴⁶. Unravelling the underlying mechanism of DUX4 cytotoxicity would help identify therapeutic targets. We hypothesized that inhibition of key downstream DUX4 effectors would slow or abrogate the cytotoxic process, and set out to identify such genes by performing an exhaustive genome wide CRISPR/Cas9 screen. We know the screen was exhaustive because we picked up rare rearrangements disabling the DUX4 transgene or the rtTA inducer. The goal of the screen was to identify targets that can mitigate DUX4 induced toxicity. While the screen's technical execution went very well and displayed high sensitivity, specificity toward candidate editing events that indeed mitigated cytotoxicity in our transgene model, none of the obtained hits had a direct effect on DUX4 its downstream transcriptional network. Rather, screen hits seemed to specifically affect the experimental system itself, either by affecting the tetracycline-inducible system responsible for DUX4 transgene induction, or by mutations of the DUX4 transgene. The main contributor is likely a rare Cas9-induced chromosomal truncation event, that removes the transgenes when targeted to the chromosomal arm to which they have integrated. Although these events appear to be rare, nearly all guides that targeted genes located on the chromosomal arm to which rtTA had integrated (5q) were robustly enriched, underwriting the sensitivity of this screening method. Most remaining hits did not appear to effect DIE cell survival upon individual validation, but members of the Mediator complex did show a positive effect on survival. Unfortunately, these mediator subunit genes seemed to generally suppress rtTA-mediated transcription so their mitigating effect was not mediated by specifically altering DUX4 cytotoxicity. We therefore concluded, that based on the conditions used in this study, there to be no individual target (other than DUX4 itself) that upon knockout can provide a strong inhibition of DUX4 induced cytotoxicity. Efforts should therefore be redirected to the direct modulation of DUX4.

While our library only targeted protein-coding genes, we believe we would have picked-up any mitigating non-coding RNAs as well, had they provided a strong rescue from the DUX4 cytotoxic effects. In that case, one would have expected to see a similar hotspot of gRNAs on and around the true target sites, as we observed for, MED16, where a hotspot of gRNAs was observed on the p-arm of chromosome 19, corresponding to the location of MED16. Another hotspot can be seen on the q arm of chromosome 19, corresponding to the location of MED25. The hotspot on chromosome 15 can be explained by the genetic makeup of the KBM7 cell line. KBM7 cells not only have the Philadelphia chromosome, but also an integration event where a region of chromosome 15 integrated on chromosome 19p⁴⁷. The hotspot on chromosome 15 correlates to the region that has integrated on chromosome 19p, close to the MED25 site.

Our screen results shown here do not corroborate previous findings of Shadle et al⁴⁵. However, their siRNA screen differs in many aspects to the performed genome wide CRISPR/Cas9 screen we executed. Their screen was knocking-down the druggable genome, using Lipofectamine RNAiMAX to deliver the siRNA library, in Rhabdomyosarcoma derived cells; whereas our screen was knocking-out protein coding genes genome wide, using a viral library, in chronic myeloid leukemia derived cells. These differences could explain why results between the two screens are not correlating with one another. Furthermore, A side by side comparison study of CRISPR/Cas9 and a next generation RNAi screens reveals that the screening methods seem to effect different biological aspect of the cells, therefore finding little correlations between results. The authors also in part attribute these differences to the

technical differences between the two techniques⁴⁸.

Another recently published genome wide CRISPR/Cas9 study, where a similar methodology was used in a DUX4 inducible immortalized myoblast line, identified the HIF1 oxidative stress pathway as a modulator of DUX4-induced apoptosis⁴⁹. This study, as well as previous reports, clearly demonstrate the role of the HIF1 hypoxia pathway in DUX4-mediated cytotoxicity⁴⁹⁻⁵¹. The HIF1 pathways did not come up in our screen (Fig. S7). This demonstrates that changes in this pathway are likely not the only DUX4 induced cellular changes that push cells towards apoptosis. The fact that the HIF1 pathway did not come up in our screen could also indicate differences in sensitivity to oxidative stress between cellular systems. Different cell types experience and respond differently to oxidative stress, with differences in culturing conditions as further attributing factors, for example the concentration of 2-mercapto ethanol to cell culture media.

While our screen did not identify target genes that can mitigate DUX4 cytotoxicity, it does illustrate some important aspects that need to be considered when performing phenotypic CRISPR/Cas9 screens. One being the large chromosomal truncations that can be induced by Cas9, a phenomena also recently reported by Cullot et al⁵². While these are rare events in a cell population, our results demonstrate that in a sufficiently sensitive screening system, they are robustly identified and can crowd potential positive hits. Sufficient selection should at least help in this aspect by removing cells that had their resistance marker (linked to the transgene) deleted. Another aspect that needs consideration are the endogenous genes that have a general effect on transcription and translation, in this case effecting the inducible system, like subunits of the mediator complex identified in this study. Potential hits will always need to be validated individually in such a way that can exclude this possibility, like shown here, or by Shadle et al. where some of the same genes were identified effecting their inducible Tet-On system⁴⁵.

This study started out with the aim of trying to contribute to the understanding of the underlying molecular mechanism of FSHD, by performing a genome wide CRISPR-Cas9 phenotypic screen. However, with no significant hits that can explain their contribution to the apoptotic phenotype, this story also tells a cautionary tale for knockout screens through the use CRISPR-Cas9, which will benefit future groups planning to execute similar screens.

Methods

Cell culture

HAP1 cells were cultured in IMDM media (Fischer Scientific) supplemented with 10% FBS. DIE cells were cultured in IMDM media supplemented with 10% Tet system approved FBS (Clontech), 5µg/ml Puromycin, 6µg/ml Blastidicin, and 100 µM Beta-mercaptoethanol.

Cloning p2T-Cas9, p2T-ieGFP and sgRNA constructs, and generating DIE-Cas9 and DIE-Cas9-ieGFP cell lines

The p2T-CAG-spCas9-NeoR mammalian expression plasmid was created by replacing the Blastidicin resistance gene (BlastR) in the p2T-CAG-spCas9-BlastR (Addgene: 107190)⁵³ with a Neomycin resistance gene (NeoR). The p2T-CAG-spCas9-BlastR plasmid is contained in a p2Tol2 backbone⁵⁴. The BlastR gene was removed using restriction digestion, using MfeI and

AflIII (NEB). Cloning the NeoR DNA fragment into the p2T-CAG-spCas9 backbone was done in similar fashion as described above. The p2T-CAG-SpCas9-BlastR was a gift from Richard Sherwood. The p2T-TetO-eGFP-HygroR plasmid was generated in a similar way as the p2T-CAG-spCas9-NeoR. In short, all sequences between transposable elements of a p2T plasmid were removed by restriction digestion using AclI and EcoRI (NEB). The TetO-eGFP-HygroR cassette was created by amplifying each subunit individually, and thereafter cloned into the empty p2T backbone, using in-fusion cloning.

Both p2T-CAG-spCas9-NeoR and p2T-TetO-eGFP-HygroR were introduced in the DIE cell line by using Transposase. The p2T-CAG-spCas9-NeoR was introduced into DIE cells together with a plasmid encoding for transposase, using Polyethylenimine (PEI) transfection reagent (4ug PEI per 1ug DNA). The DIE cells were exposed to the transfection mixture for 14-16h, after which the transfection media was replaced with growth media. Geneticin g418 selection was started two days post transfection, generating the DIE-Cas9 line. The DIE-Cas9-ieGFP cell line was created by adding Transposase and p2T-TetO-eGFP-HygroR the DIE-Cas9 line, described as above.

spCas9-sgRNA constructs were cloned using a plasmid containing a U6 promoter, 2 BsmBI sites with directly adjacent the tracrRNA sequence, and a Hygromycin resistance gene (made in house). This U6-2xBsmBI-Tracr-HygroR plasmid was digested with the BsmBI restriction enzymes (NEB), after which the CRISPR inserts were ligated in using T4 DNA ligase (NEB). CRISPR inserts were generated by annealing two complementary oligos containing a 4bp adapter serving as the BsmBI sticky end.

All plasmids mentioned in this study were transformed in chemically competent Stbl3 Escherichia coli (E.coli), and prepped using a HiPure plasmid Midi or Maxi kit (Invitrogen).

Doxycycline titration curve

200.000 cells were seeded into wells of a 24-wells plate and incubated overnight at 5% CO₂ and 37°C. When cells reached a density of 90-100% confluency, different concentrations of doxycycline were added to the vertical lanes (100 ng/ml, 250ng/ml, 500ng/ml, 750ng/ml, 1000 ng/ml), with the horizontal lanes experiencing different exposure times (48h, 36h, 24h, 12h). After a recovery period of 96h (after doxy exposure was ended), cells were washed with DPBS, and fixed with Methanol for 10 minutes. Giemsa stain, modified solution (Sigma) was subsequently added for 45 minutes, after which it was removed and the wells were washed with demineralized water.

Genome-wide CRISPR screen

The screen on the DIE line was performed as previously described by Doench et al.⁴³. Due to a shared selection marker between the DIE line and the Brunello lentiviral library, transfected cells could not be selected for, thus the total number of cells was raised to 1500 cells per guide, when considering an average transfection efficiency of 30-50% in all cell lines tested by Doench et al.⁴³. The transfection efficiency was determined and calculated using the DIE parental line, the rediplodized HAP1 cells. With 1500 cells per guide (total of 77.441guides), each of the technical three replicates contained 120*10E6 cells, that were spin transfected for 2h at 1000g with 82*10E6 Brunello virus particles (LentiCRISPRv2, Addgene 73179-LV). With a multiplicity of infection (MOI) of 0.65, transfection efficiency reached 60% upon testing the viral library on the diploid HAP1 parental line. After transfection the 120*10E6 transfected cells (contained in 40 wells of 12-well tissue culturing plates) were trypsinized and passaged to 60 145mm TC plates. Mutagenized cells were maintained for 6 days,

before inducing a set of 24 plates with either a low or high doxycycline concentration (low: 250ng/ml, high: 1000ng/ml). The remaining 12 plates were harvested for cryofreezing (7 plates) and for determining library coverage (5 plates). After a 24h doxycycline induction period, 12 plates were given a 24h recovery period (early harvest) of both the low and high doxycycline exposed sets. The remaining 24 plates received an additional 48h of recovery time (late harvest), before harvesting the surviving cells for sequencing (Fig. S1). Cell Pellets were stored at -80°C until further processing. The Human Brunello CRISPR knockout pooled lentiviral prep library was a gift from David Root and John Doench.

Library prep, sequencing and analysis

Genomic DNA (gDNA) was isolated using NucleoSpin Blood Mini (less than 5 million cells), Midi (L) (5-20 million cells) and Maxi (XL) (more than 20 million cells) kits, depending on the size of cell pellet. Libraries were prepared and sequenced on a HiSeq2000 (Illumina) as described by Doench et al. Analysis was conducted using “STARS”, gene-ranking method to generate FDR values developed by Doench et al. that was used to generate p-values and FDR rates.⁴³ Chromosomal ideogram were generated by using the PhenoGram webtool from the Ritchie lab from the university of Pennsylvania⁵⁷.

Individual knock outs in DIE cells

DIE-Cas9 and DIE-Cas9-ieGFP were seeded in a 24-well setting. Next day, when the cells had reached 70-90% confluency, cells were transfected with 500ng guide plasmid per well using 4ug PEI per 1ug DNA. During the overnight transfection no selection markers were presents in the media, however growth media was supplemented with 100U/ml pen-strep. Cells were passaged with or without selection markers during a period of 6-7, after which doxycycline was added (100, 250 or 1000 ng/ml) for a 24h period. Wells were washed with DPBS to remove dead cells and debris. Remaining cells were given the opportunity to grow out, or to perish (if they had already entered the apoptotic pathway) for an additional 48-96 hours. The wells were stained using Giemsa modified solution, as described previously.

Flowcytometry sorting (FACS) and analysis

DIE-Cas9-ieGFP cells were induced with 250ng/ml doxycycline 24h prior to FACS analysis. After the 24h doxycycline exposure, cells were trypsinized using 0.25% Trypsin-EDTA, resuspended in iMDM media supplemented with Tet approved FBS and DAPI nuclear staining, and strained using a Cell-strainer capped tubes (Falcon). Cells were analyzed using the Beckman coulter Cytomflex S flow cytometer.

Data Resources

Data containing the Genome wide CRISPR/Cas9 samples in triplicate are available from the GEO data base, accession number: GSE155034.

Acknowledgements

This study was supported by Stichting FSHD nd the SingelSwim Utrecht. The authors like to thank Nune Schelling, Peng Shang, Stefan van der Elst, en Reinier van der Linden for excellent technical assistance.

References

1. Deenen, J. C. W. *et al.* Population-based incidence and prevalence of facioscapulohumeral dystrophy. *Neurology* **83**, 1056–1059 (2014).
2. Tawil, R. & Van Der Maarel, S. M. Facioscapulohumeral muscular dystrophy. *Muscle and Nerve* **34**, 1–15 (2006).
3. Pastorello, E., Cao, M. & Trevisan, C. P. Atypical onset in a series of 122 cases with FacioScapuloHumeral Muscular Dystrophy. *Clin. Neurol. Neurosurg.* **114**, 230–234 (2012).
4. Dixit, M. *et al.* DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18157–18162 (2007).
5. Lemmers, R. J. L. F. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* **329**, 1650–1653 (2010).
6. Snider, L. *et al.* Facioscapulohumeral dystrophy: Incomplete suppression of a retrotransposed gene. *PLoS Genet.* **6**, 1–14 (2010).
7. Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* **49**, 935–940 (2017).
8. Hendrickson, P. G. *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934 (2017).
9. Das, S. & Chadwick, B. P. Influence of repressive histone and DNA methylation upon D4Z4 transcription in non-myogenic cells. *PLoS One* **11**, 1–26 (2016).
10. Geng, L. N. *et al.* DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy. *Dev. Cell* **22**, 38–51 (2012).
11. Dmitriev, P. *et al.* DUX4-induced constitutive DNA damage and oxidative stress contribute to aberrant differentiation of myoblasts from FSHD patients. *Free Radic. Biol. Med.* **99**, 244–258 (2016).
12. Iaco, A. De *et al.* A family of double-homeodomain transcription factors regulates zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945 (2017).
13. Knopp, P. *et al.* DUX4 induces a transcriptome more characteristic of a less-differentiated cell state and inhibits myogenesis. *J. Cell Sci.* **129**, 3816–3831 (2016).
14. Bosnakovski, D. *et al.* High-throughput screening identifies inhibitors of DUX4-induced myoblast toxicity. *Skelet. Muscle* **4**, 1–11 (2014).
15. Bosnakovski, D. *et al.* DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. *Exp. Neurol.* **214**, 87–96 (2008).
16. Winokur, S. T. *et al.* Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum. Mol. Genet.* **12**, 2895–2907 (2003).
17. Bosnakovski, D. *et al.* An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J.* **27**, 2766–2779 (2008).
18. Bosnakovski, D. *et al.* p53-independent DUX4 pathology in cell and animal models of facioscapulohumeral muscular dystrophy. *DMM Dis. Model. Mech.* **10**, 1211–1216 (2017).
19. Feng, Q. *et al.* A feedback loop between nonsense-mediated decay and the retrogene DUX4 in facioscapulohumeral muscular dystrophy. *Elife* **2015**, 1–13 (2015).
20. Rickard, A. M., Petek, L. M. & Miller, D. G. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum. Mol. Genet.* **24**, 5901–5914 (2015).
21. Jagannathan, S. *et al.* Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum. Mol. Genet.* **25**, ddw271 (2016).
22. Van Den Heuvel, A. *et al.* Single-cell RNA sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. *Hum. Mol. Genet.* **28**, 1064–1075 (2019).
23. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
24. Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
25. Christian, M. *et al.* Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 756–761 (2010).
26. Li, T. *et al.* TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **39**, 359–372 (2011).
27. Mussolino, C. *et al.* A novel TALE nuclease scaffold enables high genome editing activity in

- combination with low toxicity. *Nucleic Acids Res.* **39**, 9283–9293 (2011).
28. Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143–150 (2011).
 29. Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1156–1160 (1996).
 30. Kim, Y. G., Kim, P. S., Herbert, A. & Rich, A. Construction of a Z-DNA-specific restriction endonuclease. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12875–12879 (1997).
 31. Smith, J. *et al.* Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res.* **28**, 3361–3369 (2000).
 32. Bibikova, M. *et al.* Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Mol. Cell. Biol.* **21**, 289–297 (2001).
 33. Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted Chromosomal Cleavage and Mutagenesis in *Drosophila* Using Zinc-Finger Nucleases. *Genetics* **161**, 1169–1175 (2002).
 34. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2579–2586 (2012).
 35. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2013**, 1–9 (2013).
 36. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).
 37. Choi, S. H. *et al.* DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **44**, 5161–5173 (2016).
 38. Vuoristo, S. *et al.* DUX4 regulates oocyte to embryo transition in human. *Biorxiv* (2019) doi:http://dx.doi.org/10.1101/732289.
 39. Jones, T. I., Himeda, C. L., Perez, D. P. & Jones, P. L. Large family cohorts of lymphoblastoid cells provide a new cellular model for investigating facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* **27**, 221–238 (2017).
 40. Banerji, C. R. S., Zammit, P. S. & Panamaraova, M. Lymphocytes contribute to DUX4 target genes in 1 FSHD muscle biopsies. *Biorxiv* (2019) doi:http://dx.doi.org/10.1101/717652.
 41. Andersson, B. S. *et al.* Ph-positive chronic myeloid leukemia with near-haploid conversion in vivo and establishment of a continuously growing cell line with similar cytogenetic pattern. *Cancer Genet. Cytogenet.* **24**, 335–343 (1987).
 42. Mulherkar, N. *et al.* Ebola Virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340–343 (2012).
 43. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).
 44. Soutourina, J. Transcription regulation by the Mediator complex. *Nat. Rev. Mol. Cell Biol.* **19**, 262–274 (2018).
 45. Shadle, S. C. *et al.* DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. *PLoS Genet.* **13**, 1–25 (2017).
 46. Tawil, R. *et al.* Evidence-based guideline summary: Evaluation, diagnosis, and management of facioscapulohumeral muscular dystrophy. *Neurology* **85**, 357–364 (2015).
 47. Essletzbichler, P. *et al.* Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line. *Genome Res.* **24**, 2059–2065 (2014).
 48. Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34**, 634–636 (2016).
 49. Lek, A. *et al.* Applying genome-wide CRISPR-Cas9 screens for therapeutic discovery in facioscapulohumeral muscular dystrophy. *Sci. Transl. Med.* **12**, 9–11 (2020).
 50. Banerji, C. R. S. *et al.* β -catenin is central to DUX4-driven network rewiring in facioscapulohumeral muscular dystrophy. *J. R. Soc. Interface* **12**, (2015).
 51. Tsumagari, K. *et al.* Gene expression during normal and FSHD myogenesis. *BMC Med. Genomics* **4**, (2011).
 52. Cullot, G. *et al.* CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. *Nat. Commun.* **10**, 1–14 (2019).
 53. Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).
 54. Urasaki, A., Morvan, G. & Kawakami, K. Functional dissection of the Tol2 transposable

- element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* **174**, 639–649 (2006).
55. Ferreboeuf, M. *et al.* DUX4 and DUX4 downstream target genes are expressed in fetal FSHD muscles. *Hum. Mol. Genet.* **23**, 171–181 (2014).
 56. Borgström, A. *et al.* A novel multiplex qPCR targeting 23S rDNA for diagnosis of swine dysentery and porcine intestinal spirochaetosis. *BMC Vet. Res.* **13**, 1–8 (2017).
 57. Wolfe, D., Dudek, S., Ritchie, M. D. & Pendergrass, S. A. Visualizing genomic information across chromosomes with PhenoGram. *BioData Min.* **6**, 1–12 (2013).

Supplementary figures

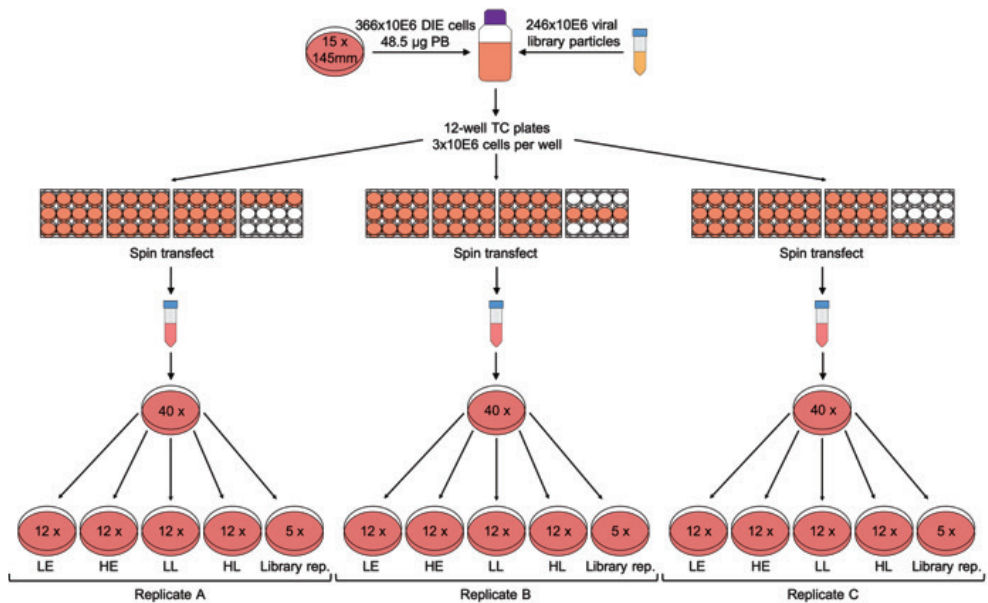


Figure S1. Execution of the CRISPR/Cas9 genome wide screens. A schematic representation of the execution of the CRISPR/Cas9 screens. PB: Polybrene, TC: Tissue culture, LE: Low doxycycline/Early harvest, HE: High doxycycline/Early harvest, LL: Low doxycycline/Late harvest, HL: High doxycycline/Late harvest, Library rep: Library representation.

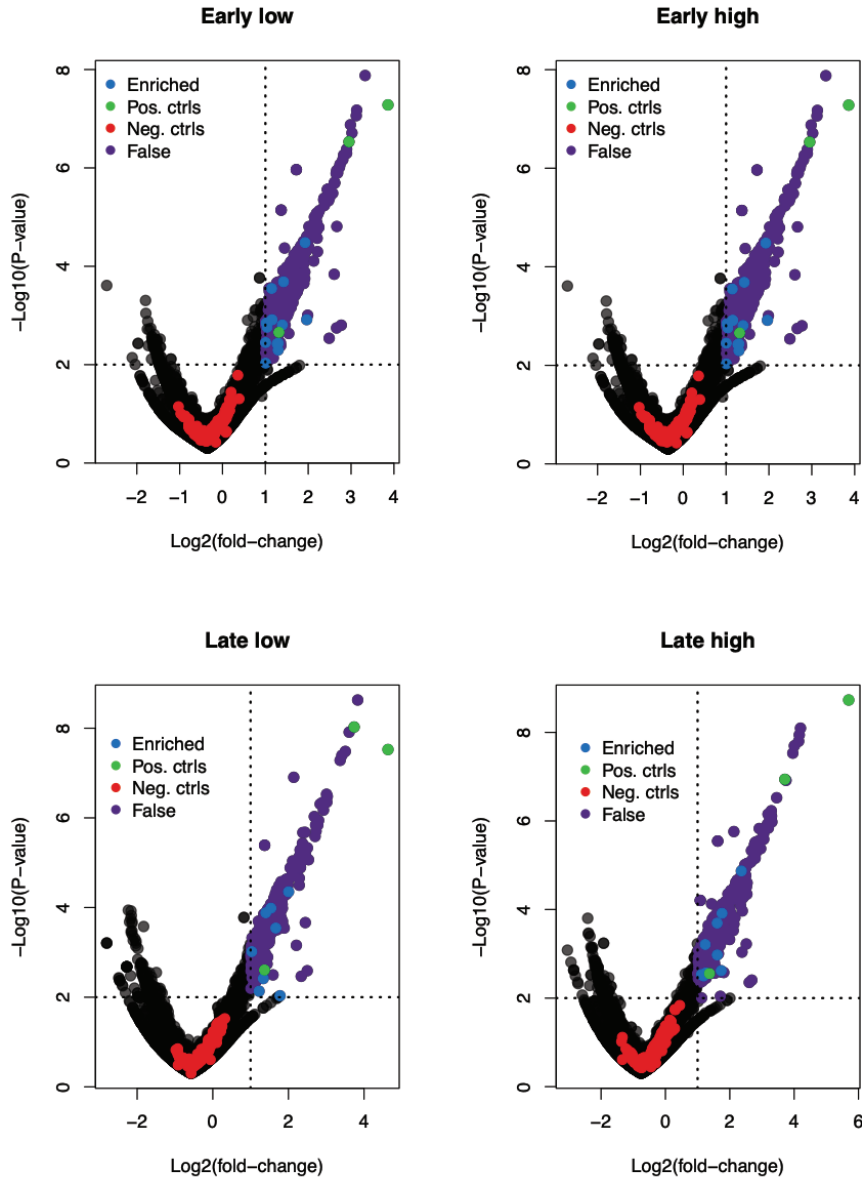
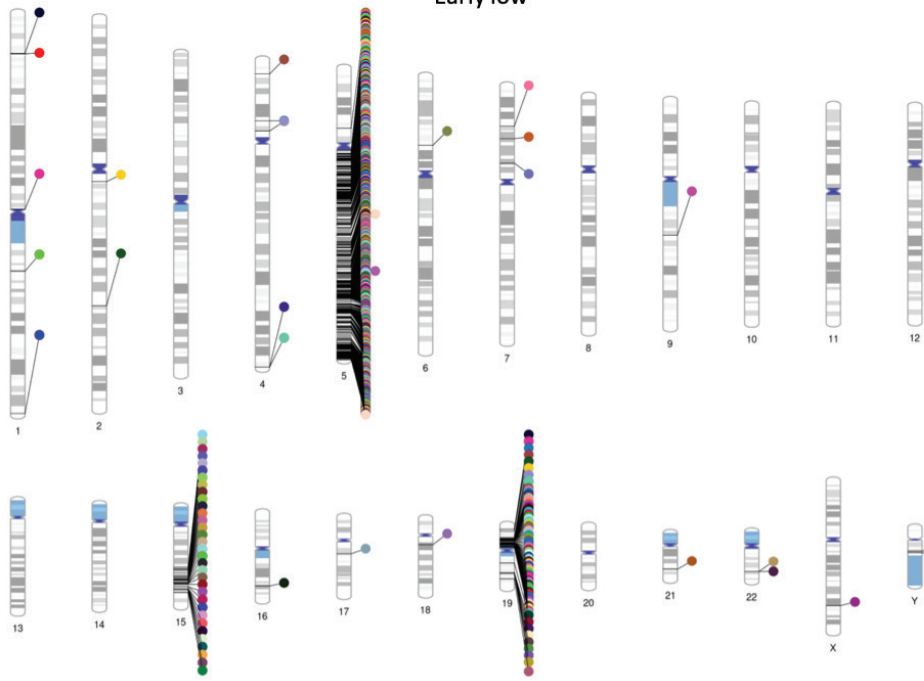
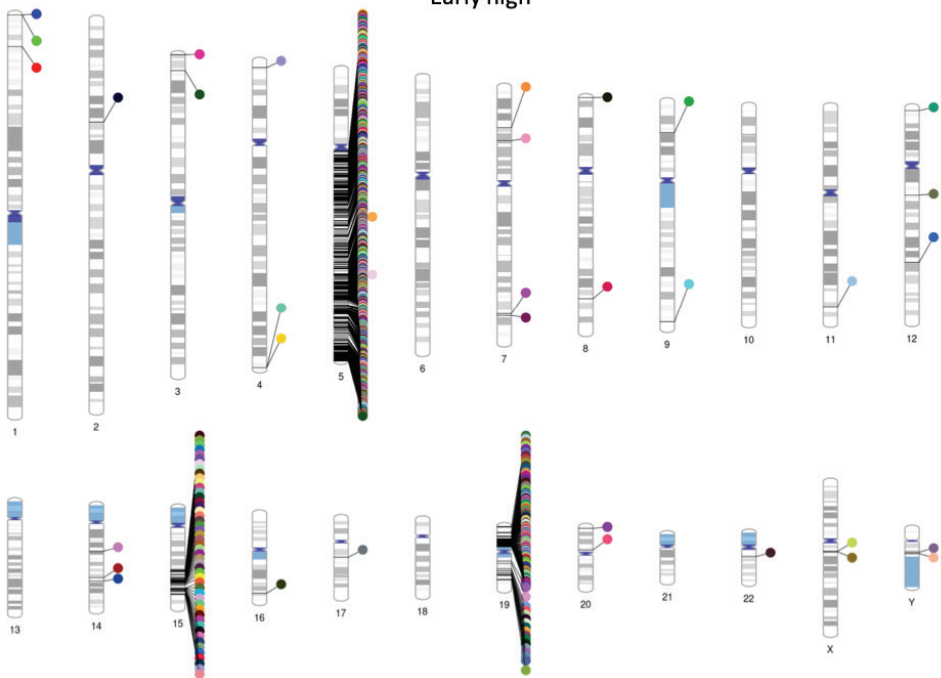


Figure S2. gRNAs enrichment from screen data processed with a one-sided analysis. Volcano plots illustrating enrichment of gRNAs in the surviving population of DIE cells of all 4 screens. Due to the one-sided analysis, depletion data should not be taken into consideration. For a two-sided analysis see supplementary figure S3. The $\text{Log}_2(\text{foldchange})$ is plotted on the x-axis and the $-\text{Log}_{10}(\text{p-value})$, is plotted on the y-axis. Data shown here shows the average $\text{log}_2(\text{foldchange})$ and $-\text{log}_{10}(\text{P-value})$ of each guide set (set: 4 guides per gene). Blue points represent guide sets that are significantly enriched ($\text{Log}_2(\text{foldchange}) \geq 1$, $-\text{log}_{10}(\text{P-value}) \geq 2$), purple points represent the false positive hits on chromosome 5q and chromosome 19p, green point are the positive controls (DUX4, MAST1, MGAT4B), red points represent the Non-Target control guides.

Early low



Early high



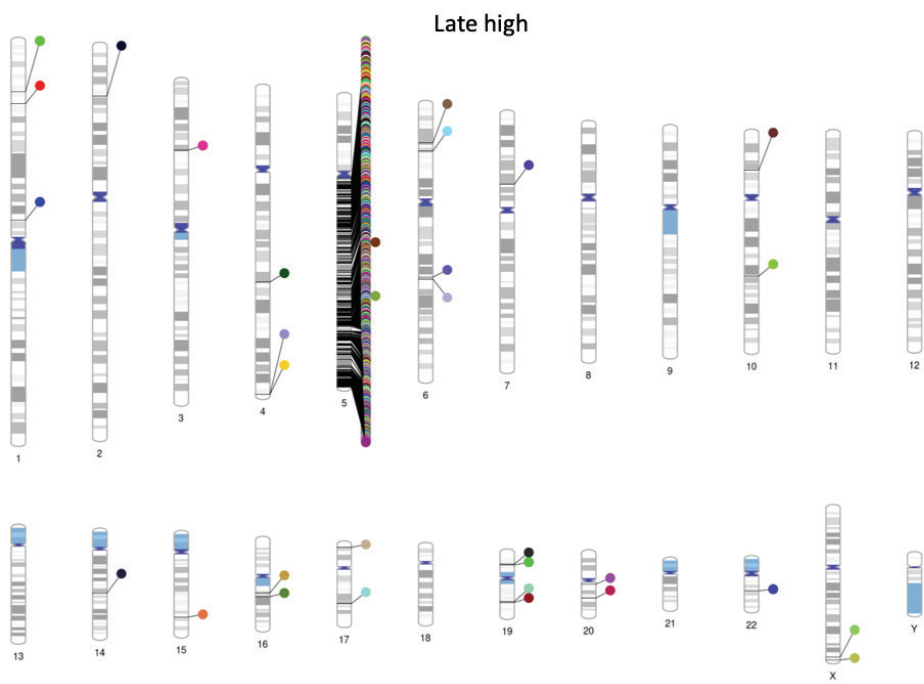
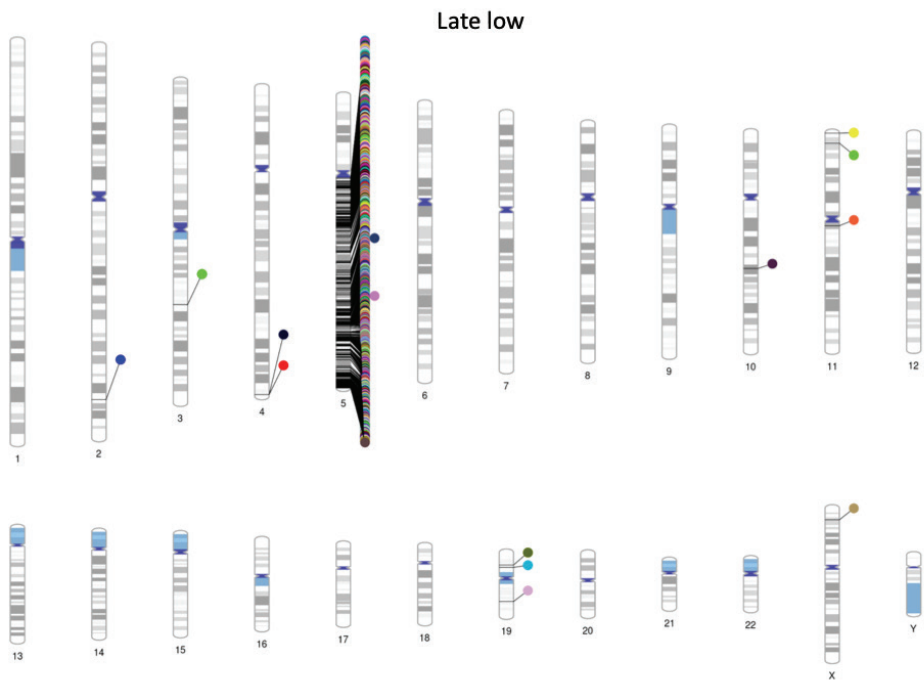


Figure S3. PhenoGrams showing enriched hits in the human genome. Chromosomal ideogram indicating the location of enriched hits in the human genome, for each of the 4 screens. PhenoGram is a software created by the Ritchie lab from the university of Pennsylvania⁵⁷.

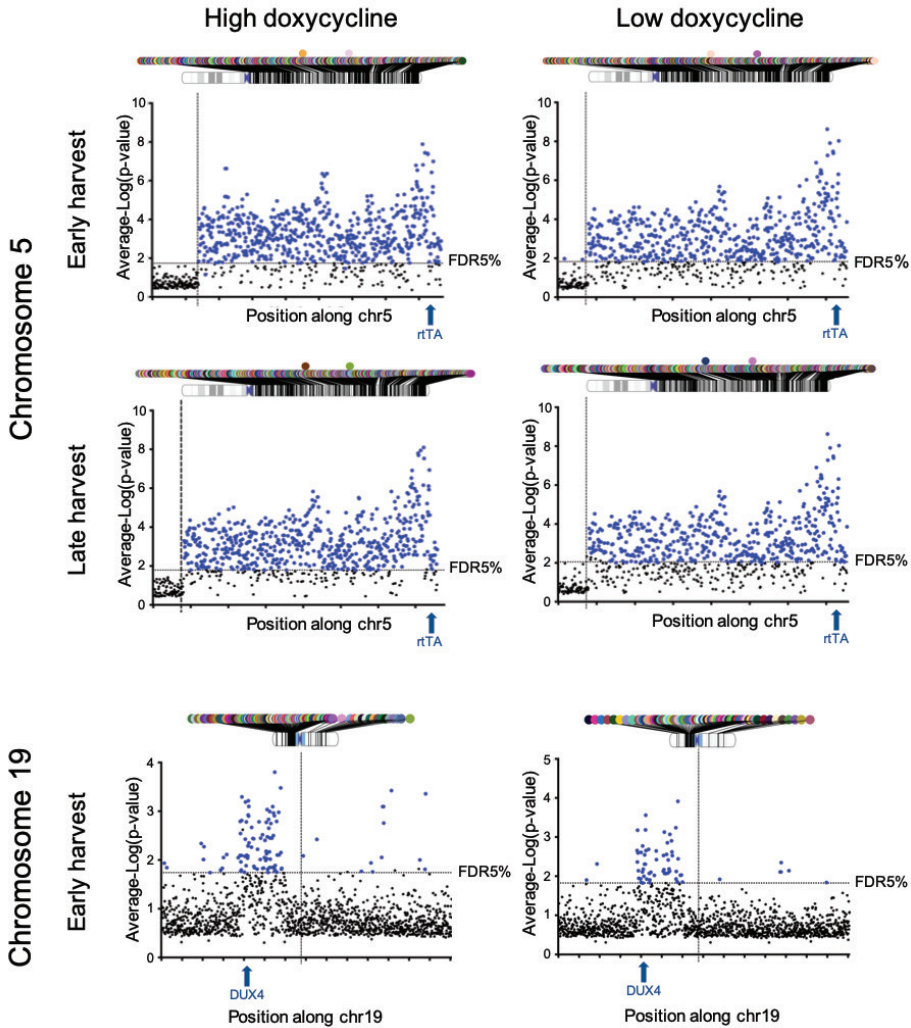


Figure S4. Data plot displaying enriched hits on chromosome 5q and chromosome 19. The average $-\log(p\text{-value})$ is plotted on the y-axis, and the x-axis is displaying the position on the chromosome. The vertical abline indicates the position of the centromere. All points above the horizontal abline (in blue) indicate significantly enriched hits that fall below the 5% False Discovery Rate (FDR) threshold. The location of the transgene is annotated with a blue arrow on the x-axis.

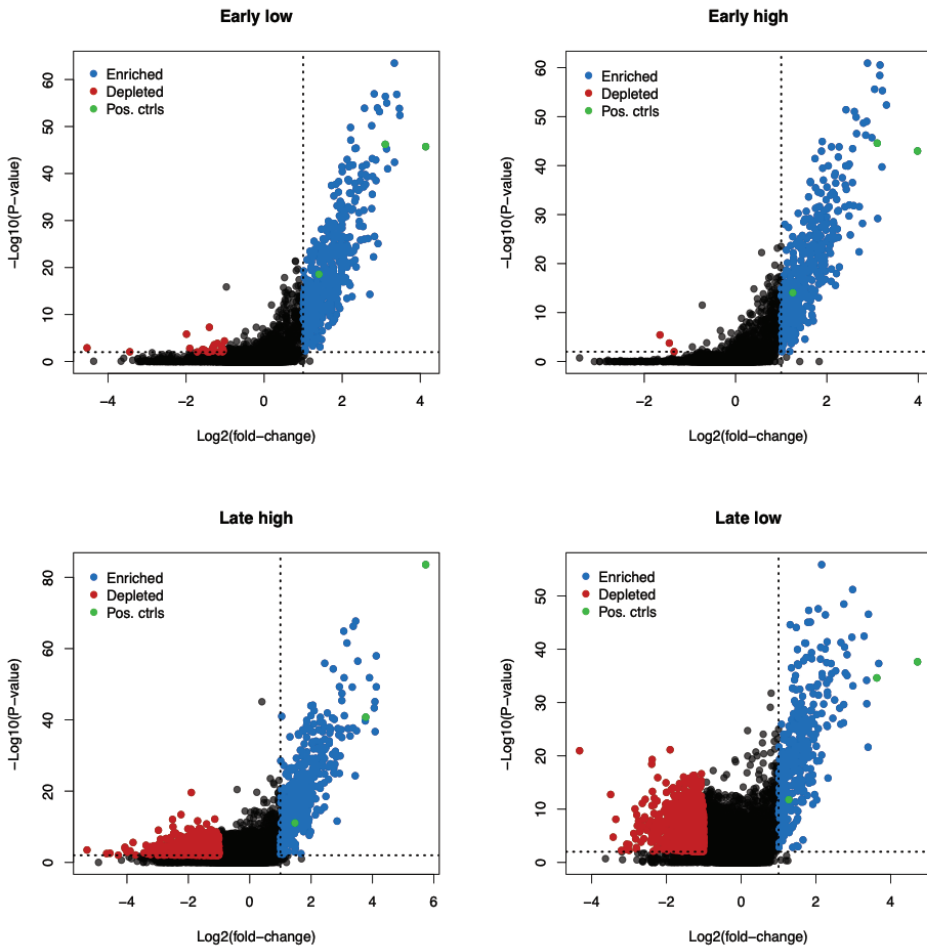


Figure S5. Analysis of enrichment and depletion of gRNAs from screen data analyzed with a two-sided analysis. Volcano plots illustrating enrichment and depletion of gRNAs in all 4 screens. The Log2(foldchange) is plotted on the x-axis and the -Log10(P-value) is plotted on the y-axis. Blue points represent significantly enriched gRNA sequences (Log2(foldchange) ≥ 1, -log10(P-value) ≥ 2). Green dots represent positive controls (DUX4, MAST1, MGAT4B), and red dots represent the depleted targets.

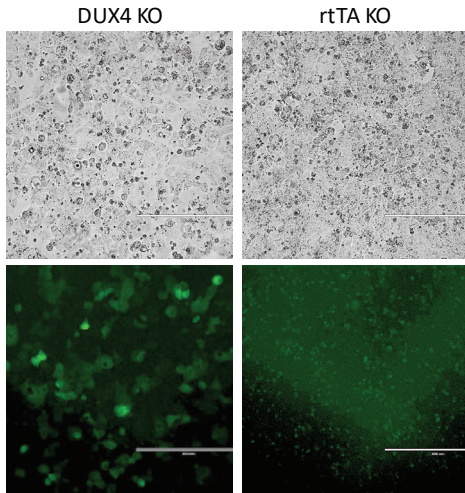


Figure S6. Individual knock-outs in DIE-ieGFP-Cas9 cells demonstrating eGFP activation in cells with a functional TetO inducible system. Phase contrast (top panel) and fluorescent images (bottom panel) of DIE cells containing a DUX4 KO (left panel), and rtTA KO (right panel) induced with 250 ng/ml doxycycline. DIE cells had reached over 100% confluency.

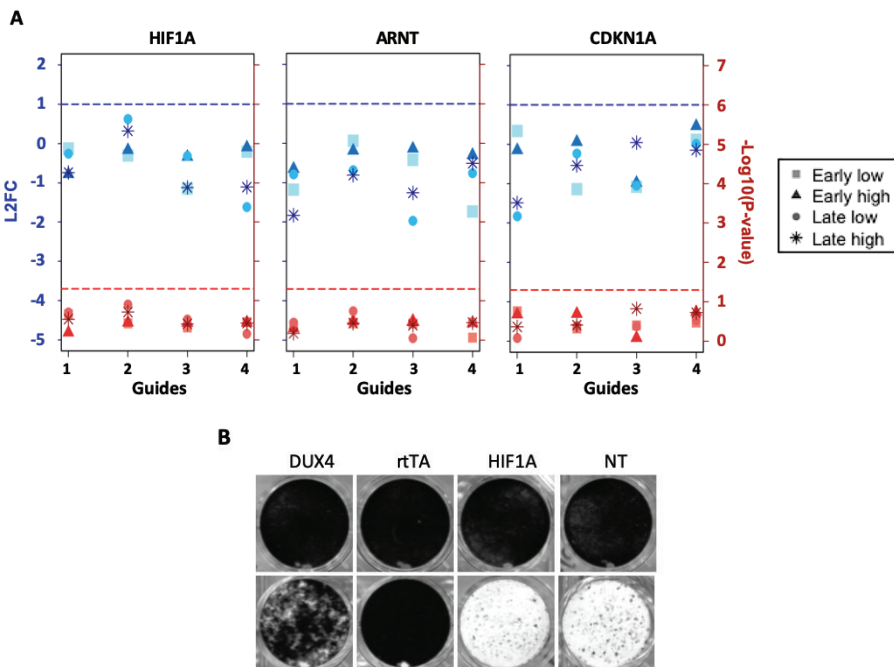


Figure S7. Validation of genes involved in the HIF1 hypoxia pathway (A) Data plots showing the enrichment of gRNA's targeting HIF1A, HIF1B/ARNT, and CDKN1A. The Log₂(foldchange) value of each individual guide is plotted on the left y-axis, indicated in blue, and the -Log₁₀(P-value) is plotted on the right y-axis, in red. Guides located above the blue and red intermittent ablines (blue: Log₂(foldchange) > 1, red: -Log₁₀(P-value) > 1.3) are considered to be significantly enriched. **(B)** Viability staining of DIE-Cas9 uninduced cells (top panel) and DIE-Cas9 100ng/ml doxycycline induced cells (lower panel), transfected with DUX4, rtTA, HIF1A and non-targeting (NT) sgRNA coding plasmids.

Table S1: Enriched sgRNA's and their corresponding genes. Screen: Low doxycycline-early harvest

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
PAFAH1B3	chr19q	0.32	1.30	-0.37	0.15	1.16	1.59	1.13	1.15
RINL	chr19q	-0.38	-0.75	1.56	0.71	0.62	0.57	1.93	1.28
LGALS7B	chr19q	-0.10	-0.33	1.26	0.99	0.69	0.59	1.88	0.93
MED25	chr19q	0.93	1.38	0.32	0.18	2.30	2.94	2.12	1.20
CD177	chr19q	0.55	-0.11	0.35	1.00	1.85	1.01	1.59	2.21
CDH10	chr5p	-0.69	1.34	0.86	-1.25	0.40	2.19	1.18	0.26
USH2A	chr1	1.01	-1.44	0.23	1.05	1.84	0.17	1.04	2.40
TMEM57	chr1	0.04	1.41	0.14	1.02	1.21	2.42	1.57	2.37
NBPF4/6	chr1	1.66	0.27	0.12	0.68	2.55	1.82	1.28	2.29
S100A7L2	chr1	1.01	0.44	0.37	0.03	2.31	2.28	1.35	1.01
UBE2D1	chr10	-0.07	0.14	1.80	0.70	1.40	1.57	1.97	1.89
TBATA	chr10	-1.22	1.20	-1.01	0.85	0.28	2.18	0.38	1.12
NCOA4	chr10	0.30	1.64	-0.41	-0.05	1.09	1.31	1.08	1.08
CISD1	chr10	-1.38	2.01	-0.63	-0.37	0.19	1.53	0.42	0.43
METTL10	chr10	1.32	-0.42	1.02	-0.98	2.41	0.54	1.17	0.42
OR5M9	chr11	-1.83	-0.27	0.77	1.59	0.05	0.70	1.29	2.04
RPS25	chr11	2.27	-0.98	-0.79	1.66	3.04	0.38	0.42	1.72
BUD13	chr11	-0.31	1.11	0.54	0.59	1.07	2.85	1.28	1.66
OR5B3	chr11	0.47	-0.16	1.55	-0.47	1.27	0.99	1.42	0.87
RAB21	chr12	1.24	-0.79	-1.93	1.82	1.41	0.38	0.03	2.66
GTSF1	chr12	-0.91	1.06	1.10	0.56	0.48	2.46	2.92	1.07
STYK1	chr12	0.28	1.51	-1.51	0.06	1.26	1.35	0.14	1.04
ZNF10	chr12	1.09	0.96	-0.48	-0.40	2.34	1.06	0.56	0.97
NFE2	chr12	-1.25	1.80	0.18	0.10	0.26	1.45	1.40	0.86
NAA25	chr12	1.66	-0.84	-0.30	-1.61	1.32	0.38	0.43	0.10
VCPKMT	chr14	0.24	1.10	0.15	0.31	1.09	2.79	1.07	1.85
GOLGA6L4	chr15	1.13	1.08	1.62	NA	2.16	1.56	6.83	NA
GOLGA6L10	chr15	-0.53	1.19	1.26	1.42	2.16	3.54	5.10	5.89
CHRNA5	chr15	0.37	0.39	0.29	1.03	1.25	2.28	1.03	3.42
GOLGA6C	chr15	0.15	0.89	0.71	1.27	1.15	2.79	2.23	3.39
GOLGA6L9	chr15	-0.76	0.82	-0.25	1.27	0.76	1.06	0.98	3.34
ST20-MTHFS	chr15	0.68	1.03	0.19	0.75	2.23	2.63	1.96	2.51
ANKDD1A	chr15	0.41	0.41	-1.35	1.09	1.06	1.30	0.21	2.41



Table S1 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
VWA9	chr15	0.70	0.05	1.21	0.75	1.98	1.12	3.38	2.38
CYP1A2	chr15	0.92	1.06	0.15	0.64	2.80	3.16	1.04	2.27
CCDC33	chr15	-0.23	0.89	-0.09	1.29	1.00	1.48	1.16	2.24
CYP1A1	chr15	-0.03	1.43	0.30	0.51	1.22	2.11	1.50	2.06
SMAD3	chr15	0.27	0.06	1.10	0.48	1.45	1.07	2.44	1.95
EMC7	chr15	1.16	0.35	-0.59	0.72	2.23	1.10	0.80	1.95
EFTUD1	chr15	-0.71	0.11	-0.46	1.99	0.51	0.76	0.66	1.51
RASL12	chr15	0.56	1.31	-0.07	0.01	1.60	1.97	1.17	1.22
UBE2Q2L	chr15	1.12	1.35	1.20	0.91	2.62	6.04	4.36	1.18
ADAMTS7	chr15	0.52	0.79	1.07	0.36	2.06	2.79	3.72	1.05
GOLGA8R	chr15	-0.30	2.25	-0.39	-0.56	0.87	1.41	0.84	0.70
ODF3L1	chr15	-0.02	0.92	1.40	-1.23	1.16	1.21	2.27	0.27
RPL4	chr15	0.24	1.05	1.30	-1.43	1.16	1.87	2.45	0.17
HSD3B7	chr16	0.11	-0.24	0.67	1.89	1.46	1.45	1.49	1.83
PDZD9	chr16	-0.87	1.69	-0.21	-1.20	0.38	1.33	0.45	0.28
CCL8	chr17	1.62	-0.95	-0.66	0.07	1.30	0.41	0.44	0.70
SUMO2	chr17	-0.51	-0.03	1.66	-1.28	0.48	0.58	1.32	0.24
MALT1	chr18	0.25	1.32	0.43	0.81	1.07	1.91	1.61	1.67
SEPT10	chr2	1.19	-0.93	0.41	0.93	2.41	0.46	1.11	2.29
CGREF1	chr2	-0.21	0.57	-0.05	1.50	1.07	1.53	1.25	1.63
KANSL1L	chr2	0.35	0.12	-0.45	1.67	1.32	1.18	1.01	1.49
AMER3	chr2	-0.28	1.78	-0.76	0.48	0.69	1.45	0.61	1.39
SPAG16	chr2	0.10	-0.52	1.62	-0.08	1.02	0.74	1.31	0.90
C2orf80	chr2	-1.92	1.68	0.33	-1.29	0.03	1.33	1.14	0.47
MGME1	chr20	1.90	0.24	-1.62	0.76	2.02	1.46	0.10	1.86
GNAS	chr20	0.07	1.05	0.11	0.21	0.92	2.46	1.03	1.47
NSFL1C	chr20	0.21	-1.10	0.07	1.11	1.07	0.35	0.92	1.37
TTC3	chr21	0.35	0.41	0.33	1.03	1.30	2.22	1.03	3.59
BAGE2	chr21	0.34	1.13	0.07	0.23	1.82	2.48	1.08	1.15
DOPEY2	chr21	1.69	-0.59	-0.46	0.00	1.33	0.51	0.61	0.80
SNRPD3	chr22	-1.13	0.29	0.85	1.28	0.33	1.16	2.03	2.18
SREBF2	chr22	-1.08	-0.02	-1.66	1.83	0.40	0.59	0.08	1.42
L3MBTL2	chr22	-0.26	1.80	0.30	-1.27	0.72	1.40	1.09	0.25
C22orf46	chr22	0.18	0.29	1.25	-1.35	1.05	1.14	1.69	0.21
OR5K1	chr3	0.18	1.53	0.38	0.51	1.18	2.33	1.31	1.97

Table S1 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
MRAS	chr3	-1.21	-1.12	1.76	-1.57	0.44	0.90	1.37	0.11
RPSA	chr3	0.48	1.56	NA	-1.86	1.40	1.70	NA	0.03
PTTG2	chr4	1.10	0.41	0.51	1.40	2.75	1.21	2.51	3.96
PRSS48	chr4	-0.73	0.59	0.60	1.15	0.64	1.09	1.70	3.01
PDE6B	chr4	0.30	-1.74	-1.05	2.15	1.07	0.06	0.39	1.62
MRPL2	chr6	0.06	0.73	0.54	1.03	1.02	2.44	1.95	2.85
RPF2	chr6	0.19	0.82	-0.78	1.03	1.02	1.70	0.59	2.13
NR2E1	chr6	-1.19	-1.27	1.17	0.47	0.43	0.25	1.43	1.10
IGF2R	chr6	1.46	1.25	-1.48	-0.88	2.68	1.24	0.15	0.38
HECA	chr6	-0.97	0.29	1.68	-1.50	0.38	1.06	1.33	0.14
GJA1	chr6	-0.04	1.73	0.30	-1.93	1.08	1.36	1.11	0.03
CYCS	chr7	1.81	1.41	0.43	0.51	4.05	2.82	1.41	2.76
NUB1	chr7	0.11	-1.30	-0.20	1.79	0.80	0.23	0.76	1.39
ZNF/626/680	chr7	-0.13	1.56	-0.43	0.97	0.61	1.98	0.56	1.11
PDAP1	chr7	0.44	-1.38	1.37	-0.48	1.20	0.19	1.35	0.49
ISCA1	chr9	1.72	0.87	-0.35	-0.24	2.20	1.35	0.73	1.21
DMRT2	chr9	0.65	-1.18	1.35	0.29	1.79	0.30	2.01	1.19
GPR21	chr9	-0.83	1.16	1.01	0.15	0.55	2.41	1.60	1.09
UXT	chrX	0.85	-0.66	-0.13	1.23	1.13	0.71	0.93	2.18
PLCXD1	chrX	0.60	-0.29	-0.12	1.72	1.35	0.95	1.34	1.70
UPRT	chrX	0.67	1.23	0.02	0.05	1.83	2.30	1.13	1.32
BMX	chrX	0.57	1.52	-1.13	0.38	1.62	2.31	0.33	1.26
FUNDC2	chrX	1.23	0.44	0.22	0.10	2.58	1.79	1.37	1.13
RNF113A	chrX	0.29	1.02	0.83	-0.23	1.48	2.14	2.01	1.02
CXorf36	chrX	0.42	1.22	-0.24	-0.41	1.12	1.31	1.08	0.74

Table S2: Enriched sgRNA's and their corresponding genes. Screen: High doxycycline-early harvest

Gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
PLAUR	chr19q	1.09	-0.10	0.51	0.33	2.46	1.09	1.79	1.67
HPN	chr19q	1.21	-0.12	-0.17	0.29	1.38	1.14	0.81	1.23
SNRPD2	chr19q	0.34	0.43	-0.09	1.43	1.60	1.74	1.23	2.50
MED25	chr19q	0.60	1.24	1.73	0.91	1.37	4.17	5.47	2.70

Table S2 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
NBPF4/6	chr1	1.39	0.40	0.60	0.86	2.73	0.94	1.74	2.44
OR4F29	chr1	1.04	0.10	0.46	0.49	2.97	1.73	2.26	2.38
CISD1	chr10	-0.06	1.22	0.33	0.28	1.15	2.27	1.88	1.33
NCOA4	chr10	-0.72	1.76	-0.21	0.11	0.23	1.39	0.61	0.80
RPS25	chr11	1.70	-0.45	-0.25	1.80	1.41	0.55	0.58	3.12
MTRNR2L2/8	chr11	-0.09	-0.87	1.27	2.09	0.68	0.23	1.37	2.20
OR5B3	chr11	0.30	0.25	1.05	-0.30	1.26	1.08	2.11	0.96
CYB561A3	chr11	-0.06	1.26	0.17	-0.08	0.99	1.77	1.16	0.94
SNRPF	chr12	0.21	0.28	0.53	1.11	1.10	1.83	2.25	3.35
PRMT8	chr12	0.59	-1.15	-0.65	1.03	1.07	0.04	0.43	1.96
LRP6	chr12	1.01	0.38	-0.06	0.24	2.08	1.85	1.06	1.47
SLC2A13	chr12	1.12	0.76	-0.44	-0.08	2.25	1.10	0.61	0.92
GOLGA6L4	chr15	1.40	0.63	0.98	NA	8.62	2.25	4.99	NA
GOLGA6L10	chr15	0.75	0.49	1.34	0.91	2.25	0.99	6.18	3.51
PPCDC	chr15	0.80	0.55	-0.12	1.14	2.30	1.58	1.11	3.32
GOLGA6C	chr15	0.03	0.96	0.36	1.33	1.19	2.47	2.29	2.59
GOLGA6L9	chr15	0.32	-0.77	1.11	0.77	1.12	0.24	3.36	2.37
THAP10	chr15	1.06	0.19	0.06	0.57	2.49	1.88	1.08	1.93
PEAK1	chr15	0.54	1.02	0.16	0.17	1.85	3.01	1.07	1.78
ADAMTSL3	chr15	-0.11	1.21	0.81	0.05	1.14	2.30	1.62	1.33
RASL12	chr15	1.01	0.52	-0.19	0.41	2.83	1.81	1.06	1.31
UBE2Q2L	chr15	0.97	0.78	1.43	0.57	3.93	2.48	5.29	1.23
HERC1	chr15	0.17	0.02	1.31	0.06	1.36	0.94	2.25	1.17
OAZ2	chr15	0.46	0.89	1.18	0.25	2.40	3.02	3.58	1.13
ODF3L1	chr15	0.14	0.52	1.02	-0.40	1.07	1.67	1.82	0.69
RPL4	chr15	0.12	1.18	0.90	-0.48	1.13	2.41	1.59	0.52
SUMO2	chr17	-0.28	0.81	1.04	-1.55	0.51	1.08	2.31	0.01
MYEOV2	chr2	-0.60	0.97	0.17	1.05	0.35	1.78	1.08	2.49
C22orf46	chr22	-0.89	0.15	1.11	0.09	0.12	1.10	1.49	0.87
RHOA	chr3	0.30	-0.39	0.62	1.12	1.10	0.72	2.02	2.36
GJA1	chr6	-0.33	1.70	0.31	0.11	0.85	1.54	1.36	1.29
HECA	chr6	0.08	-0.94	2.10	-1.51	0.73	0.73	1.61	0.01
CYCS	chr7	1.83	1.85	0.40	0.93	4.20	4.48	1.44	3.26
NDUFA5	chr7	-0.26	-0.34	0.01	1.65	0.61	0.54	0.84	1.33
DNAJB9	chr7	-0.33	1.40	-0.98	0.76	0.46	2.24	0.08	1.21

Table S2 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
SLC12A9	chr7	1.00	0.36	0.18	0.03	2.31	1.85	1.43	1.06
PDAP1	chr7	0.36	-0.61	1.28	-0.49	1.16	0.34	1.41	0.38
ACER2	chr9	0.21	0.01	0.66	1.02	1.98	1.07	2.09	2.20
TOMM5	chr9	-0.10	0.20	1.20	0.27	1.14	1.17	1.93	1.68
ISCA1	chr9	1.20	0.13	0.19	-0.58	1.65	0.98	1.14	0.37
RBMY1J	chrY	0.34	0.07	0.14	1.00	1.69	1.06	1.36	2.54

Table S3: Enriched sgRNA's and their corresponding genes. Screen: Low doxycycline-late harvest

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
HPN	chr19q	2.07	0.14	-1.13	0.32	1.80	1.21	0.45	1.65
ZNF599	chr19q	0.42	0.24	-0.01	1.00	2.04	1.37	0.98	2.60
PPP1R15A	chr19q	0.27	-0.98	-1.04	1.66	1.11	0.54	0.38	1.36
MED25	chr19q	1.68	2.38	2.63	-0.03	2.54	4.23	5.30	2.10
CEACAM20	chr19q	-1.29	1.21	2.20	-0.64	0.33	1.75	2.61	0.55
MARK4	chr19q	-1.12	1.54	-0.56	-0.78	0.43	1.30	0.46	0.45
ZNF534	chr19q	1.01	1.10	-0.96	0.20	1.95	2.34	0.64	1.04
DAB2	chr5p	-0.76	0.96	1.89	0.03	0.89	1.53	2.26	1.51
C1QTNF3	chr5p	0.04	1.02	0.44	0.52	1.35	2.23	1.87	2.15
GDNF	chr5p	-0.44	1.57	-1.31	-1.11	0.44	1.31	0.32	0.38
OR2L2	chr1	0.64	1.24	0.34	1.00	2.11	3.15	1.04	2.78
PTGER3	chr1	0.16	0.28	-0.04	1.28	1.15	1.83	1.13	2.51
RBM15	chr1	-0.10	-0.18	0.75	1.11	1.27	1.05	1.93	2.08
PMVK	chr1	-0.49	-0.10	0.79	1.04	1.00	1.27	1.36	2.00
NLRP3	chr1	1.22	0.32	-0.36	0.77	2.24	1.64	1.12	1.96
SYNC	chr1	0.25	-0.58	-0.59	2.18	1.18	1.08	0.60	1.73
B3GALT6	chr1	0.48	1.83	-0.42	-0.23	1.47	1.52	1.03	1.47
RSG1	chr1	-0.06	1.06	-0.07	-0.06	1.01	2.41	0.69	1.35
TCHH	chr1	1.80	-0.08	-0.98	0.27	1.46	1.12	0.61	1.30
CASQ1	chr1	-1.50	-0.78	-0.86	1.54	0.20	0.47	0.42	1.30
CRTC2	chr1	0.28	-0.27	-1.47	1.55	1.13	0.98	0.22	1.30
RWDD3	chr1	-1.35	1.59	-0.12	0.10	0.29	1.33	0.87	1.24
CDC14A	chr1	-0.20	1.05	-0.61	1.03	1.09	2.36	1.00	1.13



Table S3 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
MCOLN2	chr1	1.06	-1.11	1.08	-0.17	1.14	0.47	2.41	1.03
UBE4B	chr1	-0.19	0.08	1.61	-0.44	1.10	1.34	1.47	0.85
CD34	chr1	0.50	1.61	-0.40	-0.54	1.34	1.50	1.26	0.80
MR1	chr1	-1.79	1.60	-0.69	-0.57	0.09	1.33	0.43	0.50
RNF115	chr1	1.57	0.31	-0.07	-1.52	1.33	1.31	1.17	0.19
FAM213B	chr1	-0.69	0.07	1.57	-2.14	0.50	0.83	1.31	0.03
RBBP5	chr1	-0.76	1.62	0.54	-2.29	0.46	1.57	1.35	0.02
NCOA4	chr10	0.06	1.07	-0.62	0.72	1.12	1.88	1.02	1.59
LIPK	chr10	-1.32	-1.09	-0.22	1.99	0.31	0.38	0.55	1.58
PPP3CB	chr10	1.02	-1.34	0.32	0.33	2.26	0.30	0.99	1.21
CISD1	chr10	-0.34	2.17	-0.61	0.23	1.05	1.72	0.87	1.13
CFAP43	chr10	1.54	-1.36	-0.84	-0.82	1.30	0.28	0.43	0.48
ARFIP2	chr11	0.49	0.49	0.65	1.05	1.77	1.01	2.73	4.37
RPS25	chr11	1.56	0.13	-0.08	1.11	2.48	1.78	1.31	2.38
TRIM68	chr11	1.00	-0.85	-0.43	1.58	1.32	0.76	0.76	2.32
RIC8A	chr11	0.49	-0.15	1.82	0.91	2.17	1.47	2.73	2.19
FAM76B	chr11	-0.70	-1.23	0.81	1.05	0.50	0.37	1.00	2.04
DCPS	chr11	-0.12	0.71	-0.53	1.70	1.29	1.39	1.23	1.87
MTRNR2L2/8	chr11	-0.39	0.32	0.65	1.86	0.86	1.30	1.47	1.75
MICALCL	chr11	-1.15	-1.74	-2.24	1.55	0.74	0.71	0.02	1.30
PAFAH1B2	chr11	1.35	0.24	-0.04	-0.20	2.04	1.38	1.19	1.07
OR5B12	chr11	1.14	-1.49	-1.15	1.01	2.33	0.21	0.38	1.06
DHCR7	chr11	0.74	1.34	0.43	-0.84	1.91	2.54	1.18	0.79
KMT2A	chr11	-0.72	1.26	1.43	-1.10	0.49	1.24	2.67	0.49
SLC29A2	chr11	-2.21	1.68	-0.66	-1.60	0.02	1.38	0.58	0.43
C11orf88	chr11	-1.20	0.04	1.77	-0.97	0.39	0.80	1.44	0.39
TRIM29	chr11	0.92	1.20	-1.23	-1.12	1.10	2.19	0.37	0.38
KRT83	chr12	0.38	-0.77	0.53	1.32	1.17	0.88	1.57	2.43
LOH12CR1	chr12	-0.96	0.29	0.34	1.35	0.64	1.19	1.24	2.18
PRB1	chr12	0.49	1.29	-0.30	0.71	1.66	2.37	1.27	1.66
LIMA1	chr12	-0.50	-0.20	-0.20	1.68	0.56	1.34	1.09	1.38
ATF7	chr12	-0.37	-0.11	1.57	-0.39	0.83	1.31	1.58	0.64
C12orf65	chr12	1.85	-0.85	-1.44	-0.32	1.49	0.42	0.24	0.49
RERG	chr12	-0.61	-1.48	2.40	-0.74	0.48	0.21	1.91	0.43
NR4A1	chr12	1.31	1.12	-1.85	-0.85	2.49	1.17	0.08	0.42



Table S3 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
GLT8D2	chr12	0.30	1.66	0.44	-1.19	1.36	2.20	1.39	0.40
RASSF8	chr12	1.76	-1.16	-0.70	-1.67	1.43	0.38	0.44	0.13
MTMR6	chr13	-1.67	-0.02	2.01	-0.41	0.13	0.78	1.60	0.73
ATG2B	chr14	-1.45	-1.09	-1.05	1.70	0.23	0.38	0.65	1.39
PNP	chr14	0.97	-0.48	1.52	-0.16	1.38	1.17	2.27	1.29
RNASE1	chr14	0.19	2.10	0.98	-0.79	1.67	2.29	1.91	0.85
KCNH5	chr14	0.43	1.69	-1.11	-0.66	1.38	1.38	0.47	0.53
TECPR2	chr14	-0.11	-0.33	1.59	-1.48	0.89	0.64	1.33	0.21
GOLGA6L4	chr15	-1.82	-0.22	-0.85	NA	0.14	1.62	1.30	NA
CRTC3	chr15	-0.12	1.28	0.03	0.31	1.15	2.26	1.18	1.53
NR2F2	chr15	-0.02	1.58	-0.53	0.08	1.28	1.43	0.85	1.32
CHRNA5	chr15	0.57	1.14	0.28	-0.74	1.62	2.15	1.07	0.93
ESRP2	chr16	0.37	0.95	-1.60	1.09	1.03	2.24	0.16	2.38
ZNF720	chr16	-0.54	-0.11	-1.60	2.15	0.64	0.64	0.16	1.70
PSMB10	chr16	-0.02	-0.85	0.07	1.07	0.83	0.77	1.02	1.44
FOXL1	chr16	-1.08	0.82	1.04	-0.05	0.50	1.38	2.05	1.00
PLA2G15	chr16	1.25	-0.05	1.17	-1.67	2.57	1.13	1.37	0.13
MED31	chr17	0.10	-1.19	-0.98	1.59	0.87	0.38	0.41	1.33
ZZEF1	chr17	0.15	0.06	1.02	-0.14	1.59	0.99	2.21	0.94
HOXB8	chr17	1.70	-0.68	-0.96	-0.51	1.39	0.51	0.43	0.63
SUMO2	chr17	0.24	0.04	1.06	-1.64	1.07	1.01	1.55	0.14
ALPK2	chr18	-1.40	-0.37	0.97	1.01	0.25	0.83	0.98	2.27
FBXO15	chr18	1.03	-0.38	0.51	0.36	2.36	0.99	1.60	1.52
MBD2	chr18	-0.83	1.01	0.98	-0.45	0.73	2.28	0.98	0.79
PARD6G	chr18	-0.75	1.58	-1.27	-1.67	0.46	1.32	0.40	0.13
UGT1A1	chr2	0.43	0.27	-1.63	1.09	1.37	1.03	0.15	2.13
CCNT2	chr2	0.59	-0.21	-0.55	1.58	1.32	1.24	1.06	1.66
DTNB	chr2	1.63	-0.14	-0.33	-0.09	1.71	1.20	0.65	1.35
LPIN1	chr2	1.18	0.58	-0.56	0.23	2.02	1.65	1.09	1.22
GPR35	chr2	0.32	1.01	0.29	0.19	2.19	3.27	1.20	0.98
HSPA12B	chr20	1.10	0.62	-1.67	0.68	3.09	1.03	0.13	1.82
CST1	chr20	-1.59	-1.86	0.19	2.21	0.57	0.07	1.00	1.77
GNAS	chr20	1.03	0.94	0.59	-0.24	2.99	2.23	1.94	1.00
NSFL1C	chr20	0.75	1.20	-1.63	-0.84	1.11	1.93	0.15	0.43
CYP24A1	chr20	1.62	-0.82	0.50	-1.37	1.50	0.43	1.35	0.28



Table S3 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
JAM2	chr21	-0.14	-1.37	0.89	1.19	1.09	0.28	1.19	2.15
KCNE2	chr21	0.29	-0.23	1.48	0.68	1.82	1.26	2.18	1.95
KCNE2	chr21	0.29	-0.23	1.48	0.68	1.82	1.26	2.18	1.95
PLAC4	chr21	0.11	-0.72	1.73	0.50	1.41	0.95	1.72	1.50
COL18A1	chr21	0.14	1.52	-0.55	-0.02	1.28	1.43	0.92	1.24
EMID1	chr22	-0.40	-1.00	1.69	-2.17	0.46	0.38	1.38	0.02
TSC22D2	chr3	0.53	0.35	0.09	1.15	2.33	1.56	1.07	2.90
ADPRH	chr3	0.20	0.84	0.01	1.05	1.95	2.07	1.00	2.65
TUSC2	chr3	-0.06	1.16	-0.63	1.43	1.23	1.35	1.10	2.55
RAB5A	chr3	-1.31	0.03	0.16	1.86	0.32	0.95	1.49	1.53
PRR23B	chr3	1.06	0.58	0.59	-0.16	2.99	1.67	2.15	1.02
ZNF501	chr3	0.64	1.77	-0.63	-1.15	1.43	1.76	0.55	0.43
RSRC1	chr3	1.35	-0.42	1.22	-1.31	2.62	0.77	1.19	0.31
SNRK	chr3	-0.10	-0.37	1.57	-1.42	0.83	0.64	1.31	0.24
ERC2	chr3	1.15	-0.65	1.07	-1.51	2.42	0.54	1.07	0.20
NGLY1	chr3	1.61	-0.45	-0.81	-1.51	1.34	0.44	0.44	0.20
FAM47E-STBD1	chr4	0.30	0.31	0.28	1.02	2.03	2.75	1.99	2.79
PPA2	chr4	1.26	0.06	-0.19	0.04	2.06	1.55	0.81	1.14
G3BP2	chr4	1.66	-0.27	-0.59	-0.49	1.37	1.16	0.52	0.68
TMEM155	chr4	0.99	-1.15	1.28	-0.50	1.15	0.44	2.32	0.67
EPGN	chr4	-0.13	1.54	-0.21	-0.59	1.17	1.30	1.07	0.62
WHSC1	chr4	-0.13	-0.77	1.97	-0.82	0.81	0.62	1.57	0.46
GUCY1A3	chr4	-0.65	-0.76	1.59	-0.76	0.89	0.47	1.33	0.43
FBXW7	chr4	-0.80	1.89	-0.17	-1.21	0.45	1.51	0.58	0.39
FBXO30	chr6	0.69	-0.38	0.59	1.13	1.84	1.06	1.61	3.00
ULBP2	chr6	0.74	-0.32	-0.01	1.98	1.48	1.17	1.23	1.72
CCNC	chr6	-0.67	0.58	1.08	0.60	1.03	1.04	2.98	1.67
CCDC170	chr6	-0.82	-0.93	-1.95	1.80	0.49	0.39	0.06	1.46
CLVS2	chr6	-1.05	-1.30	-2.05	1.72	0.65	0.41	0.04	1.40
AGPAT4	chr6	1.16	0.48	0.25	-0.15	2.18	2.07	1.47	1.08
GJA1	chr6	0.37	1.28	0.33	-0.80	1.28	2.29	1.15	0.84
SUMO4	chr6	2.06	-0.20	-1.04	-0.60	1.64	0.57	0.54	0.56
B3GAT2	chr6	-1.45	-0.06	1.62	-0.81	0.23	0.69	1.34	0.44
ZPBP	chr7	0.92	0.23	-1.45	1.40	2.01	1.22	0.23	2.20
NUB1	chr7	-0.50	0.41	-0.03	1.04	1.00	1.35	1.34	1.42

Table S3 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
RADIL	chr7	0.35	-0.95	1.82	0.29	1.47	0.64	2.19	1.25
HIPK2	chr7	-0.58	1.57	-1.27	0.19	0.59	1.32	0.35	0.99
LMBR1	chr7	0.06	1.07	1.11	-0.91	1.05	1.61	2.42	0.69
HUS1	chr7	-1.17	1.56	-0.99	-0.97	0.38	1.31	0.42	0.58
C1GALT1	chr7	-0.96	0.71	1.40	-1.77	0.39	1.22	1.87	0.10
LRRD1	chr7	1.66	-1.26	-0.56	-2.12	1.37	0.40	0.43	0.03
RALYL	chr8	0.99	-1.04	-0.47	1.09	1.03	0.55	0.70	2.30
STAR	chr8	-0.72	-0.06	-0.36	1.54	0.68	0.95	0.84	1.30
PXDNL	chr8	1.25	0.56	0.14	-0.10	2.32	1.79	1.62	1.13
ZNF707	chr8	-0.28	-0.42	1.56	-0.62	1.12	0.77	1.31	0.51
QSOX2	chr9	0.01	2.03	-0.49	0.17	1.36	1.62	0.97	1.49
PTPRD	chr9	-1.50	-0.15	0.42	1.51	0.20	1.18	1.28	1.36
TOMM5	chr9	-0.53	-0.41	1.83	-0.08	0.67	0.78	1.48	1.29
LURAP1L	chr9	1.59	0.26	0.42	-0.55	2.09	1.33	1.36	1.25
IFNE	chr9	0.92	-0.88	1.09	-0.82	1.03	0.43	2.20	0.73
CXorf51A/B	chrX	1.57	-0.43	0.46	0.82	2.28	1.45	1.73	1.77
MED12	chrX	0.05	1.30	0.01	0.07	1.16	2.65	0.83	1.57
KCND1	chrX	0.25	0.37	1.60	-0.08	1.33	2.08	2.40	1.29
OGT	chrX	-0.05	0.37	1.43	0.33	1.24	2.28	2.47	1.28
HSFX1/2	chrX	0.33	1.62	0.56	-0.07	1.23	1.42	1.35	1.08
FUNDC2	chrX	1.69	-0.82	-0.66	-0.19	1.38	0.53	0.57	0.81
CACNA1F	chrX	-0.97	1.85	-0.68	-0.09	0.51	1.49	0.62	0.66
ARAF	chrX	0.21	1.26	1.32	-1.03	1.17	1.97	2.68	0.55
CXorf36	chrX	-0.76	1.55	0.17	-0.85	0.76	1.30	0.97	0.46
KDM5D	chrY	-1.37	1.93	-0.06	-1.56	0.44	1.54	0.69	0.17
RPS4Y2	chrY	1.58	-1.19	0.07	-1.70	1.32	0.38	0.84	0.12

Table S4: Enriched sgRNA's and their corresponding genes. Screen: High doxycycline-late harvest

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
PAFAH1B3	chr19q	-0.23	1.22	0.43	-0.36	1.27	2.00	1.44	1.06
GPATCH1	chr19q	0.20	0.99	-0.86	1.10	1.04	2.12	0.99	2.24
HPN	chr19q	2.56	-1.27	-0.17	-0.79	1.73	0.55	0.70	0.58

Table S4 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
ZNF793	chr19q	0.14	1.58	-0.13	-0.17	1.46	2.49	1.21	1.04
EGLN2	chr19q	-1.98	-0.90	1.90	-0.43	0.12	0.50	1.36	0.51
MED25	chr19q	0.00	0.76	1.66	0.18	1.25	2.09	2.94	1.93
ZNF221	chr19q	0.78	-0.45	-1.36	1.78	1.30	0.93	0.46	1.96
LIN7B	chr19q	-1.23	2.31	0.02	-0.11	0.59	1.57	1.49	0.90
SIGLEC6	chr19q	-0.71	1.37	0.75	0.49	1.12	2.81	1.92	1.28
FPR3	chr19q	-0.83	1.00	0.69	0.67	0.94	3.21	1.83	1.08
KLK13	chr19q	1.09	0.35	0.24	0.15	3.39	2.22	1.33	0.98
IGLON5	chr19q	-1.57	0.51	1.50	-1.44	0.31	1.18	1.57	0.39
BIRC8	chr19q	0.29	-0.36	1.68	-0.31	1.26	1.14	2.01	1.25
NLRP4	chr19q	0.26	-0.32	1.61	-0.45	1.22	1.20	1.78	1.12
C1QTNF3	chr5p	-0.62	-0.58	0.63	1.86	0.98	1.25	1.49	1.55
C5orf22	chr5p	-1.15	1.02	-0.09	1.09	0.67	1.51	0.99	2.26
C5orf49	chr5p	2.37	-0.95	-0.46	-1.94	1.61	0.47	0.50	0.13
IRX1	chr5p	-0.04	1.20	-2.20	1.03	1.05	2.28	0.06	1.63
CD53	chr1	0.66	0.06	-0.05	1.30	1.83	1.79	1.09	2.80
GSTM3	chr1	-0.42	0.66	0.27	1.10	0.99	1.85	1.79	2.29
RCOR3	chr1	-1.63	-0.71	0.99	1.33	0.27	0.64	1.10	2.24
RRAGC	chr1	0.84	0.20	-0.40	1.37	2.03	1.91	1.12	2.13
USH2A	chr1	0.15	-1.30	0.62	1.84	1.33	0.52	1.73	2.02
OR2T35	chr1	-0.18	0.47	-0.23	1.60	1.06	1.18	1.00	1.79
LCE2C/D	chr1	-0.70	0.91	-0.33	1.78	0.72	1.13	0.79	1.77
NBPF4/6	chr1	1.82	0.42	-0.71	1.52	2.13	1.04	0.81	1.65
B4GALT3	chr1	-0.39	-0.65	-0.63	1.56	1.20	0.54	0.72	1.39
CASQ1	chr1	-1.58	-1.33	0.36	1.34	0.31	0.38	1.11	1.35
SMCP	chr1	1.18	-0.23	-0.20	0.00	2.34	0.88	1.04	1.33
WDR47	chr1	-0.94	-1.32	1.61	0.54	0.51	0.48	1.61	1.22
TMEM167B	chr1	0.96	1.20	-1.78	-0.12	1.47	2.20	0.20	1.05
LCE3C	chr1	1.79	-1.01	-1.11	0.14	1.31	0.72	0.45	1.04
PPM1J	chr1	0.33	0.34	1.12	-0.42	1.32	1.85	2.41	1.00
SORT1	chr1	-1.42	0.97	1.10	-0.60	0.42	0.99	2.21	0.75
HPCA	chr1	0.48	1.04	0.59	-1.29	0.96	2.78	1.69	0.53
CAMK1G	chr1	-2.42	2.44	-0.87	-1.88	0.03	1.65	0.59	0.44
NES	chr1	1.94	-0.33	-0.93	-1.44	1.38	0.57	0.48	0.40
RNF115	chr1	1.76	0.24	0.08	-1.58	1.87	1.29	1.18	0.30

Table S4 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
CFHR4	chr1	0.80	1.24	0.09	-1.92	1.88	2.00	1.07	0.14
HPS1	chr10	1.17	0.47	-1.81	1.34	2.44	1.11	0.18	2.74
WNT8B	chr10	-0.30	0.55	0.27	1.15	1.02	2.16	1.63	2.29
MSMB	chr10	-1.51	1.00	0.06	1.72	0.35	1.84	1.28	2.24
DHX32	chr10	-0.04	1.21	0.05	0.12	1.02	2.83	1.05	1.81
LIPN	chr10	-0.02	1.85	0.63	0.16	1.34	2.90	2.03	1.74
ZNF485	chr10	-1.33	-1.46	1.75	0.54	0.39	0.38	1.60	1.29
SLC18A2	chr10	0.10	1.57	0.19	-0.19	1.21	2.43	1.90	1.10
AKR1C2	chr10	-0.68	1.53	-1.00	0.72	0.51	1.52	0.50	1.03
CISD1	chr10	-1.18	1.98	-0.35	-0.72	0.56	1.40	0.64	0.63
MSRB2	chr10	1.63	1.75	0.90	-1.87	2.87	3.71	1.29	0.16
RPS25	chr11	1.87	-1.73	-2.02	1.60	2.84	0.49	0.10	1.34
CNTN5	chr11	-1.02	1.80	-2.08	0.69	0.44	1.83	0.09	1.31
OR6T1	chr11	0.44	1.58	0.76	-0.82	1.21	2.67	1.94	1.11
MMP1	chr11	-0.90	1.92	0.42	-0.34	0.99	1.43	1.37	1.10
ATG16L2	chr11	0.96	-2.15	1.02	-0.57	0.95	0.07	2.20	0.78
DUSP8	chr11	-1.66	-2.34	1.85	-0.30	0.45	0.04	1.34	0.60
ANKRD13D	chr11	0.76	1.91	-0.37	-1.26	1.36	1.93	1.04	0.56
OR52L1	chr11	-0.51	2.16	-0.99	-1.40	0.47	1.49	0.45	0.43
COLCA2	chr11	0.93	0.10	1.12	-1.58	1.91	1.00	2.15	0.30
PUS7L	chr12	-1.63	0.83	-0.31	1.09	0.27	1.15	0.98	2.03
CLEC4A	chr12	0.55	1.22	-0.34	-0.22	1.63	2.04	1.06	1.29
LRCOL1	chr12	-0.94	0.52	1.61	0.11	0.92	1.58	1.93	1.23
PRMT8	chr12	2.24	-1.21	-1.52	0.08	1.53	0.38	0.35	0.97
STAC3	chr12	-2.07	1.31	0.86	-0.88	0.09	2.07	1.10	0.51
THSD1	chr13	-0.65	-0.05	0.18	1.66	1.09	1.25	1.39	1.59
CDC42BPB	chr14	-1.63	-1.39	0.37	1.78	0.27	0.38	1.30	1.36
ITPK1	chr14	1.01	-0.72	0.14	0.16	2.00	0.94	1.07	1.26
GPR33	chr14	-0.83	1.38	-0.88	1.38	1.01	2.65	0.54	1.13
RNASE1	chr14	1.04	0.66	0.06	-1.64	1.84	1.79	0.96	0.27
EXD1	chr15	1.47	-0.58	-0.64	1.36	2.62	1.17	0.77	1.41
SNRPN	chr15	-0.33	0.87	1.02	-0.27	0.95	2.07	2.08	1.20
LRRK1	chr15	1.93	0.17	-0.50	-0.87	1.37	1.08	1.02	0.87
PKM	chr15	2.37	-1.73	-1.97	-0.89	1.61	0.44	0.12	0.49
ADGRG1	chr16	0.75	-0.75	0.89	1.14	1.22	1.01	2.11	3.40

Table S4 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
FTO	chr16	1.04	0.19	-0.01	0.39	2.93	1.39	0.96	2.11
ST3GAL2	chr16	-0.61	0.34	1.46	0.52	1.16	1.46	2.46	1.58
VPS4A	chr16	-1.43	0.29	-0.21	1.51	0.41	1.24	1.18	1.31
MT1X	chr16	-1.15	1.66	-0.22	0.50	0.57	1.36	1.02	1.16
PSMB10	chr16	0.23	1.43	-0.23	-0.28	1.27	2.19	1.16	1.15
NOMO3	chr16	1.62	-0.28	0.40	-1.07	1.34	0.89	0.96	0.81
OR1D2	chr17	0.61	0.42	1.21	0.64	1.75	1.05	4.25	3.10
APPBP2	chr17	-0.13	0.12	1.13	0.94	1.01	1.95	2.60	2.17
KRT33A	chr17	0.17	-1.27	0.24	1.17	1.03	0.55	1.17	2.05
MED24	chr17	1.28	0.24	-0.34	0.25	2.22	1.18	1.09	2.05
UNK	chr17	-1.02	-1.67	-0.03	2.23	0.44	0.25	0.84	1.53
CCL2	chr17	-0.73	1.12	-0.25	0.98	1.00	2.23	1.23	1.24
RECQL5	chr17	-1.08	0.18	2.04	-0.26	0.75	1.22	1.43	1.10
KRTAP4-11	chr17	-2.28	1.38	-1.58	-2.25	0.92	2.08	1.54	1.04
CYB5A	chr18	0.73	-0.01	-0.19	1.12	1.89	1.69	1.00	2.44
RAB31	chr18	0.10	-1.02	0.90	1.13	1.01	0.82	1.91	2.12
GALNT1	chr18	-0.29	-1.14	-1.82	1.83	0.60	0.40	0.18	1.32
DYM	chr18	0.22	2.57	-1.57	-0.23	1.27	1.74	0.31	1.14
TEX261	chr2	1.02	-1.95	-0.17	1.34	1.39	0.13	1.11	2.27
CAPN14	chr2	-0.29	1.42	0.25	0.88	1.15	2.24	2.10	2.16
SUPT7L	chr2	-0.62	-0.62	0.20	1.69	0.73	1.12	1.26	1.44
KCNK12	chr2	1.26	-0.87	0.31	0.40	2.37	1.02	1.07	1.39
SPC25	chr2	2.28	-1.04	-0.64	0.27	1.55	0.71	0.80	1.22
CALCLRL	chr2	-0.74	1.05	-0.75	0.93	0.97	2.15	0.62	1.21
ORC4	chr2	1.28	0.91	-0.43	-0.37	2.14	1.83	1.04	1.09
INHBB	chr2	1.23	0.04	1.18	-0.53	2.45	1.61	1.79	1.06
LPIN1	chr2	1.05	1.22	-1.03	-0.09	1.52	2.31	0.81	1.06
ATG16L1	chr2	2.08	-0.62	-1.03	-0.09	1.46	0.78	0.73	0.81
ALLC	chr2	-0.11	-1.39	1.59	-0.16	1.22	0.44	1.40	0.75
MAL	chr2	-0.85	1.87	-0.99	-1.78	0.45	1.34	0.43	0.20
DEFB119	chr20	0.59	1.58	-0.73	0.62	1.25	3.03	1.21	1.74
LKAAEAR1	chr20	0.24	1.12	-0.94	0.27	1.00	2.21	0.92	1.22
TMEM230	chr20	-0.03	0.17	1.20	-0.28	1.08	1.64	2.20	1.05
SH3BGR	chr21	0.74	0.94	-1.77	1.68	1.26	2.18	0.20	3.37
RSPH1	chr21	1.65	0.55	-1.27	-0.28	1.62	1.25	0.55	1.19



Table S4 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
KRTAP11-1	chr21	-1.50	-0.75	2.18	0.24	0.36	0.61	1.50	1.18
PDXK	chr21	0.45	-1.17	1.37	0.19	1.48	0.65	2.10	1.12
ABCC5	chr3	-0.78	1.98	-0.21	0.45	1.17	1.47	1.31	1.40
P2RY1	chr3	-1.14	1.58	-0.85	0.35	0.53	1.32	0.68	1.21
SKIL	chr3	1.30	-0.85	0.55	0.17	2.05	1.05	1.63	1.09
GRIP2	chr3	-0.23	0.10	1.22	-0.06	1.00	1.58	2.32	1.06
FAM19A1	chr3	-0.23	1.81	0.14	-1.09	1.05	1.31	1.27	0.74
OR5H14	chr3	0.44	0.72	1.08	-0.87	0.83	1.52	2.06	0.69
LSMEM2	chr3	-0.44	1.12	0.88	-2.24	0.94	2.09	1.00	0.05
HPGD	chr4	-0.49	-1.19	0.95	1.04	0.87	0.62	0.96	2.18
MAD2L1	chr4	0.91	-0.27	1.38	0.07	2.13	1.12	2.23	1.84
PSORS1C1	chr6	-2.44	-1.27	-1.60	2.03	0.03	0.63	0.43	1.42
GUCA1A	chr6	-1.79	2.23	-1.74	0.35	0.19	1.52	0.49	1.32
CD24	chr6	1.64	0.12	1.30	-0.53	2.58	1.63	1.96	1.24
NR2E1	chr6	1.60	1.17	0.63	-0.52	3.14	2.45	1.64	1.22
SMLR1	chr6	1.86	-0.37	0.39	-0.71	1.39	1.29	1.34	1.04
ID4	chr6	-1.39	0.63	1.52	-0.58	0.44	1.18	1.75	0.78
GJA1	chr6	-1.86	1.52	0.33	-0.59	0.16	1.30	1.18	0.76
EEF1E1	chr6	1.13	0.86	0.78	-0.20	2.40	1.47	1.06	0.75
LTV1	chr6	1.36	-0.66	0.40	-1.21	1.40	0.69	1.12	0.61
COA1	chr7	1.09	0.27	-0.33	0.90	2.29	2.07	0.98	2.13
PEX1	chr7	-3.15	1.18	0.23	0.63	0.00	2.20	1.04	1.75
CHCHD2	chr7	1.97	-0.58	-0.77	-0.46	1.39	0.77	0.50	1.18
NOS3	chr7	0.07	-0.63	1.67	-0.29	1.25	0.96	1.42	1.17
RADIL	chr7	-1.02	1.32	0.70	-0.06	0.83	1.85	1.57	1.10
C1GALT1	chr7	-2.64	1.04	0.23	-0.11	0.01	1.49	1.16	0.96
TMEM140	chr7	0.19	0.97	1.51	-1.00	1.18	2.10	2.21	0.85
AMPH	chr7	1.16	1.58	-0.59	-0.66	1.36	2.43	1.21	0.77
FAM71F1	chr7	1.03	-0.31	0.75	-1.16	1.92	0.95	1.13	0.66
ZPBP	chr7	2.14	-0.80	-0.68	-1.01	1.48	0.57	0.84	0.43
DUSP4	chr8	0.43	0.41	-1.58	1.06	1.44	0.97	0.30	2.60
SLA	chr8	0.38	-0.72	0.27	1.22	1.37	1.06	1.26	2.28
TDRP	chr8	0.25	1.18	0.23	-0.37	1.96	2.19	1.18	1.04
EXT1	chr8	0.40	1.69	-0.43	-0.87	1.26	1.40	1.02	0.96
EXD3	chr9	1.11	-1.58	-1.41	1.15	1.02	0.30	0.38	2.37

Table S4 continued

Target gene	chr	LFC 1	LFC 2	LFC 3	LFC 4	-LogP 1	-LogP 2	-LogP 3	-LogP 4
OR1B1	chr9	0.32	0.26	-1.04	1.03	1.28	0.96	0.80	2.26
COL5A1	chr9	0.20	0.17	-0.41	1.14	1.88	1.11	1.02	2.06
GTF3C4	chr9	0.41	1.17	-0.21	0.20	2.13	2.38	1.03	1.42
DNAJB5	chr9	0.24	1.76	-0.84	0.29	1.24	2.23	1.06	1.29
C9orf62	chr9	0.05	1.39	0.52	-0.21	1.57	2.39	1.80	1.14
TLN1	chr9	1.50	-2.37	0.51	-0.54	1.56	0.04	1.18	0.81
TMEM8B	chr9	-1.60	1.80	-0.59	-1.73	0.43	1.31	0.45	0.22
CXorf51A/B	chrX	2.13	0.35	-0.84	0.30	1.72	1.53	0.84	1.41
FAM127A	chrX	0.93	-0.68	1.06	0.13	1.97	0.97	2.15	1.33
FUNDC2	chrX	1.70	0.34	1.39	0.13	3.35	2.45	2.67	1.27
NLGN3	chrX	-1.24	1.32	-1.47	0.89	0.38	2.11	0.38	1.10
RPA4	chrX	2.12	-0.39	0.20	-0.54	1.61	1.13	1.47	1.01
SPANXN2	chrX	1.83	-0.88	0.17	-1.08	1.33	0.75	1.07	0.51
ARAF	chrX	0.02	-1.35	1.81	-1.63	0.90	0.38	1.31	0.28
Non-Target	unknown	1.21	0.14	0.50	-0.07	2.76	1.55	1.99	1.05
Non-Target	unknown	-0.31	1.82	0.07	-0.43	1.14	1.82	1.32	0.96

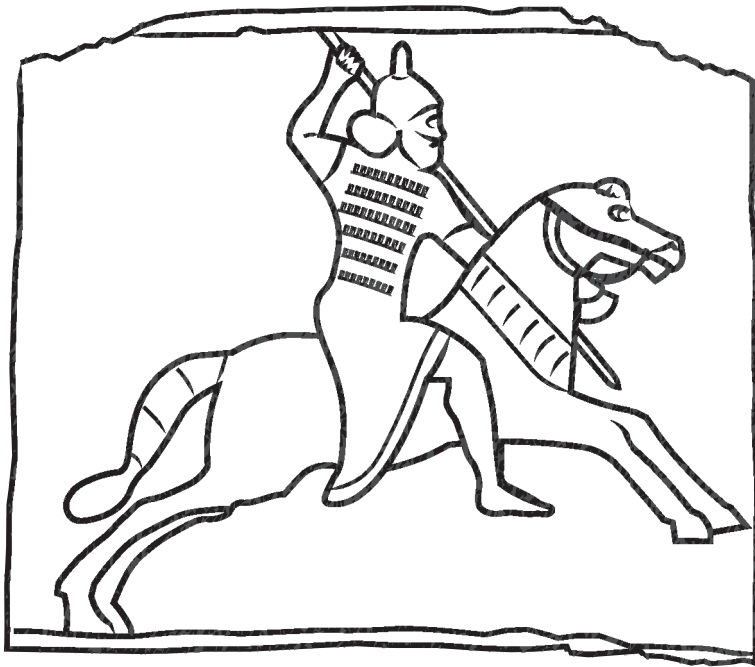


Illustration based on a stone carving on display at the British museum

Chapter 5

CRISPR-mediated functional silencing of DUX4

Ator Ashoti, Richard Lemmers, Silvère van der Maarel and Niels Geijsen

Abstract

Facioscapulohumeral muscular dystrophy (FSHD) is a genetic muscle disorder that leads to progressive muscle weakness predominantly of the face, shoulders and upper arms. The origin of FSHD lies in the D4Z4 repeat in the subtelomeric region of chromosome 4q, which contains the DUX4 gene. Ectopic expression of DUX4 is cytotoxic to muscle cells. The open reading frame of DUX4 is present in all D4Z4 sequences within the human genome, which hampers the application of conventional CRISPR/Cas9 genome-editing strategies to try and block DUX4 expression. Here we report an alternative targeting strategy that does not target the coding sequence of the DUX4 gene itself, but instead makes use of a relative unique region in the 3' UTR, upstream of the polyadenylation signal. This could potentially destabilize the DUX4 transcript, and thus minimize the possibility of DUX4 translation. Our DUX4-inducible cell line contains the genomic DUX4 sequence, and we were able to rescue cells from DUX4-induced apoptosis. This demonstrates the feasibility that these types of targeting strategies may abrogate DUX4 expression, and the potential for FSHD treatment in the future.

Introduction

Double homeobox 4 (DUX4), a pioneer transcription factor^{1,2}, is the main cause for the development of facioscapulohumeral muscular dystrophy (FSHD) due to its inappropriate expression in muscle³⁻⁵. The open reading frame (ORF) of the DUX4 gene can be found in D4Z4 repeats, which are present at several distinct loci in the human genome and in a tandem repeat sequence on chromosomes 4 and 10⁶⁻⁸. The pathological expression of DUX4 is caused by multiple genetic and epigenetic events that initiate the epigenetic de-repression of DUX4 at the subtelomeric region of chromosome 4. Contraction of the D4Z4 tandem repeat array on chromosome 4 to less than 10 D4Z4 repeats removes chromatin features needed for effective repression of this locus⁹ (FHSD1). Alternatively, loss-of-function mutations in chromatin modifier genes, such as structural maintenance of chromosome hinge domain 1 (SMCHD1) or DNA-methyltransferase 3 beta (DNMT3B)¹⁰⁻¹² can also cause de-repression of the D4Z4 repeat array and its embedded DUX4 gene (FSHD2). These two modifier genes collaboratively establish and maintain the hypomethylated state of their target genes, including DUX4, thereby repressing their expression¹³⁻¹⁸. Insufficient epigenetic repression of the D4Z4 repeats (due to a contracted D4Z4 repeat array, and/or the loss of epigenetic modifier genes) results in chromatin relaxation, which in itself is not enough to cause FSHD, but does render the DUX4 gene permissive for transcription. Only when de-repression occurs in a 4qA genetic background can it lead to the development of FSHD. The DUX4 transcript is stabilized by a polyA sequence that is present in exon 3 on the 4qA allele. This stabilized DUX4 transcript can then be translated into DUX4 protein and lead to the development of FSHD⁴. The 4qB variant on the other hand that does not possess this polyA sequence, diminishing pathological DUX4 expression, and is therefore generally classified as non-pathogenic^{19,20}.

In general, silencing disease-causing genes in gain-of-function disorders is a relatively straightforward approach with new genome-editing techniques such as CRISPR/Cas9²¹⁻²³. Although this may be an option for many genetic disorders, FSHD has a much more complex genetic and epigenetic structure, which complicates a simple targeting approach. The presence of multiple copies of DUX4 throughout the human genome complicates the use of genome-editing techniques to silence the gene. Targeting DUX4 with CRISPR/Cas9 can lead to shortening of the D4Z4 repeat sequence, and possibly aggravate the pathophysiology of both FHSD1 and FSHD2. In FSHD1, it can shorten an already contracted sequence, which results in further loss of repressive chromatin. In FSHD2, it can shorten a normal-sized D4Z4 allele to a contracted D4Z4 allele, in addition to the mutation in the SMCHD1/DNMT3B chromatin-modifier gene. Together with the above risks of using CRISPR/Cas9 to target the DUX4 locus, the occurrence of D4Z4 repeats throughout the genome will result in Cas9-induced double-strand breaks at multiple places in the host's genome, which can have unpredictable and unwanted outcomes including off-target insertions and deletions (indels) or translocation events. Recent attempts at targeting DUX4 directly include systems that do not lead to DNA damage, such as the use of antisense morpholino oligonucleotides to target and knock-down the DUX4 transcript²⁴. In another study, a catalytically disabled Cas9 fused to a *Krüppel*-associated box (dCas9-KRAB) was used to target the promotor of DUX4, inducing epigenetic repression of DUX4²⁵. Both studies show the ability to successfully diminish DUX4 expression in patient-derived cells. While these studies show promising results, unless a gene-therapy approach is taken to introduce these systems in vivo (which has other practical and ethical issues), the fact that these approaches have transient effects makes them less ideal for the long-term treatment of FSHD.

We therefore explored options to target DUX4 directly, in order to permanently disable expression of this gene. One promising approach is to use CRISPR/Cas9 to target a region of DUX4 that is not in the ORF of the gene and that does not frequently occur in other regions of the human genome. Lemmers et al. recently described a relative unique sequence present in the most distal copy of DUX4, termed the E3 sequence^{4,26}. The E3 sequence is located downstream of the DUX4 stop codon and upstream from the polyA signal. Here, we explored if CRISPR/Cas9-mediated targeting of the E3 sequence can abrogate DUX4 expression in our DUX4-inducible in vitro model.

Results

Direct targeting of DUX4-E3 with CRISPR/Cas9 modified systems

The E3 sequence upstream of the polyadenylation signal (PAS) is found in a subpopulation of patients with the 4qA allele²⁶. The E3 sequence is a promising region to directly target the DUX4 gene. CRISPR/Cas9-generated indels at this region can potentially disrupt regulatory function and destabilize the DUX4 transcript. To explore the possibility of abrogating DUX4 expression by targeting the E3 sequence, we designed several guide RNA (gRNA) sequences targeting this relatively unique region using the online WU-CRISPR gRNA design algorithm^{27,28}. Of the possible gRNA sequences, two were selected for further analysis (Fig. 1A) based on their predicted effectiveness as well as the low predicted chance of off-target editing events at other genomic loci²⁹.

Using our DUX4 inducible expression (DIE) cell model system (described in detail in chapter 2), in which DUX4 expression is induced in a doxycycline-dependent manner, we tested the ability of these two gRNAs to inhibit DUX4 expression. DIE cells were transduced with recombinant CRISPR/Cas9 ribonucleoprotein complex using the iTOP transduction method³⁰. DUX4 was induced 48-96 hours (h) after CRISPR/Cas9 transduction. 24h after doxycycline-mediated induction of DUX4 expression, survival of DIE cells was measured by fluorescence-activated cell sorting (FACS) analysis. As shown in Figure 1B, CRISPR/Cas9 targeting of the E3 sequence significantly increased cell survival post-DUX4 induction compared to the control (Cas9 protein, no guide) (Fig. 1B; gRNA1: 24.3% \pm 3.1%, p-value = 7E-05; gRNA2: 18.2% \pm 1.7%, p-value = 2E-05). gRNA1 was significantly more efficient at promoting cell survival after doxycycline induction of DUX4 expression, compared to gRNA2 (p-value = 0.008) (Fig. 1B). To further increase knock-out efficiency, synthetic single guide RNA (sgRNA) were used from Synthego (California, USA) that carry prime-end thiol modifications to increase RNA stability. Furthermore, spCas9 was optimized to contain four SV40 nuclear localization signals (NLSs) at the protein's N-terminal and two SV40 NLSs at its C-terminal, to improve its nuclear import ability³¹. These modifications further increased editing efficiency and significantly increased DIE cell survival to around 51.6% (\pm 1.56%, p-value = 0.004) (Fig. 1C). Unexpectedly, repeated targeting of the E3 sequence either by the same gRNA or by using different gRNAs in each round of targeting, only incrementally enhanced DIE cell survival (Fig. 1D). However, these data could be an overestimation of the DIE cell rescue due to proliferation of the positively-targeted cells between the time of doxycycline administration and FACS analysis (16-18 h). To examine this, we analyzed CRISPR/Cas9-mediated disruption of DUX4 at the clonal level. DUX4 targeted DIE cells were single-cell sorted into a 96-well plate 48 h after CRISPR/Cas9 transduction and allowed to expand. Clones were then induced with doxycycline and scored (live or dead) 48 h after doxycycline administration. Survival of

individual clones was similar to the survival of the heterogeneous cell population (Fig. 1E), indicating that proliferation and selection of targeted cells did not significantly contribute to the overall rescue effect.

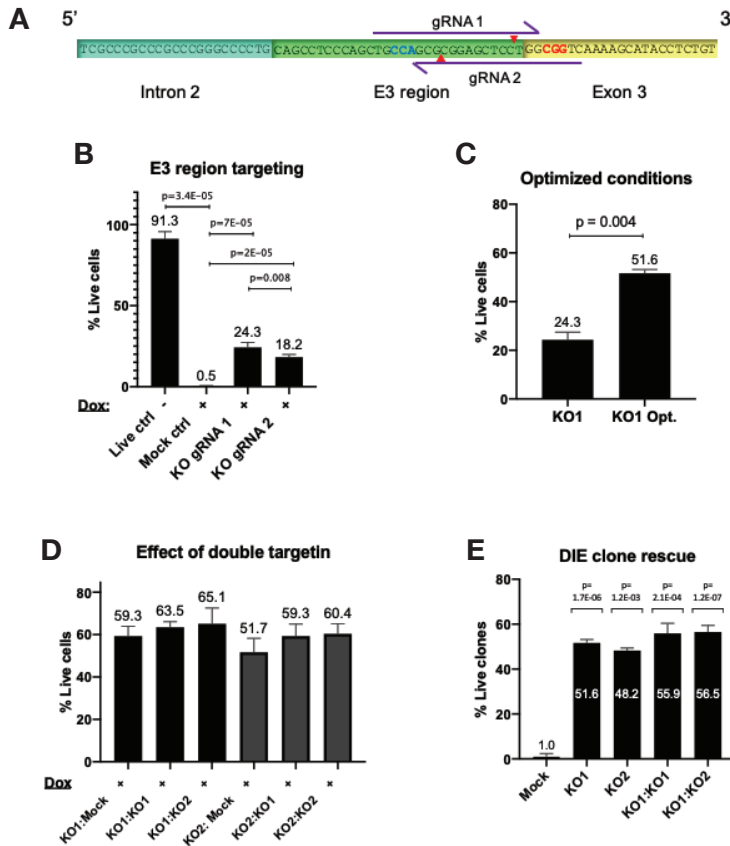


Figure 1. Silencing DUX4 with CRISPR/Cas9. (A) Schematic representation of the E3 region and context sequences. Purple half arrows indicate gRNA sequences, their location and orientation at the E3 site, with the red triangles representing the Cas9 cut sites. Red lettering is the PAM sequence of gRNA1, and blue lettering is the complementary sequence of the gRNA2 PAM, located on the anti-sense strand not shown in this figure. (B) FACS analysis of doxycycline uninduced (-) and induced (+) DIE cells. Live ctrl: DIE cells that have not been transduced with CRISPR/CAS9 or exposed to doxycycline. Mock ctrl: DIE cells transduced with only spCas9 protein. KO gRNA1/2: DIE cells transduced with spCas9 protein and DUX4 gRNA1 or gRNA2. Statistical significance of FACS data was determined by a two-tailed Student t-test. (C) Rescue efficiency of DIE cells with conventional targeting (KO1), and optimized targeting (KO opt.). Statistical significance of FACS data was determined by a one-tailed Student t-test. (D) FACS analyzed data including double targeted E3 sequence using optimized conditions. Mock: DIE cells were transduced with 4xSV40 NLS-spCas9-2xSV40 NLS protein only. KO1: DIE cells were transduced with 4xSV40 NLS-spCas9-2xSV40 NLS protein and gRNA1. KO2: DIE cells were transduced with 4xSV40 NLS-spCas9-2xSV40 NLS protein and gRNA2. (E) Rescue percentage single and double targeted DIE clones. All surviving clones were counted after 48h of doxycycline exposure. A significant increase in survival can be seen when DIE cells were transduced with 4xSV40 NLS-spCas9-2xSV40 NLS and a DUX4 specific gRNA (KO1, KO2, KO1:KO1, KO2:KO2), compared to cells that were only transduced with 4xSV40 NLS-spCas9-2xSV40 NLS protein (Mock). Statistical significance was determined by a two-tailed Student t-test. gRNA: guide RNA, DIE: DUX4 induced expression, E3: exon3 antecedent sequence, FACS: fluorescence-activated cell sorting, KO: knock-out, NLS: nuclear localization signal, PAM: protospacer adjacent motif.

Next, we examined the type of indel that was able to eliminate DUX4 expression using TIDE analysis (Tracking of Indels by Decomposition)³². Targeting the E3 sequence with gRNA1 primarily causes an insertion of one nucleotide (40.1% ± 10%, p-value = 0). Insertions of more than one nucleotide were not detected. The wildtype sequence was found at a frequency of 35.8% ± 10.4%. The remaining 3.7-44.5% consisted of deletions of different sizes, however, none reached a frequency higher than 5% (Fig. 2A). To discern which indels are responsible for the rescue, the same population of targeted DIE cells were exposed to doxycycline for a period of 24 h to obtain an enriched population of rescued cells that were subsequently also analyzed for their indel frequency. All indels previously detected were still visible in these rescued cells (Fig. 2B), suggesting that all indels presented in these graphs can contribute to silencing the DUX4 gene. Sequencing data further showed that the inserted nucleotide consisted of a cytosine in 83.4-88.2% of all gRNA1 targeted samples (Fig. 2C). This cytosine insertion can be found directly to the left side of the break (Fig. 2D, bottom panel).

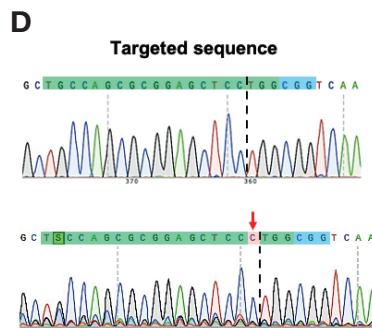
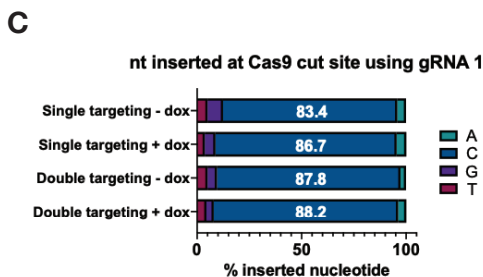
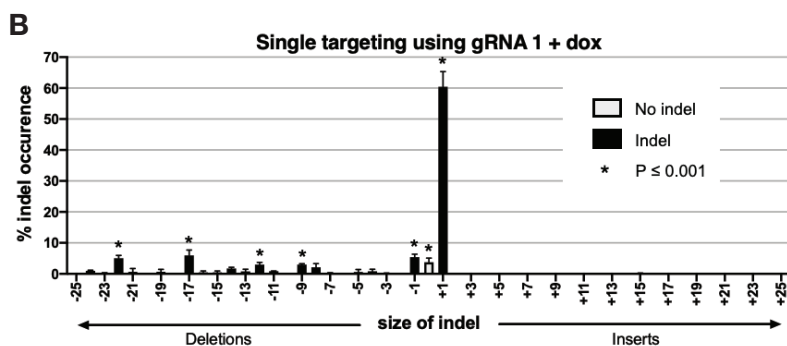
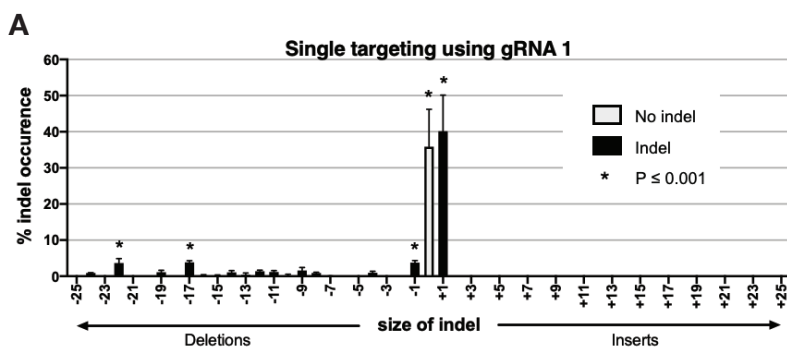


Figure 2. Type and frequency of CRISPR/Cas9-induced indel at the E3 site when targeted with gRNA1. (A & B) The percentage of inserted or deleted nucleotides found at the cut site of the gRNA1 targeted E3 region. **(A)** In DIE cells that were not exposed to doxycycline, gRNA1 demonstrates a high tendency of a one nucleotide insertion ($40.1\% \pm 10\%$, $p\text{-value} = 0$). The wildtype sequence can be found at a frequency of $35.8\% \pm 10.4\%$ ($p\text{-value} = 0$). Deletions of 1, 17 and 22 nucleotides were also detected in a significant amount (-1nt: $3.8\% \pm 0.6\%$, $p\text{-value} = 4.3\text{E-}08$; -17nt: $3.8\% \pm 0.5\%$, $p\text{-value} = 9.1\text{E-}07$; -22nt: $3.6\% \pm 1.3\%$, $p\text{-value} = 0.0002$). **(B)** sequencing data of gRNA1 targeted DIE cells that were exposed to doxycycline (+1nt: $60.4\% \pm 4.9\%$, $p\text{-value} = 0$; 0nt: $3.7\% \pm 1.3\%$, $p\text{-value} = 1.8\text{E-}04$; -1nt: $5.4\% \pm 1\%$, $p\text{-value} = 1.7\text{E-}10$; -9nt: $3\% \pm 0.3\%$, $p\text{-value} = 3.7\text{E-}05$; -12nt: $3\% \pm 0.7\%$, $p\text{-value} = 1.2\text{E-}04$; -17nt: $5.9\% \pm 1.8\%$, $p\text{-value} = 3.7\text{E-}10$; -22nt: $5\% \pm 0.9\%$, $p\text{-value} = 6.9\text{E-}12$). **(C)** The nucleotide inserted when the E3 sequence is targeted with gRNA1 is predominantly a cytosine (Single targeting – dox: $83.4\% \pm 2.7\%$, $p\text{-value} = 1.7\text{E-}05$; Single targeting + dox: $86.7\% \pm 2.6\%$, $p\text{-value} = 4\text{E-}05$; Double targeting – dox: $87.8\% \pm 7.7\%$, $p\text{-value} = 0.001$; Double targeting + dox: $88.2\% \pm 2.4\%$, $p\text{-value} = 1.5\text{E-}05$). **(D)** Sanger sequencing data demonstrating the cytosine insertion at the cleavage site. Upper panel shows the wild type situation. Cut site is indicated with a black intermitted line. The spacer sequence is highlighted in green, and the PAM is highlighted in blue. The one nucleotide cytosine insertion when targeting the E3 site with gRNA1 is highlighted in red and can be found directly to the left of the cut site. gRNA: guide RNA, DIE: DUX4 induced expression, E3: exon3 antecedent sequence, PAM: protospacer adjacent motif.

Targeting E3 with gRNA2 on the other hand, mainly resulted in deletions, the most frequent being two, seven and fourteen nucleotides (Fig 3A and 3B). These deletions can be seen to the right side of the double-stranded break (Fig. 3C, right panel). The wildtype sequence can be found at a frequency of $58.6\% \pm 1.8\%$ (Fig. 3A), suggesting that gRNA2 is less efficient than gRNA1 in its genome-editing capacity. These results corroborate previous results (Fig. 1B) that gRNA2 is less effective in rescuing DUX4-induced apoptosis than gRNA1.

Results shown here also affirm that DNA repair is not random. The tendency to produce a specific type of indel at a particular target region has previously been shown to be highly reproducible, non-random and dependent on local sequence context at the break site^{33–35}. The type of indels that were generated by the gRNAs were the same in several independent experiments and are thus likely target region dependent.

Upon closer examination of the E3 region by RBPmap³⁶, motifs of different RNA binding proteins can be found (Fig. 4). These RNA binding proteins are known to play a role in RNA splicing and mRNA processing^{37–41}, which suggests the importance of this region in regulating the stability of the DUX4 transcript.

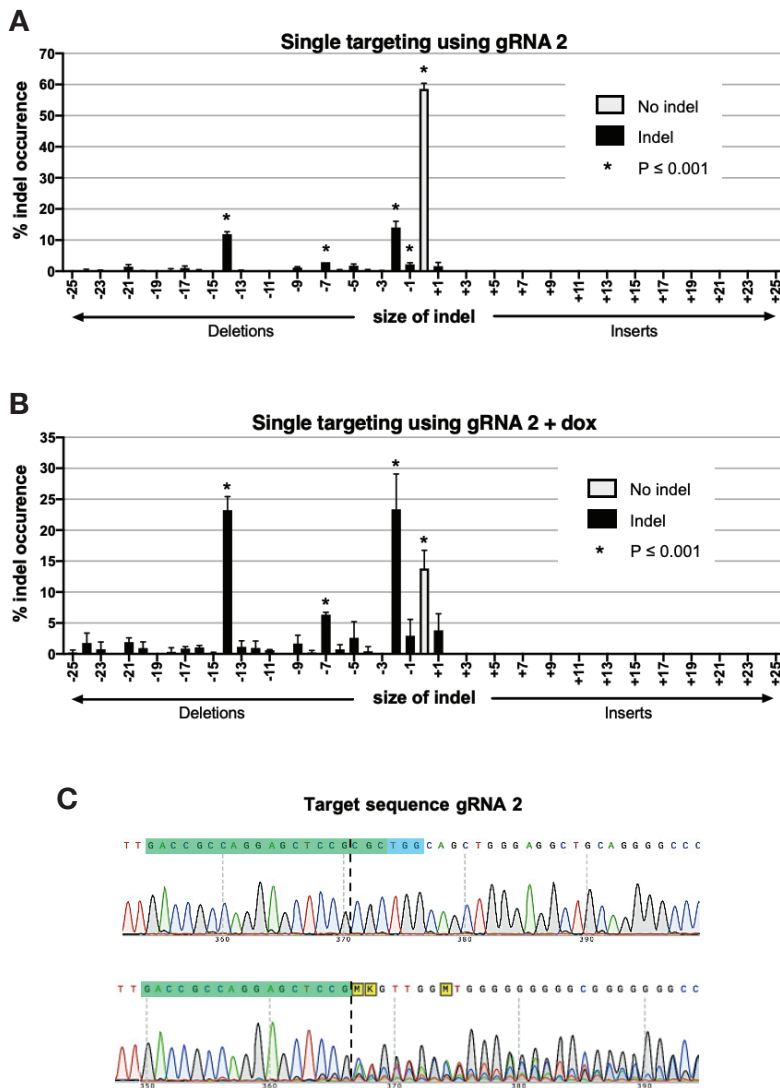


Figure 3. Type and frequency of CRISPR/Cas9-induced indel at the E3 site when targeted with gRNA2. (A) Non-doxycycline exposed DIE cells targeted with gRNA2 show a higher tendency towards deletion. The most frequent deletions are deletions of two and fourteen nucleotides (2nt: $14\% \pm 2\%$, p-value = $1.5E-83$; 14nt: $11.9\% \pm 0.8\%$, p-value = $1.1E-70$). Two other deletions were found that are less frequent, but still significantly detected among the population (1nt: $2.2\% \pm 0.5\%$, p-value = 0.0003 ; 7nt: $2.9\% \pm 0\%$, p-value = $2E-06$). The wildtype sequence can be found at a frequency of $58.6\% \pm 1.8\%$ (p-value = 0). (B) Sequencing results of doxycycline treated DIE cells targeted with gRNA2 (0nt: $13.8\% \pm 3\%$, p-value = $9.7E-36$; -2nt: $23.4\% \pm 5.7\%$, p-value = $3.3E-49$; -7nt: $6.3\% \pm 0.4\%$, p-value = $5.3E-08$; -14nt: $23.2\% \pm 2.2\%$, p-value = $1.5E-69$). (C) The deletions when targeting the E3 site with gRNA2 can be seen in the lower panel, where a deletion has occurred directly to the right of the cut site, thereby also removing the last three nucleotides of the gRNA sequence and the PAM sequence. DIE: DUX4 induced expression, nt: nucleotide, E3: exon3 antecedent sequence, gRNA: guide RNA, PAM: protospacer adjacent motif.

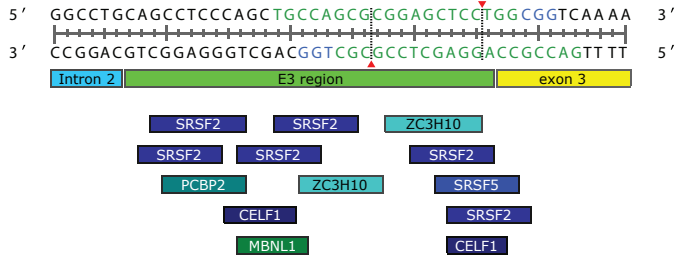


Figure 4. RNA binding protein motifs at the E3 region. Target sequence is indicated in green lettering, and the PAM sequence in blue lettering. gRNA1 targets the green sequence on the leading strand (top), and gRNA2 targets the lagging strand (bottom). Red triangle and intermitted line indicate cleavage sites from the Cas9 protein. RNA binding proteins are annotated below the DNA sequence, their location corresponding to their binding motif on the DNA. E3: exon3 antecedent sequence, gRNA: guide RNA, PAM: protospacer adjacent motif.

Discussion

In the search for a treatment for FSHD, recently developed genome-editing technologies offer interesting new possibilities to permanently shut down DUX4 expression in affected tissues. However, the repetitive nature of DUX4, the disease-causing gene in FSHD, makes it challenging to identify suitable gRNAs that specifically target the disease causing DUX4 open reading frame (ORF). We therefore designed a CRISPR/Cas9 approach to silence DUX4 without targeting the ORF. Because pathological DUX4 expression needs a stable transcript, we decided to target a site upstream of the polyadenylation sequence that is relatively 'unique', and could be important for the stabilization or processing of the DUX4 transcript. This site differs in two nucleotides from sequences found in the preceding repeats and from D4Z4 repeats at other places in the genome²⁶. Any indel created in this region could potentially destabilize the DUX4 transcript, by interfering with regulatory functions. Targeting the E3 sequence with an in vitro transcription (IVT)-generated guide and traditional spCas9 protein indeed showed some rescue in DIE cells upon doxycycline exposure, with an efficiency of approximately 24%. Because skeletal muscle fibers are multinucleated, it takes only a few DUX4-expressing nuclei to deteriorate the entire muscle fiber^{5,42}. Therefore, a high knock-out efficiency would be required to provide therapeutic efficacy. By using optimized recombinant CRISPR/Cas9 components (synthetic guides, adding additional NLSs to Cas9), the knock-out efficiency significantly increased from ~24% to ~51%. Multiple consecutive targeting's only marginally increased cell survival, despite the fact that these double-targeted DIE cells demonstrate near-90% Indel formation at the CRISPR/Cas9 target site (Fig. S1). One explanation could be that the mutations introduced at the E3 region are not as potent in functionally disrupting DUX4 expression as, for example, disruptions of the DUX4 ORF itself. Indels at the E3 site likely destabilize the DUX4 transcript, but this can also depend on other factors such as cell cycle state⁴³⁻⁴⁶ and cellular stress^{47,48}. RNA destabilization is therefore not a black or white event and the disruption of the DUX4 (pre-)mRNA stability by targeting the E3 locus is thus not sufficient for full elimination of all DUX4 transcript from cells, as depicted by the 'rescue-cap' illustrated in Figure 5. Destabilizing the DUX4 transcripts by targeting the 3'UTR may therefore not be enough to provide therapeutic benefit.

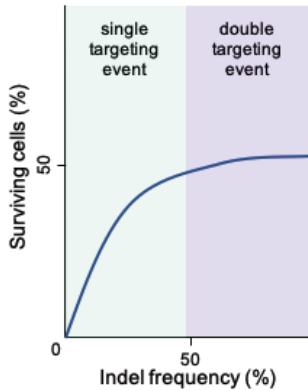


Figure 5. Schematic representation of the ‘rescue cap’ seen in DIE cells targeted at the E3 region, upon DUX4 induction. When targeting the E3 region, the frequency of indels at this region, that functionally disrupts DUX4 stability, determines the likelihood of cell survival upon DUX4 induction. A higher indel frequency beyond a certain point results in minimal effect on cell survival upon DUX4 induction. This suggests that the ability to disrupt the E3 region to destabilize DUX4 expression is limited, and maximal destabilization is almost achieved after a single round of targeting.

Analyzing the target sequence in single-targeted and double-targeted DIE cells revealed the type and frequency of indels that were generated at the E3 site. It is known that the indels generated by CRISPR/Cas9 targeting are non-random^{33–35}. Based on local sequence context of the genomic ends flanking the cleavage site, the type of indel is highly reproducible and predictable. This has led to the development of various algorithms that can predict, to a degree of certainty, the type of indel that will be produced at a specific cleavage site^{49–51}. In this particular project, targeting with gRNA1 introduced a single cytosine directly to the left of the cleavage site at the highest frequency. Interestingly, the InDelphi algorithm⁴⁹ correctly predicted that no more than one nucleotide would be inserted at this site when targeting with this gRNA. The InDelphi model also correctly predicted that this insertion would most likely be a cytosine. However, the algorithm predicted that a deletion of 12 nucleotides would be the most likely mutation to occur upon targeting, which was not the case. When targeting with the other gRNA (gRNA2), the most commonly found indels were two, seven and fourteen nucleotide deletions, which were indeed also the top three predictions with InDelphi. Although these algorithms can be extremely useful when generating a shortlist of guides for when a specific indel is required, experimental testing of gRNAs should not be omitted. In particular, since InDel size does not seem to be a good predictor of the functional effect on cell survival when targeted to other genomic regions than ORFs. Without additional context, it is perhaps surprising that a single cytosine addition in the 3’UTR of the DUX4 transcript, achieved by sgRNA1 targeting, is sufficient to achieve 50% cell survival as compared to control.

Taken together, we demonstrated here that a single base insertion at a specific intronic site can disrupt DUX4 expression greatly, despite the fact that this manipulation did not occur in the ORF of the DUX4 gene, nor a splice site for RNA splicing events. This data suggest that the E3 is important for stabilization of the pre-mRNA. This region can potentially function as a site for RNA binding proteins and thus stabilize pre-mRNA. It can also be important for the secondary or tertiary structure of the RNA, necessary for correct splicing events. Potentially it can be important for both, as structure of the RNA, and binding of RNA binding proteins are interconnected^{52,53}. Thus, by modifying this region, this extra layer of RNA stability can be lost, reducing the probability of translation, but not completely abolishing it.

Methods

Cell culturing and seeding

DUX4 inducible expression (DIE) cells were cultured in IMDM media with 10% Tet system approved FBS (Clonotech), 100 μM 2-mercapto-ethanol, 5 $\mu\text{g}/\text{ml}$ Puromycin and 6 $\mu\text{g}/\text{ml}$ Blastidicin. Cells were kept at 5% CO_2 and 37°C.

Producing IV-RT guide

For the production of single guide (sgRNA), a single stranded DNA template (supplied by IDT, California, USA) was amplified with Taq DNA polymerase by PCR. The template encodes the T7 promotor with an additional guanine at the end, a 20 nt variable spacer sequence, and the spCas9 tracr sequence with a polyT signal: 5'-TAATACGACTCACTATAGG-20nt-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGCACCGAGTCGGTGCCTTTTT-3'. The following primer set was used to amplify the double-stranded DNA template and add small adapter sequences (underlined) for improved binding of DNA and RNA polymerases: 5'-ggcactcTAATACGACTCACTATAGG -3' and 5'-cgagcgcAAAAAAGCACCGACTC-3'. The PCR product was purified using a PCR purification kit (Qiagen), and diluted to 250 ng/ μl . sgRNA was produced by reverse transcription (RT) using an IV-RT kit from NTRANS Technologies (Utrecht, The Netherlands), according to manufacturer instructions. Remaining DNA was removed by the addition of 2U of Turbo DNase to each 20 μl IV-RT reaction. Six IV-RT reactions were pooled and sgRNA was purified by phenol-chloroform extraction. The dried RNA pellet was dissolved in nuclease-free water and diluted to 10 $\mu\text{g}/\mu\text{l}$. sgRNA samples were used immediately or stored at -80°C for a maximum of 30 days.

Producing spCas9 and modified spCas9

Recombinant proteins were expressed in BL21 (DE3) *E. coli* and soluble fractions were extracted as previously described by D'Astolfo et al.³⁰. A 20°C induction temperature was used for overnight induction. The soluble fraction containing the recombinant protein was degassed and filtered through a 22 μm filter before loading it onto a HisTrap high performance column (GE Healthcare), using an AKTA Pure FPLC_{v2.0} system (GE Healthcare). The soluble fraction was loaded at a speed of 0.6 ml/min for a 1 ml column, or 3 ml/min for a 5 ml column. Protein was eluted from the HisTrap column by using increasing amounts of imidazole, and fractions of each measured peak were collected. Correct proteins fractions were confirmed by loading a sample of each fraction onto an SDS PAGE gel and staining with Coomassie blue. The fraction that contained protein of the right size was then purified and the buffer was exchanged (into 5x transduction buffer³⁰) using HiLoad Superdex 200 pg preparative SEC columns. Protein was concentrated to 75 μM using an Amicon ultra centrifugal filter unit with a 100kDa cutoff (Merck). Protein aliquots were used immediately or snap frozen in liquid nitrogen and stored at -80°C.

Transduction CRISPR/Cas9 into DIE cells

96-well tissue culture plates were coated with a Matrigel (Matrigel in PBS, 1:250). Subsequently, 15,000 DIE cells were seeded on top of Matrigel-coated wells and incubated overnight at 5% CO_2 and 37°C, until 70-80% confluency was reached. The cells were transduced with spCas9 or 4xSV40-spCas9-2xSV40 protein and sgRNAs targeting the E3

sequence using the iTOP transduction method³⁰. After a recovery period of a minimum of 24 h, cells were exposed to a high concentration of doxycycline (1000 ng/ml) for 16-48 h, depending of the type of experiment. DIE cells (dead and alive) were collected and stained with Annexin-5 FITC and DAPI and subsequently analyzed by FACS.

Flowcytometry sorting (FACS) and analysis and dead live staining

DIE cells were treated with doxycycline for 16 h prior to FACS analysis. After doxycycline exposure, the culture media was collected, as was the DBPS wash that followed, to collect all dying and detached cells. The remaining cells were trypsinized using 0.25% Trypsin-EDTA and resuspended in culture media. The trypsinized cells were added to the previously collected sample of detached cells and pelleted by centrifugation (500 g for 10 min). The supernatant was removed and the cell pellet resuspended in DPBS with 5% FBS, supplemented with annexin-V FITC. Cells were left to incubated for 15-20 min at 4 °C in a dark environment. iMDM media with 10% Tet approved FBS and DAPI nuclear staining was subsequently added and cells were strained using Cell-strainer capped tubes (Falcon) and analyzed using the BD FACSCanto II flow cytometer.

Sample preparation and indel analysis

Transduced DIE cells were harvested, pelleted and frozen at -20°C. Genomic DNA was extracted and purified using the DNeasy Blood & Tissue kit (Qiagen). A 558 basepair fragment was amplified from genomic DNA samples using high fidelity Phusion polymerase (ThermoFisher), the Phusion GC 5x buffer and the following primers flanking the E3 cleavage site: 5'- AAACGCGTCGTCCTG-3' and 5'- GCCAGAGGCCACTTGTGTAG-3'. A PCR program of 35 cycles consisted of denaturation at 98 °C, 60 seconds; annealing at 68 °C, 20 seconds; and elongation at 72 °C, 20 seconds, before visualizing on a 1% TAE agarose gel. The amplified products were gel purified and send for Sanger sequencing (BaseClear). Indel frequency was determined using tracking of indels by decomposition analysis³². Data from three biological replicates were combined and average values (per sample) are displayed.

Reference

1. Choi, S. H. *et al.* DUX4 recruits p300/CBP through its C-terminus and induces global H3K27 acetylation changes. *Nucleic Acids Res.* **44**, 5161–5173 (2016).
2. Vuoristo, S. *et al.* DUX4 regulates oocyte to embryo transition in human. *Biorxiv* (2019) doi:<http://dx.doi.org/10.1101/732289>.
3. Dixit, M. *et al.* DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18157–18162 (2007).
4. Lemmers, R. J. L. F. *et al.* A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science (80-.)*. **329**, 1650–1653 (2010).
5. Snider, L. *et al.* Facioscapulohumeral dystrophy: Incomplete suppression of a retrotransposed gene. *PLoS Genet.* **6**, 1–14 (2010).
6. Hewitt, J. E. *et al.* Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophoathy. *Hum. Mol. Genet.* **3**, 1287–1295 (1994).
7. Winokur, S. T. *et al.* The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: Implications for a role of chromatin structure in the pathogenesis of the disease. *Chromosom. Res.* **2**, 225–234 (1994).
8. Deidda, G. *et al.* Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *European Journal of Human Genetics* vol. 3 155–167 (1995).

9. Van Der Maarel, S. M. & Frants, R. R. The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **76**, 375–386 (2005).
10. Lemmers, R. J. L. F. *et al.* Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nat. Genet.* **44**, 1370–1374 (2012).
11. Larsen, M. *et al.* Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1. *Eur. J. Hum. Genet.* **23**, 808–816 (2015).
12. Van Den Boogaard, M. L. *et al.* Mutations in DNMT3B Modify Epigenetic Repression of the D4Z4 Repeat and the Penetrance of Facioscapulohumeral Dystrophy. *Am. J. Hum. Genet.* **98**, 1020–1029 (2016).
13. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases Non-invasive sexing of preimplantation stage mammalian embryos. *Nat. Am. Inc.* **19**, 219–220 (1998).
14. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
15. Blewitt, M. E. *et al.* An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7629–7634 (2005).
16. Blewitt, M. E. *et al.* SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat. Genet.* **40**, 663–669 (2008).
17. Gendrel, A. V. *et al.* Smchd1-Dependent and -Independent Pathways Determine Developmental Dynamics of CpG Island Methylation on the Inactive X Chromosome. *Dev. Cell* **23**, 265–279 (2012).
18. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
19. Lemmers, R. J. F. L. *et al.* Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* **75**, 1124–1130 (2004).
20. Wang, Z. Q., Wang, N., Van Der Maarel, S., Murong, S. X. & Wu, Z. Y. Distinguishing the 4qA and 4qB variants is essential for the diagnosis of facioscapulohumeral muscular dystrophy in the Chinese population. *Eur. J. Hum. Genet.* **19**, 64–69 (2011).
21. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.)*. **337**, 816–821 (2012).
22. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-.)*. **339**, 819–823 (2013).
23. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 2579–2586 (2012).
24. Anseau, E. *et al.* Antisense oligonucleotides used to target the DUX4 mRNA as therapeutic approaches in facioscapulohumeral muscular dystrophy (FSHD). *Genes (Basel)*. **8**, (2017).
25. Himeda, C. L., Jones, T. I. & Jones, P. L. CRISPR/dCas9-mediated transcriptional inhibition ameliorates the epigenetic dysregulation at D4Z4 and represses DUX4-fl in FSH muscular dystrophy. *Mol. Ther.* **24**, 527–535 (2016).
26. Lemmers, R. J. *et al.* Deep characterization of a common D4Z4 variant identifies biallelic DUX4 expression as a modifier for disease penetrance in FSHD2. *Eur. J. Hum. Genet.* **26**, 94–106 (2018).
27. Wong, N., Liu, W. & Wang, X. WU-CRISPR: Characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.* **16**, 1–8 (2015).
28. Hiranniramol, K. *et al.* Generalizable sgRNA design for improved CRISPR/Cas9 editing efficiency. *Bioinformatics* **36**, 2684–2689 (2020).
29. Hodgkins, A. *et al.* WGE: A CRISPR database for genome engineering. *Bioinformatics* **31**, 3078–3080 (2015).
30. D'Astolfo, D. S. *et al.* Efficient intracellular delivery of native proteins. *Cell* **161**, 674–690 (2015).
31. Staahl, B. T. *et al.* Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.* **35**, 431–434 (2017).
32. Brinkman, E. K., Chen, T., Amendola, M. & Van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, (2014).
33. van Overbeek, M. *et al.* DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-

- Mediated Breaks. *Mol. Cell* **63**, 633–646 (2016).
34. Lemos, B. R. *et al.* CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E2010–E2047 (2018).
 35. Shou, J., Li, J., Liu, Y. & Wu, Q. Precise and Predictable CRISPR Chromosomal Rearrangements Reveal Principles of Cas9-Mediated Nucleotide Insertion. *Mol. Cell* **71**, 498–509.e4 (2018).
 36. Paz, I., Kostı, I., Ares, M., Cline, M. & Mandel-Gutfreund, Y. RBPmap: A web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* **42**, 361–367 (2014).
 37. Sapra, A. K. *et al.* SR Protein Family Members Display Diverse Activities in the Formation of Nascent and Mature mRNPs In Vivo. *Mol. Cell* **34**, 179–190 (2009).
 38. Xia, H. *et al.* CELF1 preferentially binds to exon-intron boundary and regulates alternative splicing in HeLa cells. *BBA - Gene Regul. Mech.* **1860**, 911–921 (2017).
 39. Tripathi, V. *et al.* Direct Regulation of Alternative Splicing by SMAD3 through PCBP1 Is Essential to the Tumor-Promoting Role of TGF- β . *Mol. Cell* **64**, 549–564 (2016).
 40. Ho, T. H. *et al.* Muscleblind proteins regulate alternative splicing. *EMBO J.* **23**, 3103–3112 (2004).
 41. Gaudet, P., Livstone, M. S., Lewis, S. E. & Thomas, P. D. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief. Bioinform.* **12**, 449–462 (2011).
 42. Jones, T. I. *et al.* Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: Evidence for disease modifiers and a quantitative model of pathogenesis. *Hum. Mol. Genet.* **21**, 4419–4430 (2012).
 43. Trcek, T., Larson, D. R., Moldón, A., Query, C. C. & Singer, R. H. Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. *Cell* **147**, 1484–1497 (2011).
 44. Penelova, A., Richman, L., Neupert, B., Simanis, V. & Kühn, L. C. Analysis of the contribution of changes in mRNA stability to the changes in steady-state levels of cyclin mRNA in the mammalian cell cycle. *FEBS J.* **272**, 5217–5229 (2005).
 45. Choudhury, A. D., Xu, H. & Baer, R. Ubiquitination and proteasomal degradation of the BRCA1 tumor suppressor is regulated during cell cycle progression. *J. Biol. Chem.* **279**, 33909–33918 (2004).
 46. Saunus, J. M., Edwards, S. L., French, J. D., Smart, C. E. & Brown, M. A. Regulation of BRCA1 messenger RNA stability in human epithelial cell lines and during cell cycle progression. *FEBS Lett.* **581**, 3435–3442 (2007).
 47. Fan, J. *et al.* Global analysis of stress-regulated mRNA turnover by using cDNA arrays. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10611–10616 (2002).
 48. Lü, X., De La Peña, L., Barker, C., Camphausen, K. & Tofilon, P. J. Radiation-induced changes in gene expression involve recruitment of existing messenger RNAs to and away from polysomes. *Cancer Res.* **66**, 1052–1061 (2006).
 49. Shen, M. W. *et al.* Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **563**, 646–651 (2018).
 50. Allen, F. *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**, 64–82 (2019).
 51. Chen, W. *et al.* Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res.* **47**, 7989–8003 (2019).
 52. Soemedia, R. *et al.* The Effects of Structure on pre-mRNA Processing and Stability. *Methods* **125**, 36–44 (2017).
 53. Boo, S. H. & Kim, Y. K. The emerging role of RNA modifications in the regulation of mRNA stability. *Exp. Mol. Med.* **52**, 400–408 (2020).

Supplementary material

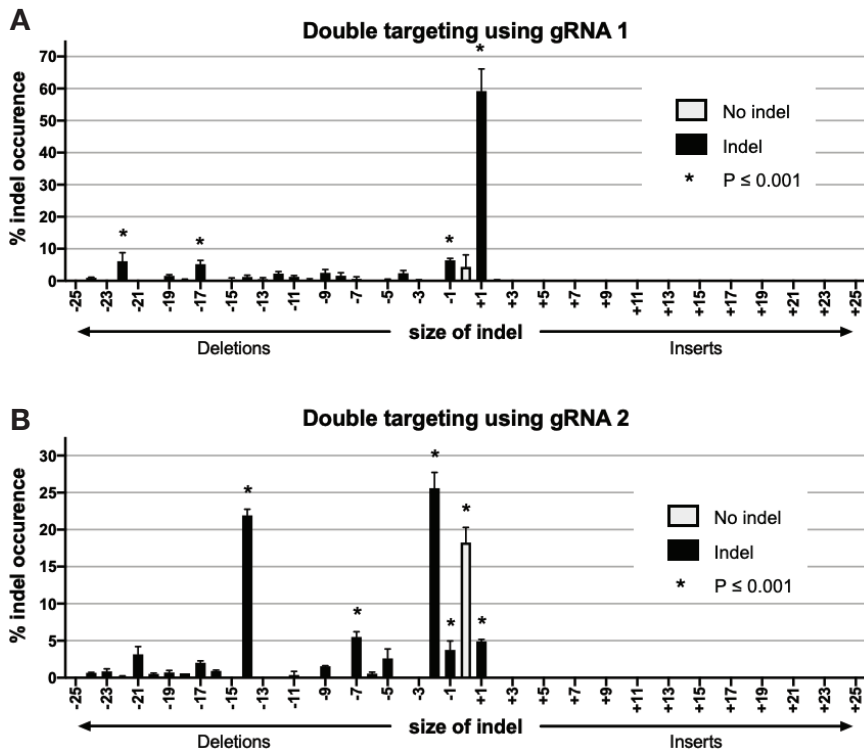


Figure S1. Indel frequency of double-targeted DIE cells not exposed to doxycycline. (A) Double targeting of the E3 region with gRNA1. The wildtype sequence can be found at a frequency of $4.3\% \pm 3.8\%$ (p -value = 0.007). A one nucleotide insertion has a frequency of $59.2\% \pm 6.9\%$ (p -value = 0). Three deletions of one, seventeen and twenty-two nucleotides are significantly present at a frequency of $6.4\% \pm 0.6\%$ (p -value = $5.7E-17$), $5.2\% \pm 1.2\%$ (p -value = $5.7E-06$), and $6.1\% \pm 2.7\%$ (p -value = $4.3E-11$), respectively. **(B)** Double targeting of the E3 region with gRNA2. The wildtype sequence is present at a frequency of $18.3\% \pm 2.1\%$ (p -value = $9E-09$). The one nucleotide insertion is present at a frequency of $4.9\% \pm 0.3\%$ (p -value = $3.6E-08$). Deletions of 1, 3, 7 and 14 nucleotides were significantly present in the DIE cells at frequencies of $3.8\% \pm 1.2\%$ (p -value = $4.6E-05$), $25.6\% \pm 2.1\%$ (p -value = $2.9E-212$), $5.5\% \pm 0.7\%$ (p -value $3.4E-12$), and $21.9\% \pm 0.8\%$ (p -value = $6.5E-134$), respectively.



Illustration based on a stone carving on display at the MET

Chapter 6

General discussion

It has been over a 136 years since Facioscapulohumeral muscular dystrophy (FSHD) was first described¹ and still we continue to search for treatment options for this disorder. As we unravel its underlying mechanisms, we move closer to solving more pieces of this very large and complex puzzle. For the past few decades, researchers have narrowed their focus to the transcription factor DUX4, as its misexpression lies at the center of the pathophysiology seen in FSHD patients²⁻⁴. DUX4 is part of a repeat sequence which can be found at multiple loci in the human genome. This complicates the more obvious and widely used targeting strategy, CRISPR/Cas9-mediated knock out. CRISPR/Cas9 is a genome editing technique, derived from the prokaryotic adaptive immune response⁵⁻⁸. CRISPR/Cas9 uses a guide RNA to navigate the Cas9 endonuclease to a specific site in the host its DNA, where it induces a double stranded break. Upon repair, errors can be introduced at this cleavage site, which can render a gene non-functional^{9,10}. This new genome editing strategy has revolutionized the genome editing field, and has quickly become the most widely used strategy to knock-out genes. As DUX4 is part of a repeated sequence, gRNAs that target the body of the DUX4 gene will cause multiple double stranded breaks throughout the human genome, risking further contraction and or translocation events. Some have therefore focused their efforts on modulating the expression or activity of DUX4-linked genes. These include genes that play a role in the expression of DUX4 itself such as p38 MAPK^{11,12}, epigenetic regulators¹³, and potential transcription factors^{14,15}, and genes and pathways that are regulated by DUX4 and contribute to its cytotoxic effect such as the MYC-mediated apoptotic pathway, the dsRNA innate immune response pathways¹⁶, and genes involved in the hypoxia-related HIF1 pathway^{17,18}. However, other new promising avenues for future therapeutic intervention do aim to target DUX4 expression directly, without introducing any DNA double stranded breaks. For example, the strategy of using a nuclease-dead Cas9 fused to a Krüppel-associated box (dCas9-KRAB) inhibits DUX4 transcription¹⁹, or the anti-sense oligonucleotides strategy that inhibits translation²⁰⁻²². Direct intervention at the source will naturally be an efficient and promising strategy that may save both the cells and their biological function, but these approaches will likely be transient unless a permanent approach is used, such as a gene therapy. Inhibiting epigenetic regulators, such as chromatin remodelers, epigenetic readers, methyltransferases that add activation marks, and demethylases that remove repressive marks, may also reduce expression of DUX4¹³, and can therefore also be explored as a potential therapeutic treatment for FSHD.

6

Recently, an increase in hyaluronic acid (HA) has been associated with FSHD. Accumulation of HA occurs after DUX4 expression and the inhibition of HA biosynthesis prevented FSHD-related pathologies, such as RNA granule formation, FUS (fused in sarcoma) protein aggregation, DNA damage, caspase activation, and apoptosis. The exact role of HA in FSHD pathology is unclear, but the involvement of Complement component 1 Q subcomponent-binding protein (C1QBP) and mitochondria is considered²³. Interestingly, HA inhibition has limited effect on DUX4 expression and a partial effect on the DUX4 induced transcriptional program. This could suggest that HA works relatively independent of the DUX4-induced transcriptional program, or at least it effects only a part and specific aspect of this program. Our data demonstrates that DUX4 activates the expression of a network of downstream transcriptional regulators, which seem to reprogram cells into a more stem-cell like state before pushing the cells into apoptosis (**Chapters 2 and 3**). While HA inhibition can rescue cells from DUX4-induced apoptosis, the continued dysregulation of the transcriptional program may still trigger a loss of cell identity, which may prove similarly detrimental to the

muscle fiber and or its function.

Currently, a single drug losmapimod, a p38/MAPK inhibitor, has entered phase 2 clinical trials for the treatment of FSHD. The p38 pathway has been identified as an activator of DUX4 expression, and inhibition of the p38/MAPK pathway interfered with DUX4 expression itself and prevented DUX4-induced cell loss^{11,12}. The exact molecular mechanism involved in the regulation of DUX4 by p38 is as of yet unknown, but p38/MAPK inhibitors have been shown to lower DUX4 and DUX4 target gene expression, both *in vitro* in FSHD patient-derived myoblasts^{11,12} as well as *in vivo* in a humanized mouse xenograft model¹¹.

Although all these advancements in the treatment of FSHD are very promising, it should be noted that systemic administration of drugs targeting multipurpose factors such as kinases, transcription factors and epigenetic regulators can result in undesirable side effects. These side effects can outweigh the potential benefit of the FSHD treatment, rendering it unsuitable. The unraveling of the underlying FSHD pathophysiology and the search for specific FSHD treatment options should therefore continue until safe and perhaps universal treatment options have been developed.

This thesis describes several ways that we explored the possibilities of mitigating the DUX4 cytotoxic effect. To achieve this goal, we built a human *in-vitro* cell model system in which we could induce and regulate the expression of DUX4, and which demonstrated a clear apoptotic phenotype upon induction. This model furthermore contains the endogenous DUX4 coding sequence, so specific targeting strategies could also be tested. Together with the cell line's highly proliferative nature, simple maintenance requirements, high transfectability/transduceability, and its robust induction phenotype made this cell model system versatile and multipurpose. The DUX4 inducible expression (DIE) model validated the pioneer qualities of DUX4, after performing RNA sequencing to explore the transcriptional events that follow DUX4 induction. With the DIE cell line, we aimed to uncover players in the DUX4-induced signaling cascade. DUX4 induction in our DIE cell model shows a high degree of similarity with its transcriptome to other FSHD models and FSHD-affected muscle cells²⁴⁻²⁷ (**Chapter 2**). The DIE cell system was therefore used to find genes involved in the DUX4-induced pathways that mediate the toxicity of DUX4. Initial bulk sequencing demonstrated that many of the early differentially upregulated genes found after only 4.5h hours of doxycycline induction are germline or stemness genes, or genes related to early embryonic development. These results confirm the developmental role of DUX4²⁸⁻³⁰, which have also been found by others that have studied the DUX4-induced transcriptome²⁴⁻²⁷. One assumption was that the induction of an early developmental stage in mature somatic cells would create contradictory signals within the cells, that might cause the cell to enter apoptosis. As these genes were differentially expressed after only 4.5 hours of DUX4 induction, we hypothesized that these genes are targets of DUX4 that are activated very early on in the toxic process. Logic dictates that intervening early in this toxic cascade would show a greater impact on reducing the toxic effects, rather than intervening later in the process when this cascade has already triggered the activation of many downstream pathways. The contribution of these early DUX4 targets to the cytotoxic cascade was therefore tested by individual knock out experiments. These yielded no viable hits, as none of the tested genes rescued or even slowed down the apoptotic phenotype upon their elimination and DUX4 induction.

We thus continued our search of finding major players in FSHD. To explore the dynamics of DUX4-induced cytotoxicity in more detail, we performed single cell RNA sequencing (SCS) on DIE cells induced for short and multiple consecutive time periods of 2, 3, 4 and 6 hours (**Chapter 3**). Performing dimensionality reduction on the single cell data revealed one large cell population, in which the cells orientated themselves on the y-axis of a t-SNE map, based on their induction status. The lack of well-defined clusters suggests that the induced transcriptomes in these cells are very similar, and that DUX4 activates the same program in most, if not all, induced cells. Differential expression analysis between the induced clusters and the uninduced clusters reveals lists of differentially expressed genes, many of which are shared between induction states. This indeed corroborating the notion that DUX4 activates the same cascade of events in most cells. This cascade of events starts with the activation of early developmental processes, quickly followed by a large variety of other cellular processes, and eventually leading to the activation of apoptotic processes. Significant changes in the cell's transcriptome can be seen as early as 2h post DUX4 induction, with most of the genes (94%) remaining differentially expressed at later timepoints. Approximately 33% of the differentially upregulated genes were transcription factors, some of which left an obvious signature expression profile. The 'footprint' expression profiles of transcription factors that were themselves not identified in the single-cell sequencing data were also identified. This suggests the involvement of "elusive" transcription factors, comparable with DUX4 itself, who's expression was too low and/or transient following DUX4 induction to be detected with SCS. These elusive transcription factors could potentially be of importance in the DUX4-induced cytotoxic cascade and could therefore be of interest to be studied more in depth in the context of FSHD. SCS analysis furthermore revealed a number of different expression profiles, some demonstrating an oscillating pattern during the course of induction, suggesting that DUX4 activates a complex and dynamic process. Further investigation into this dynamic process could potentially give more insight into the molecular workings of DUX4-induced cytotoxicity and apoptosis.

With both RNA sequencing experiments (bulk and SCS), it was extraordinary to see such robust and reproducible transcriptional changes in cells that had been induced for a relative short period of time (2-6 hours). This is, to the best of our knowledge, the earliest timepoints in which changes in the transcriptome of DUX4 affected cells were studied, at such a high resolution. This revealed a list of potential early target genes of DUX4 that hadn't been identified previous, or had been but were not necessarily classified as early target genes.

Next, we used our DIE cell system to try and identify modulators of DUX4 cytotoxicity by performing a genome-wide CRISPR/Cas9 knockout screen. The goal was to identify factors that could mitigate DUX4-induced toxicity. We were able to screen for such modulators based on their ability to rescue the apoptotic phenotype upon their knockout (**Chapter 4**). If any of the differentially upregulated genes induced by DUX4 are indeed playing a role in FSHD pathophysiology, or any other genes play an active role in the DUX4 induced cytotoxicity, we would expect to find them back in the genome-wide screen data. However, we did not find such modulators in this particular screen. This suggests that no single gene, other than DUX4 itself, when knocked out, can rescue DUX4 cytotoxicity.

A more direct and permanent approach of reducing DUX4 expression is to knock out the DUX4 gene directly, but since DUX4 is part of a repeated sequence, this approach is challenging. We set up a DUX4 knock out strategy that targets a relatively unique intronic sequence directly adjacent to exon three (E3)³¹(**Chapter 5**). We hypothesized that targeting

this E3 sequence could lead to the disruption of a regulatory region needed for pre-mRNA stabilization or processing, thereby destabilizing the DUX4 mRNA transcript. Indeed, with the use of optimized CRISPR/Cas9 tools, a functional knockout efficiency of ~50% was reached. However, due to the multinucleated nature of muscle fibers, and the stochastic burst-like expression of DUX4 in myonuclei, a 50% efficiency in functional depletion of DUX4 expression is likely not enough to show a significant effect on relieving FSHD symptoms. The study did show that the editing efficiency is likely much higher than 50%, as the wildtype sequence falls below 20% upon a second targeting experiment. This would suggest that not all edits at this site resulted in an efficient knockout of the gene. If so, a targeting strategy that does not rely on indel occurrence in an intron region might be more beneficial and could therefore result in a higher knockout efficiency.

New evolutions of the CRISPR/Cas9 gene editing system offer hope of a genome editing therapy that is both safe and effective. Himeda et al. has demonstrated that targeting a dCas9-KRAB epigenetic silencer to the DUX4 repeat sequence can effectively inhibit DUX4 expression without the danger of introducing multiple double-strand breaks¹⁹. However, as this approach relies on the temporary binding of the dCas9-KRAB to the regulatory region of DUX4, and not on permanently altering the coding sequence, this inhibition of DUX4 will be of a transient nature.

The recent development of CRISPR/Cas9-based base editors^{32,33} and the prime editing system³⁴ does allow the introduction of subtle changes to the DUX4 coding sequence, which can introduce non-sense mutations, thus disrupting the translational reading frame. Base-editing technologies employ a Cas9 nickase (nCas9) that is fused to nucleobase deaminase enzyme^{32,33}. These base-editing fusion proteins can facilitate the conversion of a C•G to T•A bases or T•A to G•C bases, depending on the deaminase enzyme. Since this technology does not introduce double-strand breaks, it can be targeted toward the DUX4 coding region, introducing a nonsense mutation. This would disrupt the DUX4 translational reading frame by introducing a premature translational stop (Fig. 1), resulting in a severely truncated protein that would lack both its homeodomains and its functional domain. This truncated protein will most likely render DUX4 non-functional. Furthermore, as base editing uses a nCas9, it does not introduce double stranded breaks like Cas9 does, and the risk of further contracting the D4Z4 repeat array, or cause translocation events is therefore limited.

In conclusion, we have found that DUX4 homogeneously induces a network of transcription factors that quickly triggers a cascade of transcriptional events, which are ultimately detrimental for cells.

Knocking out individual downstream target genes of DUX4 or even performing a genome-wide knockout screen did not identify individual factors that can mitigate DUX4 cytotoxicity, suggesting that either a multifactorial approach is needed or, more likely, that the only way to effectively eliminate DUX4 cytotoxicity is by eliminating DUX4 activity itself. After thoroughly studying the effects of pioneer transcription factor DUX4 in human cells, and concluding that no other individual factors plays a large enough role in the DUX4 induced cytotoxic cascade, we have come to the conclusion that the best way forward in finding treatments options for FSHD lies in targeting DUX4 directly. Our efforts of doing so by targeting the E3 sequence in the DUX4 3'UTR has shown that specific targeting of DUX4 with CRISPR/Cas9 is possible. However, due to the multinucleated nature of skeletal muscle fibers, targeting DUX4 would require a near 100% knockout efficiency to be clinically relevant, which is with

current technologies unfeasible. However, with the fast-evolving genome editing field, new strategies to target the DUX4 gene in an efficient and safe manner may come up, strategies like the base-editing strategy mentioned above, which will be a way forward to find better treatment options for FSHD.

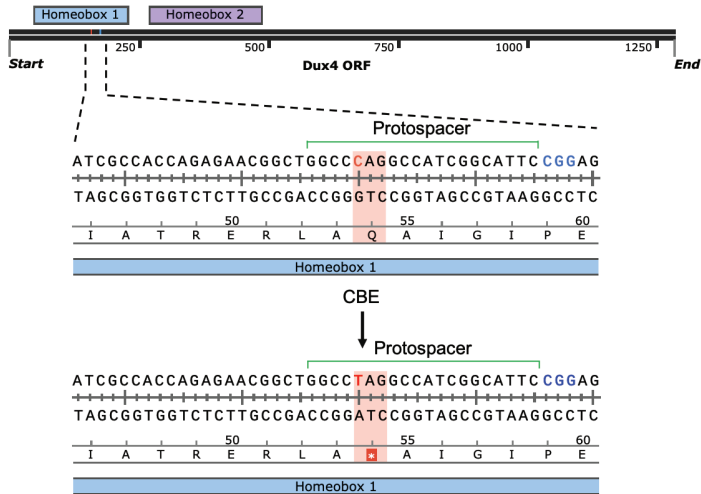


Figure 1. Base editing approach in knocking out DUX4. A schematic representation of a genome editing approach using the a Cas9 nickase fused to a nucleobase deaminase enzyme. The targeted region falls within the first homeobox of the DUX4 open reading frame, 155 nucleotides from the start codon. The PAM sequence is annotated in blue. The nucleotide that is mutated (cytosine to a thymine) is annotated in red, and the amino acid that changes from a glutamine (Q) to a stop codon (*) is highlighted in red.

References

1. Landouzy, L. & Dejerine, J. De la myopathie atrophique progressive (Myopathie héréditaire débutant, dans l'enfance, par la face, sans altération du système nerveux). *C R Acad Sci* 98, 53–55 (1884).
2. Kowaljow, V. et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul. Disord.* 17, 611–623 (2007).
3. Dixit, M. et al. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci. U. S. A.* 104, 18157–18162 (2007).
4. Snider, L. et al. RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: New candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum. Mol. Genet.* 18, 2414–2430 (2009).
5. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712 (2007).
6. Hale, C. R. et al. RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell* 139, 945–956 (2009).
7. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471, 602–607 (2011).
8. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821 (2012).
9. Endo, M., Mikami, M. & Toki, S. Multigene knockout utilizing off-target mutations of the CRISPR/cas9 system in rice. *Plant Cell Physiol.* 56, 41–47 (2015).
10. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278 (2014).

11. Oliva, J. et al. Clinically advanced p38 inhibitors suppress DUX4 expression in cellular and animal models of facioscapulohumeral muscular dystrophys. *J. Pharmacol. Exp. Ther.* 370, 219–230 (2019).
12. Rojas, L. et al. P38 α Regulates Expression of DUX4 in Facioscapulohumeral Muscular Dystrophy. *Biorxiv* 1–19 (2019) doi:<https://doi.org/10.1101/700195>.
13. Himeda, C. L. et al. Identification of Epigenetic Regulators of DUX4-fl for Targeted Therapy of Facioscapulohumeral Muscular Dystrophy. *Mol. Ther.* 26, 1797–1807 (2018).
14. Himeda, C. L. et al. Myogenic Enhancers Regulate Expression of the Facioscapulohumeral Muscular Dystrophy-Associated DUX4 Gene. *Mol. Cell. Biol.* 34, 1942–1955 (2014).
15. Campbell, A. E. et al. BET bromodomain inhibitors and agonists of the beta-2 adrenergic receptor identified in screens for compounds that inhibit DUX4 expression in FSHD muscle cells. *Skelet. Muscle* 7, 1–18 (2017).
16. Shadle, S. C. et al. DUX4-induced dsRNA and MYC mRNA stabilization activate apoptotic pathways in human cell models of facioscapulohumeral dystrophy. *PLoS Genet.* 13, 1–25 (2017).
17. Banerji, C. R. S. et al. β -catenin is central to DUX4-driven network rewiring in facioscapulohumeral muscular dystrophy. *J. R. Soc. Interface* 12, (2015).
18. Lek, A. et al. Applying genome-wide CRISPR-Cas9 screens for therapeutic discovery in facioscapulohumeral muscular dystrophy. *Sci. Transl. Med.* 12, 9–11 (2020).
19. Himeda, C. L., Jones, T. I. & Jones, P. L. CRISPR/dCas9-mediated transcriptional inhibition ameliorates the epigenetic dysregulation at D4Z4 and represses DUX4-fl in FSH muscular dystrophy. *Mol. Ther.* 24, 527–535 (2016).
20. Vanderplanck, C. et al. The FSHD atrophic myotube phenotype is caused by DUX4 expression. *PLoS One* 6, e26820 (2011).
21. Marsollier, A. C. et al. Antisense targeting of 3' end elements involved in DUX4 mRNA processing is an efficient therapeutic strategy for facioscapulohumeral dystrophy: A new gene-silencing approach. *Hum. Mol. Genet.* 25, 1468–1478 (2016).
22. Chen, J. C. J. et al. Morpholino-mediated Knockdown of DUX4 Toward Facioscapulohumeral Muscular Dystrophy Therapeutics. *Mol. Ther.* 24, 1405–1411 (2016).
23. DeSimone, A. M., Leszyk, J., Wagner, K. & Emerson, C. P. Identification of the hyaluronic acid pathway as a therapeutic target for facioscapulohumeral muscular dystrophy. *Sci. Adv.* 5, (2019).
24. Geng, L. N. et al. DUX4 Activates Germline Genes, Retroelements, and Immune Mediators: Implications for Facioscapulohumeral Dystrophy. *Dev. Cell* 22, 38–51 (2012).
25. Rickard, A. M., Petek, L. M. & Miller, D. G. Endogenous DUX4 expression in FSHD myotubes is sufficient to cause cell death and disrupts RNA splicing and cell migration pathways. *Hum. Mol. Genet.* 24, 5901–5914 (2015).
26. Jagannathan, S. et al. Model systems of DUX4 expression recapitulate the transcriptional profile of FSHD cells. *Hum. Mol. Genet.* 25, ddw271 (2016).
27. Van Den Heuvel, A. et al. Single-cell RNA sequencing in facioscapulohumeral muscular dystrophy disease etiology and development. *Hum. Mol. Genet.* 28, 1064–1075 (2019).
28. Whiddon, J. L., Langford, A. T., Wong, C. J., Zhong, J. W. & Tapscott, S. J. Conservation and innovation in the DUX4-family gene network. *Nat. Genet.* 49, 935–940 (2017).
29. Hendrickson, P. G. et al. Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* 49, 925–934 (2017).
30. Vuoristo, S. et al. DUX4 regulates oocyte to embryo transition in human. *Biorxiv* (2019) doi:<http://dx.doi.org/10.1101/732289>.
31. Lemmers, R. J. et al. Deep characterization of a common D4Z4 variant identifies biallelic DUX4 expression as a modifier for disease penetrance in FSHD2. *Eur. J. Hum. Genet.* 26, 94–106 (2018).
32. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* 533, 420–424 (2016).
33. Gaudelli, N. M. et al. Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature* 551, 464–471 (2017).
34. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* 576, 149–157 (2019).

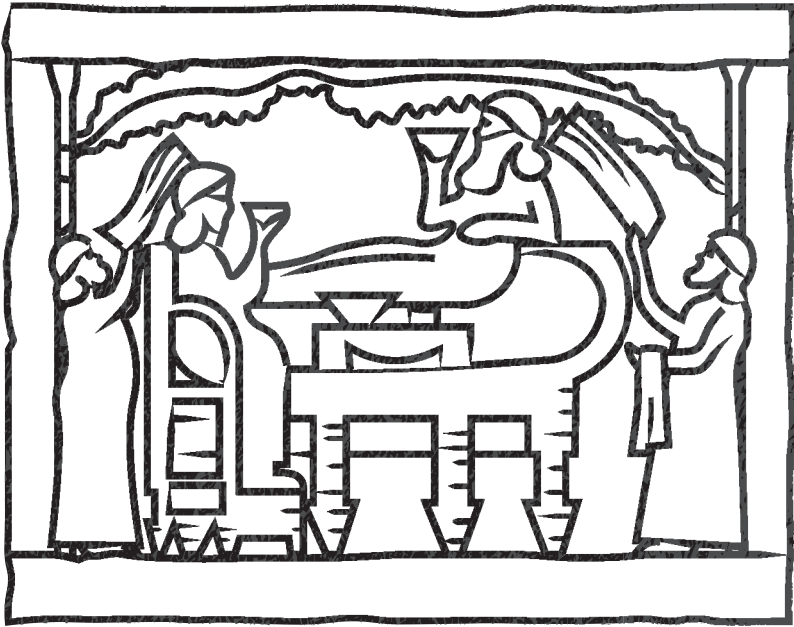


Illustration based on a stone carving on display at the British museum

Addendum

Nederlandse samenvatting

Acknowledgements

Curriculum Vitae

List of publications

Nederlandse samenvatting

Facioscapulohumerale spierdystrofie (FSHD) is één van de meest voorkomende spierziekte wereldwijd. Zoals vele andere spierziektes heeft FSHD een genetische component. De ziekte kan daarom worden geërfd, of kan zich ontwikkelen tijdens de embryonale ontwikkeling. De meeste patiënten ontwikkelen symptomen in de tweede decennia van hun leven. De ziekte openbaart zichzelf beginnend met het verzwakken van de spieren in het gelaat en de schoudergordel, wat zich vervolgens langzaam verspreid naar de boven armen, de romp en in sommige gevallen de onderbenen. FSHD wordt veroorzaakt door veel samenkomende factoren op een moleculair niveau, die vervolgens tot de ongepaste activatie van of Double homeobox 4 (DUX4) leidt. Een gen wat normaal gesproken streng wordt gecontroleerd, en alleen actief is in erg specifieke weefsels en cellen (4-cell embryo, de thymus en de testis). Het exacte moleculaire mechanisme van deze ziekte is erg complex en nog niet helemaal bekend of begrepen. Dit heeft de ontwikkeling van effectieve behandel methodes voor FSHD in weg gestaan. Momenteel worden FSHD-patiënten behandeld met ontstekingsremmers en bewegingsactiviteiten, die een erg beperkt effect hebben op het verloop van de ziekte. Hoofdstuk 1 geeft een gedetailleerd verslag van wat er bekend is over FSHD tot op heden, en hoe we hier zijn beland na 136 jaar research. Er is nog steeds onderzoek gaande naar het in kaart brengen van het moleculaire mechanisme van de ziekte, om zo de ziekte als geheel beter te kunnen begrijpen. Als dit kan worden gerealiseerd, kan er worden bepaald waar we moeten ingrijpen tijdens dit pathologische proces om zo effectievere behandelingsmethodes te ontwikkelen voor FSHD. In dit proefschrift wordt beschreven hoe wij hebben bijgedragen aan het veld, via verschillende routes. Om de ziekte beter te bestuderen hebben we een veelzijdig FSHD-cel model opgezet, waarbij het DUX4 gen naar eigen willen kan worden geactiveerd door het toevoegen van een component genaamd doxycycline. Hoofdstuk 2 legt het door ons ontwikkelde FSHD-cel model in meer detail uit, en beschrijft wat voor effect DUX4 activatie heeft op de cellen. Net als spiercellen gaan de cellen in ons FSHD model dood na de activatie van DUX4. Het effect van DUX4 op deze cellen werd ook bestudeerd op een moleculair niveau, met behulp van RNA-sequencing. Met deze techniek kunnen wij achterhalen welke gene worden beïnvloed door DUX4 activatie. Zo hebben wij vast kunnen stellen dat ons FSHD-model op moleculair niveau in veel opzichten lijkt op FSHD. Om de vroege effecten van DUX4 activatie te kunnen bestuderen is er RNA-sequencing uitgevoerd op afzonderlijke cellen waarin het DUX4 gen slechts enkele uren was geactiveerd (hoofdstuk 3). Als wij in staat zijn deze zeer vroege gebeurtenissen na DUX4 activatie in kaart te brengen, kan er geprobeerd worden in deze vroege stadia in te grijpen. Vroeg ingrijpen in het pathologisch proces zou de kans waarschijnlijk vergroten op het vertragen van het ziekte verloop. De data liet interessante veranderingen zien in de activiteit status van een specifieke set genen, bekend als transcriptiefactoren. Net als DUX4 kunnen deze transcriptiefactoren de moleculaire werking van de cellen beïnvloeden wanneer ze worden geactiveerd of onderdrukt. Sommige van deze transcriptiefactoren waren duidelijk detecteerbaar, terwijl andere alleen tekenen van hun aanwezigheid of afwezigheid vertoonden. Deze factoren zouden van groot belang kunnen zijn in het moleculaire mechanisme van FSHD, en ze zouden daarom interessant kunnen zijn voor verdere onderzoek. Hoofdstuk 4 richt zich op het vinden van sleutelfiguren in het pathologische proces die DUX4 activeert, door het uitvoeren van een knock-out screen. Dit werd uitgevoerd met behulp van de veelgebruikte genoom bewerkingstechniek, bekend als CRISPR/Cas9. Een knock-

out betekent dat een gen volledig wordt gedeactiveerd. Met CRISPR/Cas9 wordt het gen gedeactiveerd door op de plaats van het gen in het DNA te knippen. De cel zal proberen zijn DNA te repareren en daardoor fouten introduceren die kunnen leiden tot gen deactivatie. Een deel van de cellen zullen op deze manier een enkele knock-out van een gen bevatten. Door gebruik te maken van een groot aantal cellen, kunnen we er zeker van zijn dat er van elk gene in het humane genoom een klein groepje cellen zal bestaan waarvan dit gen is gedeactiveerd. Als het deactiveren van een bepaald gen het pathologische proces kan vertragen of stoppen, zouden deze cellen DUX4 activatie betere moeten overleven. Wanneer dit gebeurt, kunnen deze overlevende cellen geanalyseerd worden voor het type knock-out die zij bevatten. Zo kan er worden vastgesteld welk gen belangrijk is in pathologische proces. Zo'n gen werd niet gevonden in deze screen. Dit suggereert dat er na DUX4 activatie niet een ander gen op zichzelf een grote invloed heeft op het pathologische proces van FSHD. We kwamen daarom tot de conclusie dat de beste manier om FSHD af te remmen is om te onderzoeken hoe DUX4 efficiënt kan worden gedeactiveerd. Hoofdstuk 5 demonstreert een manier om DUX4 te deactiveren door gebruik te maken van CRISPR/Cas9. Bij het gebruik maken van CRISPR/Cas9 als een mogelijke therapeutische interventie, moet men altijd rekening houden met de risico's. Omdat CRISPR/Cas9 niet 100% accuraat is, bestaat er een kans dat er op andere plaatsen in het genoom geknipt wordt in het DNA. Dit kan ongewenste en onvoorspelbare bijwerkingen veroorzaken. Omdat er op veel plaatsen in het menselijk genoom kopieën van het DUX4 gen aanwezig zijn, is het een uitdaging om de juiste kopie te deactiveren zonder al te veel DNA-schade in te veroorzaken. De knock-out-methode in hoofdstuk 5 is opgezet om de kans te verkleinen dat andere kopieën van DUX4 worden aangetast, door zich te richten op een relatief unieke sequentie in de ziekteverwekkende kopie van DUX4. Met deze strategie waren we in staat DUX4 in ongeveer de helft van de cellen te deactiveren. Ten slotte zal hoofdstuk 6 een overzicht geven van alle onderzoeks- hoofdstukken en eventuele problemen bediscussieren.

Alles bij elkaar heeft aangetoond hoe gecompliceerd FSHD is, en dat het vinden van een behandelmethode nog erg lastig zou kunnen zijn. Wat wel duidelijk naar voren is gekomen is dat DUX4 inderdaad een grote rol speelt in FSHD, en dat alleen de deactivatie van dit gen het ziekteverloop effectief kan vertragen of zelfs stoppen. Wij zijn daarom van mening dat er meer aandacht moet komen voor het vinden van behandelingsmethoden die zich richten op het verminderen of stoppen van DUX4 activatie.

Acknowledgements

First and foremost, I would like to thank **Niels Geijsen** for accepting me in his lab, and giving me the opportunity to work on such an interesting research field. Your almost eternal optimism and passion for science is truly inspiring and (thankfully) contagious. Your trust and support have helped raise my confidence and encouraged me to keep going forward, even in the toughest times, and I'm therefore incredibly grateful to have had you as my promotor.

I would like to thank **Wouter de Laat** for taking me in when Niels moved his lab to the LUMC to become head of Department of Anatomy and Embryology. You were kind enough to become my promotor so I could finish my work and remain at the Hubrecht institute, giving me the opportunity to finish my PhD the way I had envisioned. Thank you so very much for all the help.

Thank you to all the members from the Geijsen group that I had the pleasure of sharing a lab with, and who had to endure my occasional (or quite often) and somewhat-crazy rants about real or fictional scenarios. You really made my time in the Geijsen group unforgettable and so much fun. Starting with **Nune**. The person who I sat next to the first two years of my PhD, and who has always been so supporting and went above and beyond to help out with anything at any time. I really miss our incomprehensible dialogs, that only we thought were tremendously funny (VUUR!). I was so lucky to have you there when I started, and have learned so much from you. **Fanny**, who I could always go to with any question, be it science related or not. You introduced me to the good wines, and have turned me a bit into a wine snob (no more AH wine for me!). You are one of the nicest people I have ever met, with an immense amount of patience and endurance. You have taught me so much during my time there, and I will be forever grateful. **Peng**, who has taught me almost everything I know about cloning. You have always been such a big help, and a general fun person to be around, saying peculiar things at unexpected times which always cracked me up. The PhD bromance trio that was there when I started, and that I luckily had to share an office with: **Javier**, **Axel** and **PJ**. I loved eavesdropping in your often times strange conversations, sometimes attempting to join in, and eventually succeeding without making an absolute fool myself (right?). Thank you guys so much for being so welcoming, and making me laugh all the time. **Sonja**, the very first student I supervised (halfsies). I always admired you for wanting to eat healthy. Even bringing in plain boiled chicken and rice, pretending wholeheartedly you enjoyed it, while simultaneously gobbling down all my sweets (in stark contrast to the other person sitting next to me). Because we started in the Geijsen group at the same time, we were stuck with each other, and I (thankfully) haven't able to get rid of you since. **Nicolas**, not only an exceptional scientist but also highly skilled at foraging snacks. No food was safe when you were around. Snacks that were left unsupervised mysteriously went missing. But then again, so did the snacks that were supervised. You were so much fun to be around, and I will never

forget one of our first interactions, which was trying to figure out if the oliebollen **Nune** left on her desk were there to torment us. We eventually justified eating them anyways, seeing that we both feel very strongly about not wasting food. **Pascale Dijkers**, whom without I wouldn't have ended up in the Geijssen group. Thank you for connecting me to Niels, and always helping out anyway you could, even before starting my PhD during my masters where you had supervised me during one of my internships. You also provided me with the much-needed brain food (loads of chocolate) during the final years of my PhD, when my main treats provider (**Nune**) had already left. **Melissa**, who I'm convinced woke up every morning thinking: which snack shall I take with me today that will make the most noise while eating (healthy veggie things). But all jokes aside, you are also the person who was always asking me if you could help out in any way. Which you have done in a major way, by not in the least going to Boston during winter, excellently performing a major experiment in my stead, all the while trying not to freeze your behind of. Thank you so much for all your help, and for tricking me into taking a scuba diving crash course. **Lin**, my gym buddy, kickboxing sparring partner, and fellow victim of **Melissa's** scuba diving crash course scheme. You were a great person to have around in the lab, because you weren't shy to give your upfront and unsalted opinion about anything. I genuinely appreciate all the advice and comments you have given me during my time as a PhD, and will take them along in my following endeavors. **Ouafa**, for being the other strong Dutch women with an immigration background (as we say in PC Dutch) in the group. We come from such similar backgrounds, and I loved cracking jokes about all the similar things we had to go through during our upbringing. You were always so uplifting, trying to keep me positives when I was at my limit. Thank you so much for that. **Jorik** and **Darnell** for always entertaining my crazy end of the world, but mainly zombie apocalypse, scenarios. **Clara**, for being such a grounded and hardworking person, yet so much fun on our nights out. **Ada**, for being an absolutely lovely person, with whom I swapped many stories about culture, music and fun TV shows. A big thank you to **Anna**, for helping analyze most of my large data sets, and always answering all my questions, despite being incredibly busy yourself. Although you officially started in the Geijssen group at the end of my PhD, you had already been helping me out with things long before that. I am so thankful that you made time for me in such a busy period in your career, where you had so many different things going on, and even taken the time to actually explain the analysis to me, step by step. Thanks to all the other members in the Geijssen group and the Ntrans team, **Stefan, Zeliha, Chen, Manda, Ive, Julie, Marco, Ruud, Marieke, Kim, Anna, Kelly, Jan**, and all the students that have come and gone in the time that I was there, who all contributed in making my time there so happy.

I have to thank everybody who made the Hubrecht institute such a great environment to work at. I can truly say that I have never before experienced such a sociable institute, performing at such a high level. My time working at the Hubrecht was absolutely wonderful, meeting so many amazing people. **Ilia**, my cell culture companion. I loved our complaining session in the cell culture lab, complaining about all things in life and PhD. You were a great person

to have around on eventful, or even less eventful days. Other members of the Creighton group, **Caroline, Bas, and Mirna**, for our fun and informative chats at the coffee machine, and for answering many of my questions about coding, with a special thanks to Mirna who kindly helped me analyze my first large dataset. Thank you for your patience and all the help. My PV buddies! **Jimmy, Ajit, Stijn, Brian, Laura, and Lotte**. My time in the PV was a blast, and it had everything to do with all the people in it. We got along amazingly well, and meetings were never boring. My TGIF drinking buddies, **Jimmy and Arjan**. Jimmy, I consider you one of the first friends I made at the Hubrecht institute. Together with Arjan we spend many Friday afternoons and evenings drinking beers, chatting and joking away. Playing rocket league together on some of those famous Hubrecht TGIFs are some of my fondest memories of my time at the Hubrecht. Thank you guys for being there for me at the end of many long, and though weeks, and always managing to cheer me up when needed.

My PMC writing buddy **Evelien**, for keeping me motivated during those first few weeks of writing. Our super-efficient writing sessions were unfortunately cut short due to the pandemic, but kickstarted the writing part of my PhD in such a nice way

I would like to thank all the people that have helped me out to get me here and all the people I have collaborated with in and outside of the Hubrecht. All the co-authors that helped putting this work together, **Menno Creighton, Kevin Eggen, Francesco Limone, Federica Piccioni, Judith Vivé, Silvere van der Maarel, Richard Lemmers, Anna Alemany, Mirna Baak, Fanny Sage, and Melissa van Kranenburg**. Thanks to **Richard Sherwood** for advising me about performing and setting up genome wide screens. **Frank van Steenbeek** for being part of my PhD committee, these past 5 years. All my thesis committee members, **Roos Masereeuw, Eva van Rooij, Joost Sluiter, Marie-José Goumans, Bart Spee, and Richard Lemmers**, for taking the time to go through my thesis, and for being there on the day of my defense.

A special thank you to my game night buddies! **Sonja, Javi, Axel**, for all the fun times we had outside of the lab. You guys were and are an essential part of my almost non-existing social life. I am so happy we have kept up with these awesome get togethers, albeit many of them having to be virtual due to the pandemic. I'm always looking forward to seeing all of you again, together with Rel, Eldebar, Fonkin and Nin.

Thank you **Massimiliano and Enrico** for giving me the opportunity to start a new position in your department and group, while finishing up my PhD, and to all the people of the pharmaceuticals department of the David de Wied building for making me feel so welcome and helping me get started in the new lab.

To all the uncles, aunts and cousins that are spread around the world. Thank you for the love and support, and taking care of me during my studies/rotations abroad.



Mijn twee goeie vriendinnen en paranimfen, **Sevgim Ugurlu** en **Chiwan Chiang**. Wie had gedacht dat een vriendschap die 14 jaar geleden zo terughoudend is begonnen, nu zo sterk zou zijn? En hoewel we allemaal een ongelooflijk druk leven leidden nadat we de universiteit hadden verlaten, en elkaar niet zo vaak zagen als we gewend waren, slaagden we er altijd in om tijd voor elkaar te maken wanneer het er echt toe deed. Ik vind het geweldig dat onze vriendschap onverminderd sterk is gebleven, ook al kunnen er maanden voorbijgaan zonder dat we elkaar hebben gesproken of gezien. Bedankt voor alle emotionele steun door de jaren heen en alle geweldige afleidingen tijdens mijn hectische PhD-leven.

Maar het meest dankbaar ben ik mijn geweldige familie. De mensen waarvan ik weet dat ze alles voor me over hebben, en ik voor hun. Ik had dit natuurlijk nooit gekund zonder jullie steun en liefde. Mijn zussen, **Batul** en **Shamiran**, de twee personen waarbij ik werkelijk alles kwijt kan. Mijn grote zussen die ook meteen mijn beste vriendinnen zijn, en die mij altijd hebben gesteund en hun trots voor mij hebben laten blijken. **Sham**, dank je wel dat je altijd je nonchalante zelf bent, altijd klaar om op je meest enthousiast "*ik ben blij voor je*" te zeggen. **Betta**, dank je wel dat je het 5 jaar lang met mij hebt uitgehouden hier in Utrecht, en natuurlijk voor het geweldige design van dit boekje. Mijn grote broer **Daniel**, die mij als kind zijnde altijd heeft beschermt, op de eerste paar maanden van mijn leven na waar hij aanvankelijk er alles aan deed om van mij af te komen. Ik aapte hem altijd na, en volgde hem waar hij ook heen ging, of hij dat wilde of niet. Dank je wel dat jij de meest evenwichtige persoon bent in de familie, die mij altijd kan geruststellen wanneer mijn hoofd weer op standje onrealistisch rampscenario staat. Mijn schoonzusje **Marloes**, en de drie schattigste super boefjes die er bestaan, mijn neefjes en nichtje, **Jeshua**, **James** en **Amara**. Ik kijk er altijd zo naar uit om weer thuis te komen om jullie allemaal te zien, met de kinderen te spelen, en me vol te proppen met het eten wat jullie hebben bereid (Marloes en mams, ook wel bekend als oma Bia). Het doet me altijd zo veel goed! En dan als laats mijn ouders, **Rafael** en **Harbia**. Geboren en opgegroeid als minderheden in een land dat altijd in oorlog was, hebben jullie besloten alles wat jullie kenden, alles wat jullie bezaten, en iedereen die jullie liefhadden achter te laten om jullie kinderen een beter leven te gunnen. Jullie zijn met helemaal niks, behalve jullie 4 kleine kinderen, naar een land gekomen waar jullie niemand kenden, niets hadden, en op dat moment nog niks van af wisten. Jullie hebben jezelf vervolgens kapot gewerkt om jullie gezin alle mogelijkheden te kunnen bieden die jullie als kind nooit hebben gehad. Door alles wat jullie hebben opgeofferd, ben ik nu zover gekomen. Met het volste vertrouwen hebben jullie mij vrijgelaten mijn eigen beslissingen te nemen, en me ook altijd gesteund in die beslissingen, met alleen de boodschap om voorzichtig te zijn en goed op mezelf te passen. Mijn dankbaarheid is immense, en niet in woorden uit te drukken, en ik zeg dit bij lange na niet genoeg: paps, mams, ik hou ontzettend veel van jullie.

Curriculum Vitae

Ator Rafael Odisho Ashoti was born in 1989, on the 25th of November in Dihok, Irak. With her parents and 3 siblings, she moved to the Netherlands in 1993 at the age of 3½. She grew up in Assen, where she attended primary school de Driemaster, and high school Dr. Nassau college Quintus. She started her bachelor of applied science at the Hanze Univeristy in Groningen in 2007. In the last year of her bachelor's program, she obtained an Erasmus scholarship for an internship at the University of Oxford, with the Childhood Cancer Research group, in Oxford (UK). With the supervision of Dr. Kate O'Neill, she tried to uncover a potential link between contracting an Adenovirus or Epstein-Barr virus infection during development or infancy, and childhood leukemia. She graduated from her bachelor's program in 2011, and entered into a pre-master program, with the University of Groningen. She started the Master's program Biomedical Sciences at the University of Groningen in 2012, during which she completed two internships. The first was with the group of prof. Ody Sibon, in the department of Cell Biology, at the University Medical Center Groningen, and under the direct supervision of Dr. Pascale Dijkers. During this 6-month internship Ator studied the role of the enzyme Vanin in immunity and ageing of *Drosophila melanogaster*. Her second internship was with the Children's Medical Research Institute in Sydney Australia, within the group of prof. Patrick Tam in the department of Embryology, and under the direct supervision of Dr. David Loebel. Here Ator studied downstream targets of Twist1 in murine cranial mesoderm. Ator graduated Cum Laude from the Biomedical Sciences Master's program in 2014. In April 2015, she was awarded a 4-year PhD fellowship in Regenerative Medicine from the Netherlands Organization for Scientific Research (NWO). She moved to Utrecht and started her PhD in September 2015 under the supervision of prof. Niels Geijsen, at the Hubrecht institute in the Netherlands. The result of her PhD is described in this thesis.

List of publications

Ashoti, A.*, Limone, F.*, van Kranenburg, M., Alemany, A., Baak, M., Vivie, J., Piccioni, F., Creighton, M., Eggen, K., & Geijsen, N. Considerations and practical implications of performing a phenotypic CRISPR/Cas survival screen.

Submitted

Available in and adapted form on bioRxiv (2020).

<https://doi.org/10.1101/2020.07.27.223420>

Ashoti, A., Alemany, A., Sage, F. & Geijsen, N. DUX4 induces a homogeneous sequence of molecular changes, culminating in the activation of a stem-cell-like transcriptional network and induction of apoptosis in somatic cells. bioRxiv (2021).

<https://doi.org/10.1101/2021.05.04.442407>.

Submitted

Publication before PhD

Bildsoe, H., Fan, X., Wilkie, E. E., **Ashoti, A.**, Jones, V. J., Power, M., Qin, J., Wang, J., Tam P. P. L., & Loebel, D. A. F. (2016). Transcriptional targets of TWIST1 in the cranial mesoderm regulate cell-matrix interactions and mesenchyme maintenance. *Developmental Biology*, 418(1), 189–203.

<https://doi.org/10.1016/j.ydbio.2016.08.016>

* Equal contribution

