Double homeobox 4 activates germline genes, endogenous retroelements and immune modulators: Implications for facioscapulohumeral muscular dystrophy

Linda N. Geng

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2011

Stephen J. Tapscott, Chair

Edith H. Wang

Muneesh Tewari

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

Double homeobox 4 activates germline genes, endogenous retroelements and immune modulators: Implications for facioscapulohumeral muscular dystrophy

Linda N. Geng

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

Double homeobox 4 (DUX4) is a candidate disease gene for causing facioscapulohumeral dystrophy (FSHD), a condition characterized by progressive degeneration of specific skeletal muscle groups. While the genetic lesion associated with the disease has been known for decades, the molecular mechanism(s) leading to muscular dystrophy remained unclear. DUX4 was detected in FSHD muscle, but the abundance was extremely low and the downstream consequence of DUX4 in human muscle cells was not known. This work demonstrates that the low level of DUX4 expression in a population of FSHD muscle cells represents a relatively high expression of DUX4 mRNA and protein in a subset of cells at any given point in time. While the full length DUX4 induces toxicity in muscle cells, the shorter isoform without the C-terminal end, DUX4-s, exhibited no obvious detrimental effects on muscle cells. Genome-wide binding studies revealed that DUX4 protein binds to a DNA target sequence that contains two closely-spaced canonical homeodomain binding sequences TAAT in tandem. These binding sites are present in both unique regions of the genome as well as within LTRs of a family of endogenous retrotransposons called MaLR. DUX4-s can bind to the same sequences, but, interestingly, DUX4-s was unable to activate transcription in luciferase reporter assays and, in fact, could act as a dominant negative to inhibit full length DUX4's activity. Microarray expression analysis showed that DUX4 can upregulate over a thousand genes involved in multiple processes, such as germ cell development, RNA splicing and immune modulation. These targets are indeed

upregulated in cultured FSHD muscle cells as well as biopsies, providing further support for the causal role of DUX4 in FSHD. One of the DUX4 targets, DEFB103, not only suppresses the innate immune response to viral infection but also inhibits myogenic differentiation. Therefore, FSHD represents a disease where incomplete developmental silencing of a retrogene DUX4 results in inappropriate expression of its targets in skeletal muscle cells. These findings suggest specific mechanisms of FSHD pathology and identify candidate biomarkers for disease diagnosis and progression.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I offer my sincerest gratitude to my supervisor Dr. Stephen Tapscott for his invaluable support, guidance and wisdom. I also wish to thank everyone in the Tapscott lab for being not only my scientific colleagues but also my friends and mentors. I would like to further extend my appreciation to collaborators at the Fred Hutchinson and elsewhere who have greatly aided in the progression of this work and to my doctoral thesis committee who have helped shape me as a scientist. I stand on the shoulder of giants.

DEDICATION

To my parents and O.S.P.

Chapter 1: Introduction

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

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

Snider L*, **Geng LN***, Lemmers RJ, Kyba M, Ware CB, Nelson AM, Tawil R, Filippova GN, van der Maarel SM, Tapscott SJ, Miller DG. (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet.* Oct 28;6(10):e1001181.
*These authors contributed equally

Snider L, Asawachaicharn A, Tyler AE, **Geng LN**, Petek LM, Maves L, Miller DG, Lemmers RJ, Winokur ST, Tawil R, van der Maarel SM, Filippova GN, Tapscott SJ. (2009) RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum Mol Genet.* Jul 1;18(13):2414-30.

Facioscapulohumeral muscular dystrophy

Clinical features

Muscular dystrophy describes a group of inherited diseases characterized by degeneration and progressive weakening of affected muscle groups. There are currently no cures for any of the muscular dystrophies, though supportive therapies can help to slow the course of the disease. Facioscapulohumeral dystrophy (FSHD) is the third most common muscular dystrophy after Duchenne and myotonic dystrophies, with a prevalence of about 4 per 100,000 and incidence of about 1 in 20,000 (Orrell, 2011). FSHD is named for the muscle groups primarily affected in individuals with the disease, the face, shoulders and upper arms.

FSHD is generally a slowly progressive disease, but the clinical severity and age of onset can be extremely variable. The common classic presentation of FSHD is characterized by onset in the second or third decade of life and slow disease progression (Tawil, 2008). Most patients have a normal life span. Symptoms initiate in the face, beginning with inability to close the eyes tightly, smile or whistle, though these may be subtle in the early stages. Weakness usually progresses to the shoulder and upper arms with marked atrophy of the biceps and triceps, but sparing of the deltoids. Another characteristic feature is the winging of the scapula, affecting the patient's ability to lift, pull and push heavy objects and sometimes perform activities of daily living. The rare infantile cases of FSHD present with early onset in the first few years of life and rapid progression resulting in wheelchair confinement by the first decade in most cases (Taylor et al., 1982). Patients present with severe facial weakness, including inability to close eyes during sleep, to smile and to show other facial expressions. Weakness spreads quickly to the shoulder and hip girdles with lumbar lordosis. Non-muscular conditions of exudative telangiectasia of the retina (or Coats syndrome) and sensorineural hearing loss have also been reported to be associated with both infantile and classic cases of FSHD (Munsat et al., 1972; Small, 1968; Taylor et al., 1982; Voit et al., 1986).

Genetic features

The primary form of FSHD (FSHD1: OMIM 158900) is inherited in an autosomal dominant manner and diagnosed by a DNA test to detect deletions of a tandem array of 3.3 kb repeated sequence called D4Z4. The D4Z4 macrosatellite is located in the subtelomeric region of chromosome 4q35 and each unit contains an open reading frame (ORF) for a double homeobox transcription factor, DUX4 (Gabriels et al., 1999; Hewitt et al., 1994). The number of repeat subunits is polymorphic in the human population, and unaffected individuals have 11 to 100 copies whereas FSHD1 patients have 1 to 10 copies (Hewitt et al., 1994; Wijmenga et al., 1992). The number of residual copies of D4Z4 units is not linearly anti-correlated with clinical severity, though patients with 1-3 residual units tend to have infantile presentation and more rapid progression (Lunt et al., 1995).

Contraction of D4Z4 units is necessary but not sufficient for FSHD1. First, there are highly homologous D4Z4 repeats on chromosome 10, but deletion of D4Z4 repeats on 10q is not linked to FSHD (Bakker et al., 1995; Deidda et al., 1995). This led to the proposal of position effect variegation (Gabellini et al., 2002; Jiang et al., 2003) or looping (Bodega et al., 2009) models where the shortening of the repeat would change the expression of nearby chromosome 4-specific genes. However, at least one remaining unit of D4Z4 is necessary for FSHD, since complete deletion of 4q35 does not lead to FSHD (Tupler et al., 1996), indicating that the pathogenic agent lies within the repeat sequence itself. Second, not all contractions of 4q D4Z4 repeats were linked to the disease. The deletions had to occur in the background of specific haplotypes defined by sequences flanking the D4Z4 repeat, namely the presence or absence of a downstream β-satellite and an upstream simple sequence length polymorphism (Lemmers et al., 2010b). Recently, a polyadenylation signal for transcripts derived from D4Z4 was found in the sequence distal to the repeat (Dixit et al., 2007; Snider et al., 2009) and, it was shown that the presence of this polyadenylation signal in its intact form is necessary for FSHD and explained some of the variable disease-permissiveness of certain haplotypes (Lemmers et al., 2010a).

Molecular features

While the majority of known FSHD cases are linked to the D4Z4 contraction on chromosome 4, there is a subgroup of patients who present with the same disease phenotype but have repeat numbers in the lower end of the normal range (FSHD2: OMIM 158901). Interestingly, both FSHD1 and FSHD2 patients show reduction in DNA methylation and decreased markers of heterochromatin at the D4Z4 disease locus in addition to having one of the permissive genetic haplotypes (de Greef et al., 2010). Studies have shown that the D4Z4 macrosatellite is normally in a relatively closed chromatin state and that, as a result of either repeat contraction (FSHD1) or an as yet unknown factor (FSHD2), it loses some of its epigenetic repression (Jiang et al., 2003; Lyle et al., 1995; van Overveld et al., 2003).

The model then predicts that derepression of D4Z4 would result in aberrant transcripts derived from this locus. Initial attempts to identify D4Z4-derived transcripts were unsuccessful , so other genes proximal to the D4Z4 macrosatellite were examined, including FRG1, FRG2 and ANT1 (Gabellini et al., 2002). While some studies found an increase in the expression of these genes in FSHD samples, others did not, resulting in a lack of consensus regarding the role of these genes in disease pathogenesis (Dmitriev et al., 2009). In recent years, high sensitivity PCR was used to identify polyadenylated mRNA derived from D4Z4, including a transcript containing the full DUX4 ORF (Dixit et al., 2007; Snider et al., 2009) as well as splice forms of DUX4 transcript that truncate the coding sequence and several small RNAs (Snider et al., 2009). These findings, in conjunction with the recent genetic and epigenetic studies returned the focus to DUX4 retrogene as a primary causative factor for FSHD (van der Maarel et al., 2011), but the extremely low abundance of gene product remained a problem for this DUX4-based disease model.

Double homeobox 4

Evolution of DUXs

Homeobox genes are all characterized by the presence of a coding sequence called the homeobox which encodes a variable protein motif called the homeodomain that binds DNA (Banerjee-Basu et al., 2001; Scott et al., 1989). Double homeobox (DUX) genes are defined by the presence of two homeoboxes, usually positioned in close proximity to each other. Multiple members of the double homeobox gene family exist in the human genome, but their evolutionary history is complex because many are found in repetitive regions, are intronless and often associated with pseudogenes (Zhong and Holland, 2011). Though DUX4 is considered an intronless retrogene (Clapp et al., 2007), there are intron-containing DUX genes including DUXA and DUXB (Booth and Holland, 2007). Two other intron-containing DUX genes, DUXC and DUXBL (DUXB-like) were found in some mammals but lost in humans (Clapp et al., 2007; Leidenroth and Hewitt, 2010). Clapp et al (2007) proposed that the human DUX4 was derived from a retrotransposition of the mRNA from the ancestral DUXC gene. In contrast to many other retrotransposed sequences which have accumulated mutations and become defunct pseudogenes, the coding sequence of DUX4 has been well conserved, suggesting that it offered a selective advantage to its host organism and may have a functional role in normal biology (Clapp et al., 2007).

Normal role of DUXs

Many homeobox genes are critical to the embryonic development of metazoans and function as transcription factors that regulate morphogenesis and cell differentiation (Gehring, 1992). Single homeobox genes like the famous HOX genes have been studied to a great extent, but very little is known about the function of the double homeobox subfamily. In their classification of all human homeobox genes, Holland et al (2007) described the existence of these DUXs and concluded that "Few of these tandemly-repeated sequences are likely to be functional as expressed proteins, and all were probably derived by retrotransposition from functional DUX gene transcripts." However, as mentioned above, DUX4 contains an ORF that has been conserved through evolution making it less likely to be just a relic pseudogene. Studies of related DUXs in mice have also begun to shed light on possible functions of double homeobox genes. The mouse DUXbl gene, which encodes homeodomain sequences highly similar to those

in the human DUX4 protein, was found to be expressed in reproductive organs, more specifically in spermatogonia and oocytes from embryonic to adult stages, and in limbs and tail during embryonic development (Wu et al., 2010). Another group demonstrated that the mouse DUXl gene is upregulated by Runx1 and promotes the developmental progression of double-negative thymocytes (Kawazu et al., 2007). Together these findings lead to the hypothesis that double homeodomain transcription factors play a normal role in gametogenesis, early development and immune regulation.

Pathological function of DUX4

Several homeobox genes have been implicated in human pathologies and congenital abnormalities (Boncinelli, 1997). DUX4 is a leading candidate disease gene for FSHD and its overexpression in a variety of cell types is known to result in cytotoxicity (Bosnakovski et al., 2008b). Injection of DUX4 mRNA into zebrafish embryos severely inhibited myogenesis (Snider et al., 2009). Another group demonstrated that ectopic expression of DUX4 in *Xenopus* embryos broadly induced cell death and developmental abnormalities (Wuebbles et al., 2010). DUX4 was found to localize to the nucleus, redistribute the nuclear membrane protein emerin and induce caspase 3/7-dependent apoptosis when transfected into certain human cell lines (Kowaljow et al., 2007). Recently, it was shown that DUX4-induced apoptosis in muscle cells is at least partially effected through p53 activation, as muscles from p53-null mice (-/-) exhibited minimal damage from DUX4 (Wallace et al., 2010). Though these studies offered clues to the downstream consequences of DUX4 overexpression, the pathophysiological mechanism(s) by which low levels of endogenous DUX4 in FSHD muscle cells might lead to the muscular dystrophy remained unclear.

Chapter 2: Immunodetection of double homeobox 4

This chapter has been published as:

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

Summary

Double homeobox 4 (DUX4) is a candidate disease gene for facioscapulohumeral dystrophy (FSHD), one of the most common muscular dystrophies characterized by progressive skeletal muscle degeneration. Despite great strides in understanding precise genetics of FSHD, the molecular pathophysiology of the disease remains unclear. One of the major limitations has been the availability of appropriate molecular tools to study DUX4 protein. In the present study, we report the development of five new monoclonal antibodies targeted against the N- and C-termini of human DUX4, and characterize their reactivity using Western blot and immunofluorescence staining. Additionally, we show that expression of the canonical full coding DUX4 induces cell death in human primary muscle cells, whereas the expression of a shorter splice form of DUX4 results in no such toxicity. Immunostaining with these new antibodies reveals a differential effect of two DUX4 isoforms on human muscle cells. These antibodies will provide an excellent tool for investigating the role of DUX4 in FSHD pathogenesis.

Introduction

FSHD (FSHD1; OMIM 158900) is an autosomal dominant disease and the third most common muscular dystrophy (Tawil, 2008). It is characterized by muscle wasting and progressive weakness that usually initiates in the facial and upper trunk regions and eventually spreads through the body, sometimes resulting in wheelchair confinement (Padberg and van Engelen, 2009; Pandya et al., 2008). FSHD is associated with a contraction in a critical number of repeats of the macrosatellite D4Z4 on the haplotype 4qA161 (Lemmers et al., 2007; Wijmenga et al., 1992). Within each of these repeats resides the open reading frame for the double homeobox gene *DUX4*, a transcription factor. A recent study by Lemmers et al revealed that the allele specificity of FSHD is due to polymorphisms in the polyadenylation signal for transcripts derived from *DUX4* (Lemmers et al., 2010a). We and others have also reported the presence of polyadenylated transcripts arising from this disease locus that encode the full DUX4 protein in FSHD muscle cells, but the protein has thus far been elusive and difficult to

detect consistently (Dixit et al., 2007; Snider et al., 2009). Overexpression of DUX4 in cultured muscle cells has been shown to induce apoptosis (Kowaljow et al., 2007) and increase oxidative stress (Bosnakovski et al., 2008b), possibly reflecting mechanisms of muscle deterioration seen in FSHD. Currently, to our knowledge, there is only one anti-DUX4 monoclonal antibody previously reported in the literature called 9A12 (Dixit et al., 2007). While it has been useful, the mouse monoclonal 9A12 antibody does not distinguish DUX4 from DUX4c, another double homeobox that shares more than two thirds of its sequence with DUX4 (Ansseau et al., 2009; Bosnakovski et al., 2008a). It is important to distinguish the two species as *DUX4c* has also been proposed as a candidate gene for FSHD. Thus, we raised our antibodies against the unique C-terminus region of DUX4 in addition to the shared N-terminus. In this study, we generated three mouse monoclonal antibodies P4H2, P2G4 and P2B1 and two rabbit monoclonal antibodies E5-5 and E14-3, and we report the characterization of these novel monoclonal antibodies to human DUX4.

Results

Production of antibodies against the human DUX4 N- and C-termini

The first 159 amino acid sequence of DUX4 was used as the N-terminus antigen and the last 76 amino acid sequence of DUX4 was used as the C-terminus antigen, which is not shared by DUX4c (Fig. 2.1A). To generate the recombinant C-terminus antigen for immunization, the last 228 base pairs of DUX4 open reading frame was cloned in frame after a GST tag and expressed in *E. coli*. Induction of *E. coli* using 1 mM IPTG resulted in a high level of expression of a protein of the expected size at 36 kDa (Fig. 2.1B). The expressed product was purified on a glutathione coupled agarose column and eluted by competition with free glutathione. Peak A280 fractions eluted from the column were pooled and showed a high level of purity by SDS-PAGE (Fig. 2.1B, lanes 3 and 4). The purified GST-DUX4 C-terminus was concentrated and injected into two rabbits and two mice for production of antibodies. Production of the N-terminal antigen was performed

in the same manner and yielded similar purity (data not shown), and was injected into two rabbits and two mice for development of antibodies.

The antisera from all animals were screened for reactivity by western blots and immunofluorescence against transfected DUX4 (data not shown). After fusion, the subsequent hybridoma clones were screened by ELISA for positive reactivity for the fusion protein and negative reactivity for GST protein alone. To narrow the selection of clones, they were further screened by western blots and immunofluorescence on transfected DUX4. The clones with the best reactivity from screening were mouse monoclonals P4H2, P2G4 and P2B1 and rabbit monoclonals E5-5 and E14-3 (Table 2.1).

P4H2, P2G4, P2B1, E5-5 and E14-3 specifically recognize the DUX4 protein

C2C12 mouse myoblast cells were transfected with plasmid expressing human DUX4. The sequence contained in the expression plasmid includes an upstream alternative initiation site about 60 amino acids away, thus producing two protein species differing by approximately 10 kDa in molecular weight. P4H2, E5-5 and E14-3 were all reactive to transfected DUX4 and none showed reactivity against untransfected C2C12 lysate by western blot analysis (Fig. 2.2A). Mouse monoclonals P2G4 and P2B1 were less effective on the western blot method (data not shown). To assess for cross-reactivity against DUX4c, we transfected C2C12 cells with a plasmid expressing human DUX4c. Mouse monoclonal P4H2 and rabbit monoclonal E5-5 displayed no reactivity against DUX4c, whereas E14-3 also recognizes DUX4c which shares its N-terminus antigen sequence (Fig. 2.2A). Western blot using the previously characterized 9A12 monoclonal antibody confirmed the identity of DUX4 (Fig. 2.2B). Blots were stripped and reprobed with anti-α-tubulin antibody and showed similar loading of all samples per blot.

C2C12 cells transfected with DUX4 were also fixed and stained with the antibodies. The immunofluorescence analysis using all five monoclonal antibodies demonstrated strong nuclear staining of cells expressing DUX4 that is not seen in neighboring cells which were not expressing DUX4 (Fig. 2.3). The immunostaining indicates that these antibodies are able to detect DUX4 protein expressed in the nuclei

of mammalian cells, which agrees with previous studies on DUX4's cellular distribution using epitope tag-based experiments (Ostlund et al., 2005). The nuclear localization of DUX4 is consistent with its function as a transcription factor that is able to bind to DNA sequences via its homeodomains and regulate target gene expression.

Full length DUX4 induces cell death whereas the shorter splice form does not

To examine the effects of DUX4 when expressed at a lower level, a retroviral vector carrying *DUX4* was used to introduce DUX4 into C2C12 myoblasts. At 42 hours post-infection, cells were fixed, permeabilized and co-labeled with mouse monoclonal P4H2 and rabbit monoclonal E14-3. Double-positive DUX4 nuclei on immunofluorescence exhibited a punctate, granular pattern suggestive of chromatin condensation and possibly a pre-apoptotic state (Fig. 2.4A). Kowaljow et al (2007) demonstrated that ectopic overexpression of DUX4 in HEp-2 cells, an epidermoid carcinoma line, results in apoptosis via caspase 3/7 activation just 24 hours post-transfection, though dramatically more evident at 48 hours. The expression of DUX4 from our retroviral construct is likely closer to endogenous levels, and, thus, the effects on the murine myoblasts are more subtle, but the results are still indicative of cellular changes associated with the early stages of cell death.

To examine the effects of the different DUX4 splice forms in the context of human muscle cells that would be highly relevant to FSHD, we used a lentiviral delivery system to transduce human primary myoblasts with full length and short DUX4. An equivalent number of myoblasts were infected with either splice form, but 48 hours post-infection, more than half the cells were dead on the plates with full length DUX4, whereas there was minimal, if any, cell toxicity was seen on the plates with the short form of DUX4 (Fig. 2.4B). At 24 hours post-infection, both full length and short DUX4 show little effect on cell viability and morphology. However, at 48 hours post-transfection, full length DUX4-infected cells exhibited condensation of nuclei and cytoplasm, blebbing of membranes and fragmentation of cells, features characteristic of apoptosis (Fig. 2.4C, panel c-d). The cells infected with short DUX4 remain intact, round and full, and show homogenous staining (Fig. 2.4C, panels g-h).

Discussion

Few antibodies are currently available against DUX4. Only one polyclonal antibody is commercially available (Santa Cruz Biotechnology, Inc., Santa Cruz, CA), and it has not been characterized in the literature to our knowledge. The 9A12 mouse monoclonal antibody developed by Dixit *et al.* has been useful (Dixit et al., 2007), but it cross reacts with another highly related protein and candidate disease gene DUX4c which shares over two-thirds of its N-terminal sequence with DUX4 (Ansseau et al., 2009; Bosnakovski et al., 2008a). We have generated the monoclonal antibodies P4H2, P2B1 and E5-5 specific to the C-terminus sequence of DUX4 that is not present in DUX4c. We also generated monoclonal antibodies P2G4 and E14-3 against the N-terminus of DUX4. Three antibodies, P4H2, E5-5 and E14-3, can detect both denatured DUX4 protein on a western and fixed native DUX4 protein by immunofluorescence. Antibodies P2G4 and P2B1 work well to detect fixed native DUX4 protein by immunofluorescence, but are less effective for denatured DUX4 on western blot. These new monoclonal antibodies can be used in combination with other monoclonal and polyclonal antibodies to assess the protein expression of DUX4 in muscle cells affected by FSHD. Additionally, multiple antibody labeling experiments and combination methods such as immunoprecipitation followed by western blot can be performed with these anti-DUX4 antibodies produced from different species. Development of tools for studying DUX4 protein is critical to advancing the field of FSHD research and for elucidating the molecular pathogenesis of this debilitating disease.

We previously identified alternatively spliced transcripts from DUX4 in FSHD and wild-type muscle cells, including the canonical full length DUX4 that contains the entire coding region of DUX4 as well as a shorter splice form that removes more than two thirds of the coding region (Snider et al., 2009). While studies have shown the full length DUX4 to be toxic to cultured cells as mentioned above, there have been no reports on the effects of the shorter splice form on human muscle cells. Here we show that while full length DUX4 exerts overt toxicity on human myoblasts leading to cell

death, the short DUX4 shows no obvious negative effects, suggesting that the critical effector component of DUX4 mediating its pro-apoptotic effects is located within the C-terminal region of the protein that is unique to full length form. This is the first demonstration that the different splice forms of DUX4 result in dramatically different effects on human muscle cells. It remains to be investigated whether the short DUX4 performs a beneficial function in muscle cells.

DUX4 is the leading candidate disease gene for FSHD, but the paucity of appropriate molecular tools has greatly limited the study of DUX4 protein expression and function. We have developed five new antibodies that specifically recognize DUX4 by western blot and immunofluorescence. Immunostaining with the antibodies reveal a differential toxic effect of DUX4 isoforms on skeletal muscle cells and suggest that perhaps only specific splice forms of DUX4 are responsible for the pathogenesis of FSHD. These new tools will provide the opportunity to dissect the role of DUX4 in FSHD by various molecular methods.

Materials and Methods

Expression and purification of DUX4 fusion protein.

The sequence encoding the last C-terminal 76 amino acids of DUX4 was amplified by PCR using forward primer 5'-CGCGGATCCCCATGCAAGGCATCCCGGCGC-3' and reverse primer 5'-CCGGAATTCCTAAAGCTCCTCCAGCAGAGCCCG-3' and cloned in frame after glutathione-s-transferase in the bacterial expression vector pGEX-3x (*Glutagene*; *Amrad* Corporation, Kew, Victoria, Australia). The sequence encoding the first N-terminal 159 amino acids of DUX4 was amplified by PCR using forward primer 5'-CGCGGATCCCCATGGCCCTCCCGACACCCTC-3' and reverse primer 5'-CCGGAATTCCTGCGCGGGCGCCCTG-3' and cloned in frame after glutathione-s-transferase in pGEX-3x. The template used for both N- and C-termini cloning was the previously described pCS2+mkgDUX4 (8). Expression of the fusion protein in *E. coli* strain BL21DE3pLysS was induced by 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 37°C for 4 h. The expression product was purified

using B-Per GST Fusion Protein Purification Kit (Pierce, Rockford, IL), according to manufacturer's instructions. The purity and size of the fusion protein was determined by 10% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) followed by staining with Coomassie Blue dye.

Antibody production and screening.

Purified and concentrated fusion protein was used as immunogen. Antibody production was done in collaboration with Epitomics (Burlingame, CA) for the rabbit monoclonals and by the Antibody Development Laboratory at the Fred Hutchinson Cancer Research Center (Seattle, WA) for the mouse monoclonal. The mouse monoclonals will be commercially available. The rabbit monoclonals will be available through Epitomics. The antisera from all animals were screened for reactivity by ELISA against the immunogen, and western blots and immunofluorescence for against transfected DUX4. The best rabbit and mouse were chosen for fusion, and the subsequent hybridoma clones were screened by ELISA for positive reactivity for the fusion protein and negative reactivity for GST protein alone. To narrow the selection of clones, they were further screened by western blots and immunofluorescence on transfected DUX4. The best clones were subcloned, expanded and isotyped by the production facilities.

Expression plasmids and ectopic expression.

The pCS2+mkgDUX4 and pClneo-DUX4c expression vectors used were previously described (Kowaljow et al., 2007; Snider et al., 2009). The murine myoblast line C2C12 was maintained in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum and 1% penicillin/streptomycin (GIBCO, Carlsbad, CA) at 37°C in an atmosphere containing 5% CO_2. Transient transfections were performed using SuperFect reagent (Qiagen, Valencia, CA) according to manufacturer specifications, and cells were harvested 24 h post-transfection for lysate or direct fixation. pclBABE+DUX4 was constructed by inserting the mkgDUX4 from pCS2 construct into blunted BamHI and EcoRI sites of a pclBABE vector (similar to Addgene Plasmid 20917, in place of MyoD) and transfected into C2C12 myoblasts using SuperFect as described above, and

then collected for staining at 42 h post-transfection. Lentiviral constructs were made by inserting the coding sequences of full length and short DUX4 in place of GFP in the plasmid pRRLSIN.cPPT.PGK-GFP.WPRE (Addgene, Cambridge, MA). Full length and short DUX4 are also referred to as DUX4-fl and DUX4-s, respectively (Snider et al., 2010).

Western blot.

Cells were directly lysed in 2x laemmli sample buffer and sonicated to shear genomic DNA. Equal amounts of untransfected C2C12, transfected DUX4 and transfected DUX4c lysates were loaded onto 10% bis-tris polyacrylamide gels and separated by electrophoresis. Proteins were transferred onto nitrocellulose membranes (Invitrogen, Carlsbad, CA) and blocked with 5% non-fat dry milk in phosphate buffered saline (PBS) and 0.1% Tween-20 (PBST). The membranes were then separately incubated in appropriate dilutions of E5-5, E14-3 and P4H2 hybridoma supernatants in PBST overnight at 4°C. Blots were washed in PBST and incubated with appropriate secondary antibodies, HRP-conjugated goat-anti-rabbit IgG or HRP-conjugated goat-anti-mouse IgG (Jackson ImmunoResearch Laboratories, West Grove, PA), and subsequently washed again. Pierce ECL Western Blotting Substrate (Thermo Scientific Pierce, Rockford, IL) was added to membranes, and then blots were exposed to film and developed. Blots were stripped and reprobed with anti-α tubulin antibody (Sigma, St. Louis, MO) as a loading control.

Immunofluorescence.

Cells were gently washed in PBS and fixed in 2% paraformaldehyde for 7 min at room temperature and then washed twice with PBS. Cells were permeabilized with 1% Triton X-100 (Sigma, St. Louis, MO) in PBS for 10 min at room temperature with gentle rocking. Primary antibodies E5-5, E14-3 and P4H2 were added at appropriate dilutions in PBS overnight at 4°C and then washed with PBS. TRITC- or FITC-conjugated goat anti-rabbit or mouse secondary antibodies (Jackson ImmunoResearch Laboratories, West Grove, PA) were added for incubation for 1 h at room temperature and then

washed with PBS and counterstained for DNA with 4′,6′-diamidino-2-phenylindole (DAPI). Cells were examined under a fluorescence microscope.

Table 2.1. Properties of anti-DUX4 monoclonal antibodies

Antibody	*Species*	*Isotype*	*Antigen*	*Screening methods*
E5-5	Rabbit	n/a	C-terminus	ELISA, western blot, IF*
E14-3	Rabbit	n/a	N-terminus	ELISA, western blot, IF*
P4H2	Mouse	IgG1	C-terminus	ELISA, western blot, IF*
P2G4	Mouse	IgG1	N-terminus	ELISA, western blot[1], IF*
P2B1	Mouse	IgG1	C-terminus	ELISA, western blot[1], IF*

* IF = immunofluorescence
[1]less effective

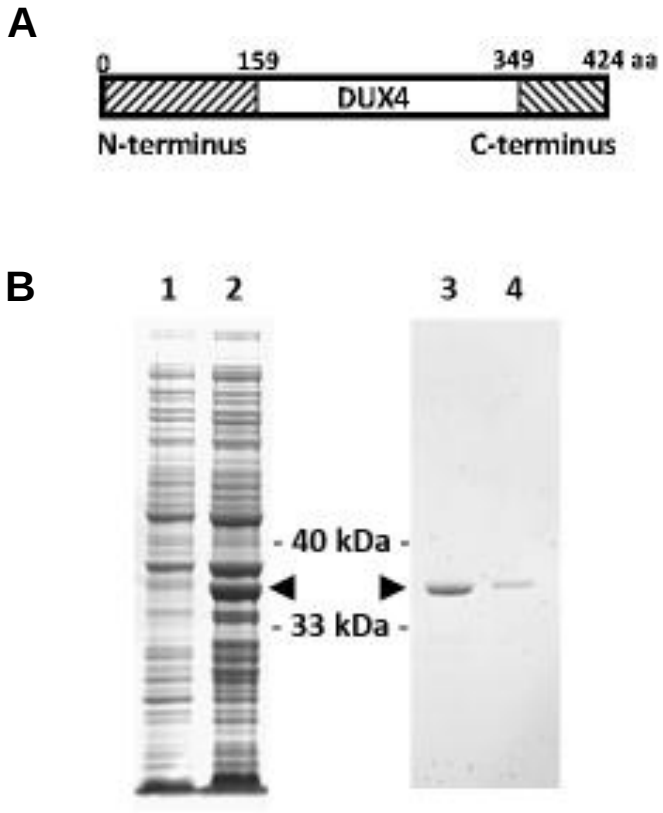


Figure 2.1. Production of recombinant DUX4 antigen.
(A) Schematic of DUX4 protein and the 159 amino acid N-terminus and 76 amino acid C-terminus antigens used to raise antibodies; (B) Expression and purification of GST-tagged recombinant C-terminus DUX4 antigen, indicated by black arrow heads. Bacterial lysates and purified fusion protein were separated by SDS-PAGE and stained with Coomassie blue. Lane 1, uninduced bacterial cells; Lane 2, bacterial cells induced with IPTG for 4h; Lane 3 and Lane 4, purified DUX4 fusion protein.

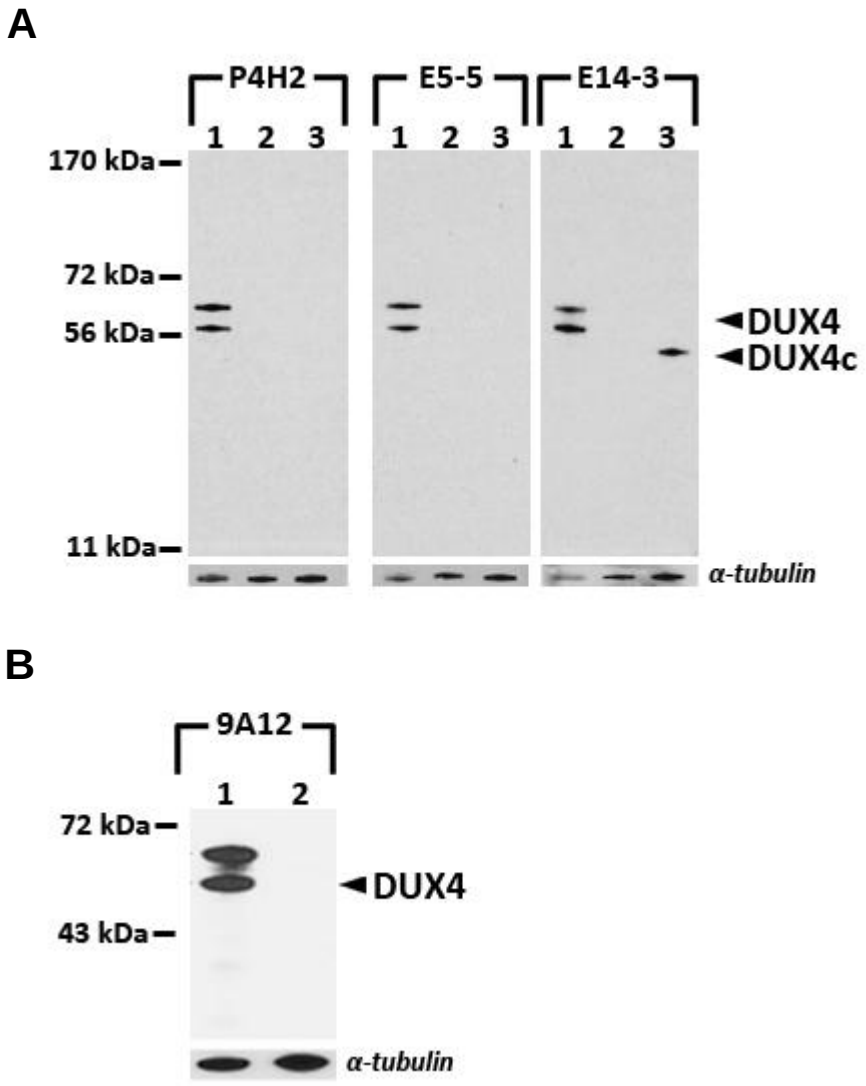


Figure 2.2. Reactivity of monoclonal antibodies on western blot.
Sample 1, C2C12 cells transfected with pCS2+mkgDUX4; sample 2, untransfected C2C12 cells; sample 3, C2C12 cells transfected with pClneo-DUX4c. Blots were probed with the appropriate antibodies and then stripped and reprobed with anti-α-tubulin antibody for loading control. (A) mouse monoclonal P4H2, rabbit monoclonal E5-5 and rabbit monoclonal E14-3; (B) mouse monoclonal 9A12(Dixit et al., 2007).

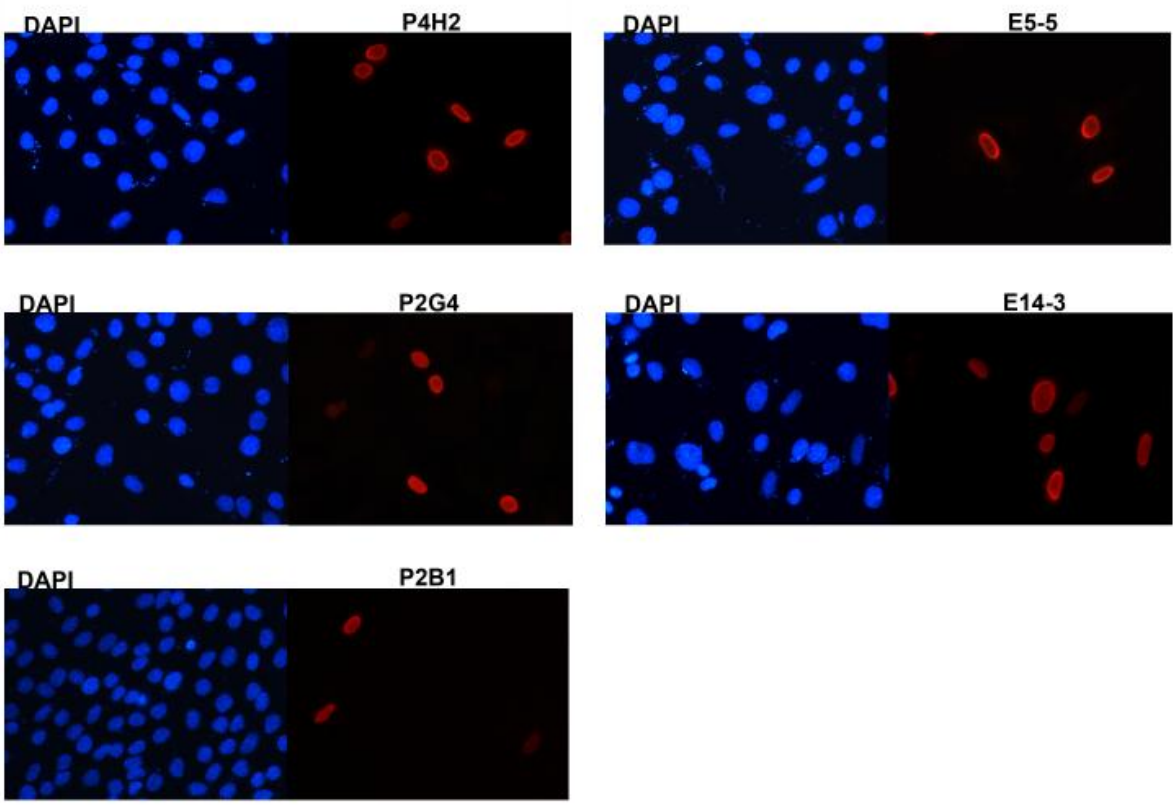


Figure 2.3. Reactivity of monoclonal antibodies on immunofluorescence.
C2C12 myoblasts were transfected with human DUX4 and cells were stained with mouse monoclonal anti-DUX4 antibodies P4H2, P2G4 and P2B1 (left), and rabbit monoclonal anti-DUX4 antibodies E5-5 and E14-3 (right). Cells were counterstained with DAPI for nuclei.

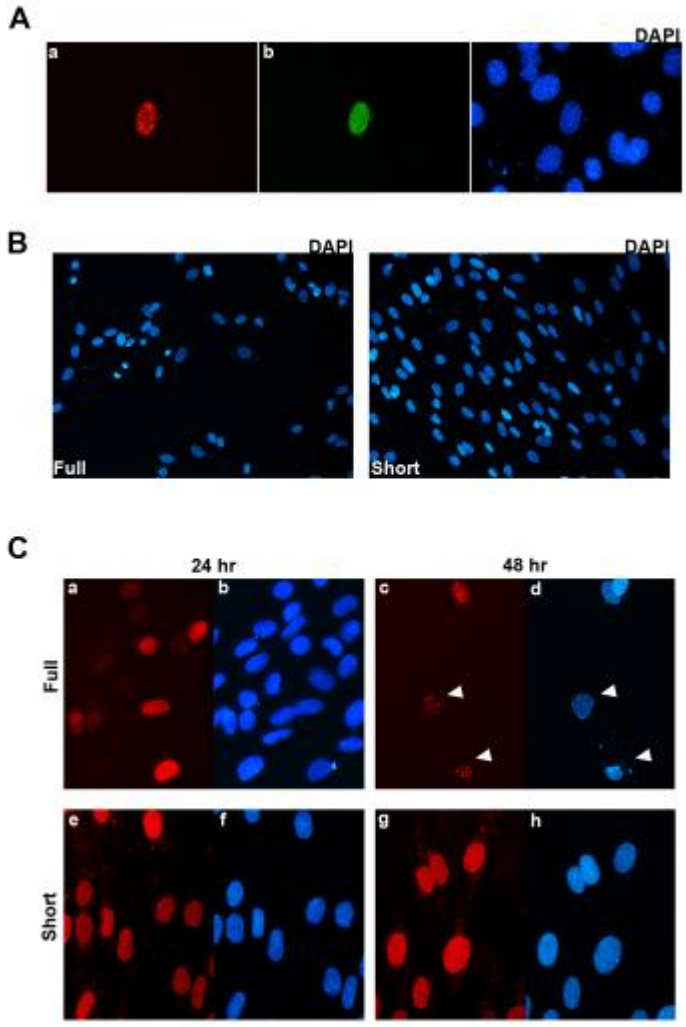


Figure 2.4. Effects of DUX4 in muscle cells.
(A) C2C12 cells transfected with pclBABE+DUX4 for 42 hr and co-immunostained with E14-3 (a) and P4H2 (b). (B) Equal number of wild-type human primary myoblasts were infected with lentivirus carrying either full length DUX4 (left) orshort DUX4 (right) and DAPI stained for nuclei 48 hr post-infection. (C) Lentiviral infection of human primary myoblasts with either full length or short DUX4 for 24 hr (a-b, e-f) or 48 hr (c-d, g-h). Cells were stained with E14-3 (red) and counterstained with DAPI (blue). Cells undergoing apoptosis are indicated by white arrowheads (c, d).

Chapter 3: FSHD: incomplete suppression of a retrotransposed gene

This chapter has been previously published:

Snider L*, **Geng LN***, Lemmers RJ, Kyba M, Ware CB, Nelson AM, Tawil R, Filippova GN, van der Maarel SM, Tapscott SJ, Miller DG. (2010) Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet.* Oct 28;6(10):e1001181.

*These authors contributed equally

Summary

Each unit of the D4Z4 macrosatellite repeat contains a retrotransposed gene encoding the DUX4 double-homeobox transcription factor. Facioscapulohumeral dystrophy (FSHD) is caused by deletion of a subset of the D4Z4 units in the subtelomeric region of chromosome 4. Although it has been reported that the deletion of D4Z4 units induces the pathological expression of DUX4 mRNA, the association of DUX4 mRNA expression with FSHD has not been rigorously investigated, nor has any human tissue been identified that normally expresses DUX4 mRNA or protein. We show that FSHD muscle expresses a different splice form of DUX4 mRNA compared to control muscle. Control muscle produces low amounts of a splice form of DUX4 encoding only the amino-terminal portion of DUX4. FSHD muscle produces low amounts of a DUX4 mRNA that encodes the full-length DUX4 protein. The low abundance of full-length DUX4 mRNA in FSHD muscle cells represents a small subset of nuclei producing a relatively high abundance of DUX4 mRNA and protein. In contrast to control skeletal muscle and most other somatic tissues, full-length DUX4 transcript and protein is expressed at relatively abundant levels in human testis, most likely in the germ-line cells. Induced pluripotent (iPS) cells also express full-length DUX4 and differentiation of control iPS cells to embryoid bodies suppresses expression of full-length DUX4, whereas expression of full-length DUX4 persists in differentiated FSHD iPS cells. Together, these findings indicate that full-length DUX4 is normally expressed at specific developmental stages and is suppressed in most somatic tissues. The contraction of the D4Z4 repeat in FSHD results in a less efficient suppression of the full-length DUX4 mRNA in skeletal muscle cells. Therefore, FSHD represents the first human disease to be associated with the incomplete developmental silencing of a retrogene array normally expressed early in development.

Introduction

Facioscapulohumeral dystrophy (FSHD) is an autosomal dominant muscular dystrophy caused by the deletion of a subset of D4Z4 macrosatellite repeat units in the

subtelomeric region of 4q on the 4A161 haplotype (FSHD1; OMIM 158900) (Lemmers et al., 2007). The unaffected population has 11-100 D4Z4 repeat units, whereas FSHD1 is associated with 1-10 units (Tawil and Van Der Maarel, 2006). The retention of at least a portion of the D4Z4 macrosatellite in FSHD1 and the demonstration that the smaller repeat arrays have diminished markings of heterochromatin (Zeng et al., 2009) support the hypothesis that repeat contraction results in diminished heterochromatin-mediated repression of a D4Z4 transcript, or a transcript from the adjacent subtelomeric region. The hypothesis that derepression of a regional transcript causes FSHD is further supported by individuals with the same clinical phenotype and decreased D4Z4 heterochromatin markings but without a contraction of the D4Z4 macrosatellite in the pathogenic range (FSHD2) (de Greef et al., 2009; van Overveld et al., 2003).

The D4Z4 repeat unit contains a conserved open reading frame for the DUX4 retrogene, which Clapp et al suggest originated from the retrotransposition of the DUXC mRNA (Clapp et al., 2007), a gene present in many mammals but lost in the primate lineage. Dixit et al (Dixit et al., 2007) demonstrated that DUX4 transcripts were present in cultured FSHD muscle cells and mapped a polyadenylation site to the region telomeric to the last repeat, a region referred to as pLAM. Lemmers et al (Lemmers et al.) recently demonstrated that the region necessary for a contracted D4Z4 array to be pathogenic maps to this polyadenylation site, which is intact on the permissive 4A chromosome but not on the non-permissive chromosomes 4B or 10, indicating that stabilization of the DUX4 mRNA is necessary to develop FSHD on a contracted allele. Our prior study (Snider et al., 2009) demonstrated bidirectional transcription of the D4Z4 region associated with the generation of small RNAs, and we suggested that these D4Z4-associated small RNAs might contribute to the epigenetic silencing of D4Z4. We also identified alternatively spliced transcripts from the DUX4 retrogene that terminate at the previously described (Dixit et al., 2007) polyadenylation site in the pLAM region. However, we identified DUX4 mRNA transcripts in both FSHD and wild-type muscle cells, as well as similar amounts of D4Z4-generated small RNAs.

Together these studies implicate a stabilized DUX4 mRNA transcript from the contracted D4Z4 array as the cause of FSHD. However, several important questions remain to be addressed:

(1) Our prior study identified two alternative splice forms of the DUX4 mRNA, which in this report we call DUX4-fl and DUX4-s, and showed that both control and FSHD muscle with a 4A chromosome contained polyadenylated DUX4 mRNA. Therefore, it is important to determine whether the overall abundance of the DUX4 mRNA or the relative abundance of the alternative splice forms is associated with FSHD.

(2) All studies reporting DUX4 mRNA associated with FSHD have used high cycle PCR to detect mRNA that are present at extremely low abundance. It remains to be determined whether the amount of DUX4 mRNA detected in FSHD cells makes sufficient DUX4 protein to have a biological consequence.

(3) DUX4 has been referred to as a pseudogene and the D4Z4 region has been referred to as "junk" DNA. The conclusion that DUX4 is not a functional gene is supported only by the absence of evidence that the DUX4 mRNA and protein is normally expressed in any human tissue. Yet, the open reading frame (ORF) of DUX4 is conserved, raising the possibility that it might have an as yet undetected role in human biology.

In this study, we address each of these important questions. Together, our data substantiate a developmental model for FSHD: full-length DUX4 mRNA is normally expressed early in development and is suppressed during cellular differentiation, whereas FSHD is associated with the failure to maintain complete suppression of full-length DUX4 expression in differentiated skeletal muscle cells. Occasional escape from repression results in the expression of relatively large amounts of DUX4 protein in a small number of skeletal muscle nuclei.

Results

Alternative DUX4 mRNA splicing distinguishes control and FSHD muscle

A recent study (Lemmers et al.) demonstrated that the sequence polymorphisms of the 4A161 haplotype necessary for FSHD include the region of the poly-adenylation signal for the DUX4 mRNA and showed that this correlated with the detection of DUX4 mRNA in three FSHD muscle cultures compared to controls. Our previous study of RNA transcripts from D4Z4 repeat units identified a full-length mRNA transcript that contains the entire DUX4 open reading frame and has one or two introns spliced in the 3-prime UTR, and a second mRNA transcript utilizing a cryptic splice donor in the DUX4 ORF that maintains the amino-terminal double-homeobox domains and removes the carboxyterminal end of DUX4 (Fig. 3.1A and B). We will refer to these two transcripts as DUX4-fl (full length) and DUX4-s (shorter ORF), respectively (see [9] for splice junction sequences). The PCR approach in the Lemmers et al study (Lemmers et al.) would not have detected the DUX4-s mRNA.

We used oligo-dT primed cDNA and a PCR strategy that would detect both DUX4-fl and DUX4-s (see Fig. 3.1B) to determine the presence of polyadenylated DUX4 mRNAs in quadriceps muscle needle biopsies from ten FSHD and fifteen control individuals (Table 3.1 and Fig. 3.1C). In general we used two cycles of PCR with nested primers to increase specificity and to detect low abundance transcripts. DUX4-fl was detected in five of the ten FSHD samples, based on primers amplifying DUX4-fl and primers amplifying the 3-prime region of DUX4-fl (DUX4-fl3') that is contained in DUX4-fl but not in DUX4-s (see Fig. 3.1B). The sequenced products matched the FSHD-permissive 4A161 haplotype polymorphisms and the variation in size of the PCR product reflected alternative splicing of only the second intron in the UTR or both the first and second UTR introns (see Fig. 3.1B). In contrast, none of the fifteen control samples expressed mRNA that amplified with primers to DUX4-fl or DUX4-fl3', including seven biopsies from individuals with at least one 4A161 chromosome. Instead, DUX4-s was detected in all control samples with 4A161 and in some of the FSHD samples. We did not detect DUX4 transcripts using these primers in six control

biopsies that do not contain the 4A chromosome. These data indicate that the 4A D4Z4 region is actively transcribed and produces alternatively spliced and polyadenylated DUX4 mRNA in both FSHD and unaffected individuals. However, the full-length DUX4 mRNA was only detected in the FSHD muscle biopsies, whereas DUX4-s was detected in muscle from controls and some FSHD individuals.

The expression of DUX4-fl mRNA in FSHD muscle biopsies could be a primary consequence of the D4Z4 contraction or a secondary response to the inflammation associated with muscle degeneration and/or regeneration. Therefore, we extended our analysis to myoblast cultures derived from four control and six FSHD individuals, including one individual with FSHD2. As seen in the muscle biopsies, the control muscle cells contained no detectable amounts of DUX4-fl mRNA, whereas muscle cells derived from both FSHD1 and FSHD2 samples expressed DUX4-fl transcripts as well as the DUX4fl-3' (Table 3.2 and Fig. 3.1D). All control and a subset of the FSHD samples expressed DUX4-s. These data are consistent with observations made in the muscle biopsies and indicate that both FSHD and control muscle cells actively transcribe DUX4. Unaffected cells produce DUX4-s from a splice donor site in the DUX4 ORF, whereas FSHD cells produce DUX4-fl with an alternative splice donor site after the translation termination codon of the DUX4 ORF.

A small fraction of FSHD muscle cells produce a relatively large amount of DUX4

In both control and FSHD cells the DUX4 mRNA transcripts, either DUX4-fl or DUX4-s, were only detected after nested PCR amplifications, indicating very low abundance of DUX4 mRNA in the FSHD and control biopsies and cells. We used the 9A12 mouse monoclonal anti-DUX4 antibody (Dixit et al., 2007) and also produced mouse and rabbit monoclonal antibodies to the amino-terminal and carboxyterminal portion of the DUX4 protein (Geng et al., 2011), but were unable to detect DUX4 protein in western analysis of FSHD muscle cultures, consistent with the very low amounts of DUX4 mRNA.

Low transcript abundance could reflect a small number of transcripts in every cell or a large number of transcripts in a small subset of cells in the population. We

assessed the presence of DUX4-fl mRNA in samplings of 100, 600, and 10,000 FSHD cultured muscle cells. DUX4-fl mRNA was present in five-out-of-ten pools of 600 cells (Fig. 3.2A) and three-out-of-20 pools of 100 cells (data not shown), as well as in the single pool of 10,000 cells. This frequency of positive pools indicates that approximately one-out-of-1000 cells is expressing a relatively abundant amount of DUX4-fl mRNA at any given time. Immunostaining of cultured FSHD and control cultured muscle cells with four independent anti-DUX4 monoclonal antibodies showed that approximately one-out-of-1000 nuclei co-stained with an antibody to the amino-terminus and an antibody to the carboxy-terminus of DUX4 (Fig. 3.2B), whereas no nuclei in the control cultures showed staining.

Both the mRNA analysis and the immunodetection indicate that approximately 0.1% of FSHD muscle nuclei express DUX4 mRNA and protein. This could represent transient bursts of expression or stochastic activation of expression that leads to cell death, or both. Forced expression of DUX4 has been shown to induce apoptosis in muscle cells (Bosnakovski et al., 2009; Snider et al., 2009). When DUX4 is expressed in control human muscle cells by lenti-viral delivery, the DUX4 protein is distributed relatively homogeneously during the first 24 hrs and then aggregates in nuclear foci at 48 hrs when the cells are undergoing apoptosis (Fig. 3.2C, panels c and d). These DUX4 nuclear foci associated with apoptosis are present in the nuclei of FSHD muscle cultures (compare panel d in Fig. 3.2C with panels a-f in Fig. 3.2B). Expression of DUX4-s in control human muscle cells does not induce apoptosis and does not accumulate in nuclear foci at 48 hrs (Fig. 3.2, panel e). Therefore, the data indicates that FSHD muscle cells that express endogenous full-length DUX4 also exhibit the nuclear foci that are characteristic of DUX4-induced apoptosis.

DUX4 mRNA and protein are expressed in human testis

Although there is no known function of DUX4 in human biology, the open reading frame has been conserved (Clapp et al., 2007). DUX4 is a retrogene thought to be derived from DUXC (Clapp et al., 2007), or a DUXC-related gene, but also similar to the DUXA family mouse Duxbl gene (Wu et al.). Therefore, if DUX4 has a biological

function it is likely to be similar to DUXC or Duxbl. Duxbl is expressed in mouse germ-line cells and we reasoned that because retrotranspositions entering the primate lineage must have occurred in the germ-line, then DUXC must be expressed in the germ-line. Indeed, we detect the canine DUXC mRNA in canine testis but not in canine skeletal muscle (data not shown). Therefore, if DUX4 has a biological function similar to DUXC, we would anticipate DUX4 expression in the human germ-line.

We obtained RNA from different adult human tissues and identified DUX4-fl in testis (Fig. 3.3A), whereas DUX4-s was present in a subset of differentiated tissues. DUX4-fl was detected in six additional testis samples, whereas only DUX4-s was detected in donor-matched skeletal muscle (Fig. 3.3B and C). Quantitative PCR (qPCR) showed that human testis samples expressed almost 100-fold higher amounts of DUX4 mRNA compared to FSHD muscle biopsies, and almost 15-fold higher amounts compared to cultured FSHD muscle cells (Fig. 3.3D). Western analysis using three different DUX4 antibodies identified a protein of the correct mobility in protein lysates from testes but not in other cells or tissues that do not express DUX4-fl mRNA, including control muscle cells (Fig. 3.3E and data not shown). Furthermore, immunoprecipitation of testis proteins with rabbit anti-DUX4 antibodies followed by western with a mouse monoclonal antibody to DUX4 detected the same protein (Fig. 3.3F). Western analysis of protein extracts from three additional human testis samples identified a similar band (data not shown). Immunostaining identified DUX4-expressing cells near the periphery of the seminiferous tubule that have the large round nucleus characteristic of spermatogonia or primary spermatocytes (Fig. 3.4a-c), and additional more differentiated appearing cells in the seminiferous tubules were also stained following antigen retrieval (Fig. 3.4e). The large numbers and nuclear morphology of the cells staining with DUX4 in the seminiferous tubules, together with expression of DUX4 in the human germ-cell cell lines SuSa and 833K (data not shown), leads us to conclude that DUX4 is expressed in the germ-line lineage. Further studies will be necessary to determine more precisely the timing and cell stages of DUX4 expression in the in the testis and to ascertain whether it has a biological function.

Chromosomes 4 and 10 produce DUX4 mRNA in human testes

The relatively high abundance of DUX4 mRNA and protein in human testes suggests a possible role for this protein in normal development. However, we have previously demonstrated that the alleles of chromosome 4 and 10 that are non-permissive for FSHD contain polymorphisms that inhibit polyadenylation of the DUX4 transcript, and, therefore, only the 4A allele would be predicted to make a DUX4 mRNA. We do not have haplotype information on the testis donors and it is possible that some might lack the 4A haplotype entirely. To determine whether only the 4A haplotype produced stable DUX4 mRNA in human testes, we sequenced mRNAs from the seven testis samples in a region with informative polymorphisms regarding transcripts from 4A, 4B, and 10. All testis mRNA had transcripts from both chromosomes 4 and 10 in approximately equal amounts (Table 3.3) based on the informative polymorphisms (Table 3.4). Some samples had 4A and 4B haplotypes.

3-prime RACE analysis on testis mRNA demonstrated that the chromosome 10 transcripts used alternative 3-prime exons with a polyadenylation signal in exon 7 that is approximately 6.5 kb further telomeric than the previously identified 4A polyadenylation site in the pLAM region (Fig. 3.5). Some 4A transcripts also use the exon 7 polyadenylation site, but the exon 3 polyadenylation site associated with the permissive allele is preferred (data not shown). The 4B transcripts do not use either the exon 3 or exon 7 polyadenylation sites since the 4B haplotype lacks these regions, however, we have not yet identified the full 3-prime sequence of the DUX4 mRNA from the 4B chromosome. Re-analysis of the muscle cell line, muscle biopsy, and somatic tissue transcripts did not identify any DUX4 mRNA utilizing the exon 7 polyadenylation site from either chromosome 10 or 4, including a control sample with a contraction to 9 copies of D4Z4 on chromosome 10 (biopsy 2318 in Table 3.1, data not shown). We conclude that chromosome 10 DUX4 transcripts in the testes use a distal exon 7 polyadenylation signal, whereas this region is not used in somatic tissues, even when the chromosome 10 D4Z4 array has contracted to ten repeats. Therefore, polyadenylated DUX4 mRNA from chromosomes 4 and 10 are present in the testis, but only chromosome 4A produces polyadenylated transcripts in somatic tissues.

Developmental regulation of alternative splicing suppresses DUX4-fl from chromosome 4

The expression of DUX4-fl mRNA in unaffected human testes and the expression of DUX4-s in some unaffected somatic tissues, including skeletal muscle, suggested a developmental regulation of splice site usage in the DUX4 transcript. To directly determine whether the transition between DUX4-fl and DUX4-s expression is developmentally regulated, we generated induced pluripotent stem (IPS) cells from FSHD and control fibroblasts by expression of *SOX2*, *OCT4*, and *KLF4* transcription factors from Moloney murine leukemia virus vectors (Takahashi et al., 2007). Stem-cell clones had normal karyotypes, exhibited the expected cellular and colony morphology, contained tissue non-specific alkaline phosphatase activity, and expressed embryonic antigens (Fig. 3.6A). RT-PCR demonstrated expression of stem cell markers *NANOG*, *HTERT*, *cMYC*, and endogenous transcripts from *OCT4*, *SOX2*, and *KLF4* (Fig. 3.6B). Pluripotency was demonstrated by the ability to form teratomas containing tissues derived from ectoderm, endoderm, and mesoderm (See Fig. 3.6A). We used these characterized IPS cells to determine the expression of DUX4-fl and DUX4-s in the parental fibroblasts, undifferentiated IPS cells, and in the IPS cells after differentiation into embryoid bodies.

DUX4-s, but not DUX4-fl, was detected in control fibroblasts. In contrast, IPS cells derived from the control fibroblasts expressed DUX4-fl, whereas differentiation of these cells to embryoid bodies resulted in a switch to the expression of DUX4-s and loss of DUX4-fl transcripts (Table 3.2 and Fig. 3.6C). In contrast, DUX4-fl was detected in FSHD fibroblasts and the IPS cells and embryoid bodies derived from FSHD fibroblasts. As expected, DUX4-fl3' was detected in samples expressing DUX4-fl. (The relative amounts of DUX4-fl in a subset of iPS cells is shown in Fig. 3.3D and a band migrating at the size of DUX4 was detected on a western with an anti-DUX4 antibody (data not shown)). DUX4-fl was detected in some human ES cell lines, but at much lower levels compared to the iPS cells (data not shown).

All of the splice donor and acceptor sites in the multiple alternative splicing events in the 3-prime UTR have consensus splice donor and acceptor sequences. In contrast, the splice donor in the ORF that produces DUX4-s is a non-canonical donor sequence and would normally not be favored for splicing. Recent studies have indicated

that repressive chromatin modifications can favor splice donor usage (Luco et al.) and we tested whether the degree of H3K9me3 correlated with the usage of the DUX4-s splice site. Chromatin immunoprecipitation showed that the control fibroblasts and embryoid bodies with DUX4-s expression had relatively higher levels of trimethylation of lysine 9 in histone H3 (H3K9me3), a repressive chromatin modification, compared to the control IPS cells, which express DUX4-fl (Fig. 3.6D). The FSHD cells maintained relatively low levels of H3K9me3 in both IPS and differentiated cells. These findings are consistent with previous studies showing decreased H3K9me3 at the D4Z4 region in FSHD1 and FSHD2 (Zeng et al., 2009) and suggest a correlation between the relatively higher levels of repressive chromatin modifications and the use of the cryptic splice donor to produce DUX4-s.

Discussion

We note that prior studies reported the presence of polyadenylated DUX4 transcripts in a small number of samples of cultured FSHD muscle cells but not in control muscle cells (Dixit et al., 2007; Lemmers et al.). Our study both confirms and significantly extends these prior studies by (a) including a larger number of FSHD muscle cell cultures, (b) assaying controls that have a permissive 4A chromosome and non-permissive 4B chromosomes, (c) extending the analysis to mRNA from primary muscle biopsies of FSHD and haplotype-matched controls, (d) identifying the DUX4-s splice form of the DUX4 mRNA in control cells and showing that the qualitative difference between control and affected muscle is splice-site usage and not production of DUX4 mRNA; (e) demonstrating that the very low abundance of DUX4 mRNA in FSHD muscle represents a small percentage of nuclei with relatively high abundance mRNA and protein; (f) demonstrating that relatively high amounts of the DUX4 mRNA are expressed in the human testes and pluripotent cells and that developmental regulation is achieved by a combination of chromatin-associated splice-site usage and polyadenylation site usage.

Together our data provide the basis for a specific model of FSHD pathophysiology: (1) full-length DUX4 is produced from the last D4Z4 unit in early stem cells; (2) in differentiated tissues, the D4Z4 array is associated with increased repressive H3K9me3 and DUX4 expression is repressed; (3) in the residual transcripts that escape repression, an alternative first-intron splice donor is utilized to produce DUX4-s instead of DUX4-fl; (4) contraction of the D4Z4 arrays impedes the conversion to repressive chromatin and the transition from DUX4-fl to DUX4-s, resulting in expression of the full-length DUX4 in skeletal muscle and possibly other tissues; and (5) the very low levels of full-length DUX4 expression in FSHD muscle reflects relatively high amounts of expression in a small sub-population of cells. Several groups have shown that expression of full-length DUX4 in muscle cells can induce pathologic features of apoptosis and expression of PITX1 (Bosnakovski et al., 2009; Bosnakovski et al., 2008b; Dixit et al., 2007; Kowaljow et al., 2007). In contrast, expression of DUX4c, a DUX4-like protein that lacks the carboxyterminal portion of DUX4, does not induce apoptosis (Bosnakovski et al., 2008a). Therefore, it is reasonable to believe that expression of DUX4-fl might induce muscle cell damage in FSHD, whereas DUX4-s expression would not be harmful to the cells. Indeed, FSHD muscle cells expressing the endogenous DUX4 have nuclear foci of DUX4 characteristic of the foci that appear during early stages of apoptosis when DUX4 is exogenously expressed in human skeletal muscle cells (see Fig. 3.2), suggesting, but not yet proving, that these DUX4 expressing cells might be initiating a process of nuclear death.

The observed association of decreased H3K9me3 of D4Z4 with detectable levels of DUX4-fl mRNA suggests a specific mechanism of regulating DUX4 splicing. Previously (Snider et al., 2009), we demonstrated bidirectional transcription of the D4Z4 repeats with the generation of small si/mi/pi-like RNA fragments and suggested that the small RNAs generated from D4Z4 might function to suppress DUX4 expression in a developmental context, a suppression mechanism observed for other retrogenes (Booth and Holland, 2007; Tam et al., 2008; Watanabe et al., 2008). A recent publication demonstrated that the small RNAs mediating heterochromatin formation also regulate splice-donor usage, either by targeting the nascent transcripts or by altering the rate of

polymerase progression through condensed chromatin (Allo et al., 2009; Luco et al.). Therefore, the repressive chromatin associated with D4Z4 in differentiated cells might facilitate the usage of the non-canonical splice donor to generate DUX4s, either through siRNAs from the region or through the impediment of polymerase progression, whereas the more permissive chromatin in FSHD and pluripotent cells might favor polymerase progression through to the consensus splice donor and generate DUX4-fl.

A recent study by Lemmers et al (Lemmers et al., 2010a) identifies sequence variants on 4A necessary to produce polyadenylated DUX4 mRNA transcripts in somatic tissues. Our results are consistent with these findings since we have not been able to identify polyadenylated transcripts from non-permissive alleles in somatic tissues. In contrast, we do find alternative distal polyadenylation usage for DUX4 mRNA from non-permissive alleles in the testis. Developmentally regulated polyadenylation site usage has been described for other genes (Ji and Tian, 2009) and appears to be one additional mechanism of silencing expression of the DUX4 retrogene in somatic cells.

Our finding that the wild-type chromosomes 4 and 10 express a full-length DUX4 mRNA in human testes, most likely in the germ-line, and that the protein is relatively abundant suggests that DUX4 might have a normal role in development. This is supported by the expression of canine DUXC in germ-line tissue (L. Geng, unpublished data). In addition, a DUX4-like gene in the mouse, Duxbl, is expressed in mouse germ-line cells in both spermatogenesis and oogenesis, as well as in early phases of skeletal muscle development (Wu et al., 2010). Similar to DUX4, Duxbl has developmentally regulated splicing to produce a full-length protein and a protein truncated after the double homoeodomains and studying the roles of Duxbl in germ-line and muscle development in mouse will likely inform our understanding of DUX4. We should note that our study describes the expression of human DUX4 in testes but we believe it is likely to be expressed in oogenesis as well. Limited access to appropriate tissue has limited our ability to carefully examine expression in cells of the ovary.

Generating new genes through retrotransposition is a common mechanism of mammalian evolution (Kaessmann et al., 2009), particularly for genes with a role in

germ cell development. Recently an FGF4 retrogene was identified as causing the short-legged phenotype in many dog breeds (Parker et al., 2009), indicating that retrogenes can direct dramatic phenotypic evolution in a population. Our study demonstrates that the expression of the DUX4 retrogene is developmentally regulated and might have a role in germ-line development, and, if similar to Duxbl, possibly in aspects of early embryonic muscle development. Maintaining the DUX4 retrogene in the primate lineage suggests some selective advantage compared to maintaining the parental gene itself. Based on current knowledge, this could be due to a function in germ-line development, or to a modulation of muscle mass in primate face and upper extremity. In this regard, it is interesting to speculate that a normal function of the DUX4 retrogene might be to regulate the development of facial and upper-extremity muscle mass in the primates, and that FSHD represents a hypermorphic phenotype secondary to inefficient developmental suppression. Alternatively, the persistent expression of full-length DUX4 might induce a neomorphic phenotype unrelated to an evolutionarily selected role of DUX4. In either case, our findings substantiate a comprehensive developmental model of FSHD and demonstrate that FSHD represents the first human disease to be associated with the incomplete developmental silencing of a retrogene array that is expressed in pluripotent stem cells and in normal development.

Materials and Methods

Muscle biopsies, cultures, and human RNA and protein: Muscle biopsy samples were collected from the vastus lateralis muscle of clinically affected and control individuals using standardized needle muscle biopsy protocol and cell cultures were derived from biopsies as described on the Fields Center website: http://www.urmc.rochester.edu/fields-center/protocols/documents/PreparingPrimaryMyoblastCultures.pdf. The sex, age, and severity score for the FSHD muscle biopsies were: F1998 (M, 43, 2); F0519 (M, 43, 4); F0515 (F, 48, 2); F0509 (M, 47, 2); F0531 (F, 47, 2); F2306 (F, 46, ND); F2331 (F, 56, 4); F2316 (F, 34, 5); F2319 (M, 52, ND); F2315 (F, 40, 3). Pathologic grading scale is 0-12

(from normal to severe) based on a score of 0-3 for each of four parameters: muscle fiber size/shape; degree of central nucleation; presence of necrotic/regenerating fibers or inflammation; and degree of fibrosis. Controls were selected in the same age range and sex representation. Muscle cell culture MB216 and muscle biopsy F2316 are from the same individual, otherwise the muscle cultures were derived from other individuals. RNA and protein lysates from human tissues were purchased from BioChain (Hayward, CA) and Origene (Rockville, MD).

RT-PCR for DUX4-fl, DUX4-s, and DUX4-fl3'. Total RNA was isolated from

muscle biopsies and cultured cells using Trizol (Invitrogen) and then treated with DNase I for 15 minutes using conditions recommended by Invitrogen with the addition of RNaseOUT (Invitrogen) to the reaction. DNase reaction components were removed using the RNeasy (Qiagen) system and RNA eluted by two sequential applications of 30 µl of RNase-free water. Volume was reduced by speed vac and 1.5-2 µg of RNA used for first strand cDNA synthesis. RNA from adult human tissues was purchased from Biochain and had been DNase-treated by the supplier. First strand synthesis was performed using Invitrogen SuperScript III reverse transcriptase and Oligo dT primers according to manufacturer's instructions at 55˚ for 1 hour followed by digestion with RNase H for 20 minutes at 37˚. Finally, the reactions were cleaned using the Qiaquick (Qiagen) pcr purification system and eluted with 50 µl of water. Primary pcr reactions were performed with 10% Invitrogen PCRx enhancer solution and Platinum Taq polymerase using 10-20% of the first strand reaction as template in a total reaction volume of 20 µl. Nested pcr reactions used 1 µl of the primary reaction as template. Primers for Dux4-fl and -s detection in biopsy and cultured cell samples were 14A forward and 174 reverse, nested with15A (or 16A) forward and 175 reverse. Primers for 3' detection were 182 forward and 183 reverse nested with1A forward and 184 reverse.

Dux4-fl and –s in adult human tissues were detected using 14A forward and 183 reverse, then nested with 15A forward and 184 reverse primers. Pcr cycling conditions were as follows for both primary and nested pcr: 94˚ 5 minutes denaturation, 35 cycles

of 94˚ for 30”, 62˚ for 30” and 68˚ for 2.5 minutes or 1 minute depending on expected length of product. A single final extension of 7 minutes at 68˚ was included. Pcr products were examined on 2% NuSieve GTG (Lonza) agarose gels in TBE.

Primers for DUX4-fl and DUX4-s

14A	5’ CCCCGAGCCAAAGCGAGGCCCTGCGAGCCT 3’	forward
174	5’ GTAACTCTAATCCAGGTTTGCCTAGA 3’	reverse
15A	5’ CGGCCCTGGCCCGGGAGACGCGGCCCGC 3’	forward
16A.	5’ GGATTCAGATCTGGTTTCAGAATCGAAGG 3’	forward
175	5’ TCTAATCCAGGTTTGCCTAGACAGC 3’	reverse

Primers for DUX4-fl3’

182	5’ CACTCCCCTGCGGCCTGCTGCTGGATGA 3’	forward
183	5’CCAGGAGATGTAACTCTAATCCAGGTTTGC 3’	reverse
1A	5’ GAGCTCCTGGCGAGCCCGGAGTTTCTG 3’	forward
184	5’ GTAACTCTAATCCAGGTTTGCCTAGACAGC 3’	reverse

Pooled PCR for DUX4. To assess for stochastic expression of DUX4 in affected muscle cells, FSHD primary myoblasts were trypsinized and collected at confluence or after differentiation for 96 hr. Cells were counted and split into pools of 100-cell, 600-cell, or 10,000-cell aliquots. RNA was extracted from individual aliquots using Dynabeads mRNA DIRECT Kit (Invitrogen) following manufacturer’s instructions. Bound polyadenylated mRNA was used directly for reverse transcription reaction with SuperScript III using on-bead oligo dT as primer. Synthesis was carried out at 52°C for 1 hr, terminated at 70°C for 15 min, followed by 15 min of RNase H treatment. 2 uL of cDNA product was used for nested DUX4-fl3’ PCR as described above.

RT-PCR for transcripts from chromosomes 10 and 4. Pcr reactions were performed on RT reactions generated as described above and using nested primer sets to sequences in exons 1 and 2 that are common to alleles on chromosomes 4 as well as 10.

Transcripts were detected using primers 1A and LS 187 followed by nesting with LS 138 and LS 188. Diagnostic polymorphisms (underlined) in the 5' end of exon 2 were used to assign allele origins of transcripts:

4A161, 159, 168 G<u>T</u>C<u>T</u>AGGCCCGGTGAGAGACTCCACAC<u>C</u>GCG

4A166 G<u>T</u>C<u>T</u>AGGCCCGGTGAGAGACTCCACAC<u>A</u>GCG

10A 166 G<u>C</u>C<u>T</u>AGGCCCGGTGAGAGACTCCACAC<u>A</u>GCG

4B163 G<u>T</u>C<u>C</u>AGGCCCGGTGAGAGACTCCACAC<u>C</u>GCG

Primers

138 5' CGGAGTTTCTGCAGCAGGCGCAACCTCTCCT 3' forward

187 5' CTGCTGGTACCTGGGCCGGCTCTGGGATCCC 3' reverse

188 5' GTACCTGGGCCGGCTCTGGGATCCCCGGGAT 3' reverse

Quantitative RT-PCR for DUX4-fl3'. For quantitative PCR, 1 ug of DNase'd RNA was used for first strand cDNA synthesis. Reverse transcription was performed as above, except at 52°C for the synthesis reaction followed by 15 minutes of RNase H treatment and the Qiaquick purification eluted in 30 µl of water. One round of PCR reactions were performed using the same reagents as above and 2 uL of purified cDNA template. Primers for full length detection were 92 forward and 116 reverse. PCR cycling conditions were as follows: 95°C 5 min denaturation, 36 cycles of 95°C for 30", 62°C for 30" and 68°C for 1 min, and final extension of 5 min at 68°C. Sequence of the product matched DUX4. A standard curve for DUX4 template copies was generated from PCR reactions using the same primers and cycling conditions but with known dilutions of a plasmid containing full length DUX4 cDNA in water. Test sample PCR reactions and standard PCR reactions were run in triplicate and examined on the same 1% agarose/TBE gels stained with SYBR Gold (Invitrogen) for 40 min per manufacturer instructions. Fluorescence was detected with Typhoon Trio Multi-mode Imager (GE Healthcare): excitation laser 488 nm; emission filter 520DP 40, PMT 500 V, 100 µm resolution. Histogram analysis was performed to ensure no signals were saturated. Gel band intensities were quantitated with ImageQuant TL v2005 (GE Healthcare) software. Estimates for the copies of DUX4 full length template in the test samples were

interpolated from the line of best fit of the dilutional standards, with the lowest visible dilutional signal setting the detection limit. The interpolated number was doubled to adjust for the single-stranded cDNA input in contrast to the double-stranded plasmid standard input. This resulted in an estimated copy number of DUX4 full-length per ug of total RNA. Final copy number estimates per cell were calculated based on assumptions of 100% efficient reverse transcription and 3.3 pg of total RNA per cell.

Primers for qPCR or DUX4-fl3'

92 5' CAAGGGGTGCTTGCGCCACCCACGT 3' forward
116 5' GGGGTGCGCACTGCGCGCAGGT 3' reverse

Open Reading Frame PCR for DUX4-fl. To assess for the full coding region of DUX4, three rounds of PCR were performed on cDNA, totaling 36 cycles. Conditions for each round were as follows: 95°C for 5', 3 cycles of 95°C for 30" and 68°C for 1'33", 3 cycles of 95°C for 30" and 65°C for 30" and 68°C for 1'33", 6 cycles of 95°C for 30" and 62°C for 30" and 68°C for 1'33". 3 uL of primary PCR was used in the secondary PCR, and 3 uL of secondary PCR were used in the tertiary PCR.

Primers for ORF PCR for DUX4-fl

133 5' ATGGCCCTCCCGACACCCTCGGACAGCACC 3' forward
134 5' CTCGGACAGCACCCTCCCCGCGGAAGCCCG 3' forward
135 5' GGAAGCCCGGGGACGAGGACGGCGACGGAG 3' forward
136 5' CTAAAGCTCCTCCAGCAGAGCCCGGTATTCTTCCTC 3' reverse
137 5' CCCGGTATTCTTCCTCGCTGAGGGGTGCTTCCAG 3' reverse
138 5' GGGGTGCTTCCAGCGAGGCGGCCTCTTCCG 3' reverse

3' RACE for Dux4 in human testes. 3' RACE was performed on total RNA using Invitrogen Gene Racer kit essentially as described. Prior to pcr with gene specific primers and the GeneRacer 3' primers the RT reaction was cleaned using Qiaquick

(Qiagen) spin columns as described above. Gene specific forward primers were 182 and 1A (nesting). Pcr products were gel purified, cloned into TOPO 4.0 (Invitrogen) and sequenced.

Generation of induced pluripotent stem (IPS) cells. IPS cells were generated by forced expression of human OCT4, SOX2, and KLF4 using the retroviral vectors essentially as previously described (1). MLV vectors (pMXs-hOCT4, pMXs-hSOX2, and pMXs-hKLF4) were purchased from Addgene (www.addgene.com, Cambridge, MA) and vector preparations were generated by transient transfection of Phoenix-GP cells (2) with pCI-VSV-G and vector plasmids (1:1 ratio), replacing the culture medium 16 and 48 hours later, harvesting and filtering (0.45 mm pore size) conditioned medium after a 16 hour exposure to cells, and concentrating 50 to 100-fold by centrifugation (3). Transduction with MLV vectors was performed with polybrene (4mg/ml concentration) (Sigma-Aldrich Corp., St. Louis, MO) added to the medium. IPS cell colonies were identified by their characteristic morphology, cloned by microdissection, and expanded on irradiated mouse embryo fibroblasts (6000 rads) for further characterization. Typically, 5×10^4 fibroblasts cultured in DMEM + 10% FBS were seeded to a 9.4 cm^2 well on day -1, the medium was replaced with medium containing vectors and polybrene on day 0, and changed again to medium with DMEM + 10% FBS on day 1. Cells were detached with trypsin and seeded to five 55 cm^2 dishes on day 2 and medium changed on day 4. On day 6 cells are again detached with trypsin and 5×10^5 cells seeded to 55 cm^2 dishes containing 7×10^5 irradiated mouse embryo fibroblasts (6000 rads) in human ES cell culture medium (see below). Medium is replaced every other day and colonies with typical morphology of IPS cells appear between day 20 and day 30 post infection. Colonies are mechanically dissected using drawn Pasteur pipettes and seeded to mouse embryo fibroblast feeder layers for culture and passaged every 2-3 days using 2 u/ml dispase.

Stem Cell Culture. IPS cells and Human ES cells were grown in a solution of DMEM:F12 (1:1) with 3.151 g / L glucose, supplemented with L-Glutamine (Invitrogen), non-essential amino acids (10 mM (100x) liquid, Invitrogen, # 11140-076), sodium pyruvate (100 mM (100x), liquid, # 11360-070), 20% knockout serum replacer (# 10828010)

(Invitrogen, Carlsbad, CA), 1mM b-mercapto Ethanol (Sigma, St. Louis MO), and 5 ng/ml basic fibroblast growth factor (Peprotech, #AF-100-18B). Cells were generally cultured in 0.1% gelatin coated dishes containing irradiated mouse embryo fibroblasts at a density of 1.3 x 10^4 cells / cm^2. When cells were used as a source of RNA, DNA, or protein, they were cultured on matrigel (1:60 dilution, BD biosciences, #356234) coated dishes in medium conditioned by exposure to confluent layers of mouse embryo fibroblasts over a 3 day period. Cells were passaged a minimum of 4 times under these conditions before DNA, RNA, or protein was harvested.

Detection of embryonic antigens in IPS cells. IPS cells were evaluated for the presence of tissue non-specific alkaline phosphatase activity by fixing colonies in phosphate buffered saline solution containing 0.5% gluteraldehyde, and washing x3 in PBS. A staining buffer containing 100 mM Tris pH 8.5, 100 mM NaCL, 50 mM $MgCl_2$, 0.1 mg / ml 5-Bromo-4-chloro-3-indolyl phosphate (xphos) and 1 mg / ml p-Nitro-Blue tetrazolium chloride (NBT) (Sigma-Aldrich, St. Louis, MO, USA) was used to detect tissue non-specific alkaline phosphatase activity. Stage Specific embryonic antigen 4 (SSEA4) was detected using mouse monoclonal MC-813-70 and goat anti-mouse FITC conjugated secondary. TRA-1-60 was detected using mouse monoclonal TRA-1-60 (Millipore, Billerica, MA), and goat anti-mouse FITC conjugated secondary (Millipore, Billerica, MA). Human NANOG was detected with a goat polyclonal flurophore (Northern Lights^{tm}) conjugated antibody (NL493, R & D systems, Minneapolis, MN). Human OCT4 was detected with a rabbit polyclonal (Abcam, Cambridge, MA) and goat anti rabbit secondary conjugated with the Alexa 488 flurophore (Invitrogen, Carlsbad, CA). Cell karyotypes were determined by the University of Washington Cytogenetics laboratory.

Teratoma formation and staining. Induced pluripotent stem cells were detached from culture dishes with dispase (2 units/ml working concentration), 2 x 10^6 cells resuspended in F12:DMEM (1:1 mixture) medium without supplements, and injected into the femoral muscle of SCID-Beige mice (CB17.B6-*Prkdc*^{scid}*Lyst*^{bg}/Crl Charles River, Stock # 250). Mice were maintained under biosafety containment level 2 conditions and palpable tumor masses developed approximately 6 weeks later. When a tumor mass

was palpable the mice were sacrificed and tumor tissue fixed for several days in phosphate buffered saline solution containing 4% formaldehyde, and imbedded in paraffin. Sections of the tumor (5 mM thickness) were placed on slides and stained with hematoxylin and eosin using standard protocols.

Embryoid body formation. Human IPS were prepared for embryoid body formation by expanding cell numbers on mouse irradiated feeder layers detaching colonies with dispase, triturating with a Pasteur pipette, and seeding colony fragments to dense layers of mouse embryo fibroblast feeders (5×10^4 irradiated mef / cm^2) prior to EB formation. Four days later densely grown colonies from a 55 cm^2 dish were treated with dispase and gently detached by pipetting or scraping. Colony fragments were washed several times and seeded (1:1) to Ultra Low Attachment 55 cm^2 culture dishes (Corning, Corning, NY) in DMEM supplemented with 20% Fetal Bovine Serum. Every three days, EB's were allowed to gravity settle and the medium was gently removed and replaced. RNA and chromatin was harvested three weeks later for analysis.

Analysis of gene expression in IPS cells. IPS cells were grown without MEF feeders for preparation of RNA to be used in gene expression analysis. Cells were seeded to matrigel coated dishes and filtered conditioned medium from mouse embryo fibroblasts was used for culture. RNA was purified from cells using standard techniques and treated with DNAse to remove residual genomic DNA from the cells. cDNA synthesis was primed with oligo dT and reverse transcriptase. In all cases a tube was processed in parallel without the addition of reverse transcriptase to serve as a control for possible DNA contamination. The presence of RNA transcripts were detected using 28 thermal cycles with the following primer pairs. RNA was replaced with water as a negative control for the reaction.

Gene	**Forward primer**	**Reverse primer**
OCT4	5'-gacaggggcaggggaggagctagg-3'	5'-cttccctccaaccagttgccccaaac
SOX2	5'-gctagtctccaagcgacgaa-3'	5'-gcaagaagcctctccttgaa-3'
hTERT	5'-cctgctcaagctgactcgacaccgtg-3'	5'-ggaaaagctggccctggggtggagc-3'
NANOG	5'-cagtctggacactggctgaa-3'	5'-ctcgctgattaggctccaac-3'

KLF4	5'-tatgacccacactgccagaa-3'	5'-tgggaacttgaccatgattg-3'
cMYC	5'-cggaactcttgtgcgtaagg-3'	5'-ctcagccaaggttgtgaggt-3'
GAPDH	5'-tgttgccatcaatgacccctt-3'	5'-ctccacgacgtactcagcg-3'

Chromatin Immunoprecipitation

The Chromatin Immunoprecipitation (ChIP) analysis of repressive histone modifications at the 5'-region of *DUX4* was performed on primary fibroblasts, induced pluripotent stem (iPS) cells and corresponding embryoid bodies (EB) derived from unaffected individuals and FSHD patients, following a previously described protocol (Nelson et al., 2006; Zeng et al., 2009). Briefly, cells were cross-linked with formaldehyde at 1.42% final concentration for 15 min at room temperature, quenched, and sonicated to generate 500-100 bp DNA fragments. 25 μg aliquots (representing approximately 500,000 cells) of chromatin were used for each immunoprecipitation with anti-Histone H3K9me3 antibodies (Abcam) and nonimmune IgG fraction used as a mock control. After reverse cross-linking and DNA purification, the IP products were analyzed by real time PCR. The 5'-region of the DUX4 gene was analyzed using the 4q-specific D4Z4 primers, 4qHox or Q-PCR, that detect internal D4Z4 units including the last repeat unit (Zeng et al., 2009). The real-time PCR signals obtained for IP antibodies were normalized to mock control IgG and to input to account for the number of D4Z4 repeats. Data are presented as mean ± stdev and represent the results of at least three independent immunoprecipitations followed by real-time PCR analysis done in triplicates.

Generation of antibodies to DUX4

We generated monoclonal antibodies to the amino- and carboxy-terminus of DUX4 for this study. The full characterization of these antibodies will be published separately (Geng et al, in preparation). Briefly, the N-terminal 159 amino acids and the C-terminal 76 amino acids of DUX4 were fused to glutathione-s-transferase tags, respectively, and injected into the animals as immunogens. Mouse monoclonals were produced at the Antibody Development core facility at the Fred Hutchinson Cancer Research Center and will be commercially available. Rabbit monoclonals were produced in collaboration

with and will be available through Epitomics (Burlingame, CA). Hybridoma clones were screened for specificity by ELISA, western blot and immunofluorescence in C2C12 myoblasts transfected with DUX4. The C-terminal antibodies P4H2, P2B1 and E5-5 are specific to DUX4 and do not recognize DUX4c, whereas the N-terminal antibodies P2G4 and E14-3 recognize both DUX4 and DUX4c.

Protein analysis

For western blotting, protein lysates were prepared by resuspension in standard Laemmli buffer and sonicated briefly. Equivalent amounts of test samples were loaded onto 4-12% gradient gel and transferred to nitrocellulose membrane, which were then blocked with 5% non-fat dry milk in PBS 0.1% Tween-20. Custom monoclonal antibodies (Epitomics) raised against DUX4 were used to probe the blots and detected by ECL reagent (Pierce). Membranes were stripped and reprobed with anti-α-tubulin antibody (Sigma-Aldrich) for loading control. Immunoprecipitation was performed on samples resuspended in PBS with protease inhibitor cocktail (Roche) by incubating overnight at 4°C with pooled anti-DUX4 rabbit monoclonal antibodies bound to protein A- and G-coupled Dynabeads (Invitrogen). Samples were eluted directly into Laemmli buffer and analyzed on western blot as described. For immunofluorescence, cells were fixed in 2% paraformaldehyde for 7 min and permeabilized in 1% Triton X-100 in PBS for 10 min at room temperature. Cells were probed with pairs of rabbit and mouse primary antibodies raised against N- or C-terminus of DUX4 diluted in PBS overnight at 4°C. Double labeling was detected with Alexa Fluor 488 goat anti-mouse IgG and Alexa Fluor 568 goat anti-rabbit IgG (Invitrogen) at 1:500 in PBS for 1 hr and counterstained with DAPI.

Dux4 IHC on Frozen tissue

Immunohistochemistry was performed by the FHCRC Experimental Histopathology Shared Resource. Six-micron sections of OCT embedded frozen de-identified human testes tissue were sectioned and fixed for 10 minutes in 10% neutral buffer formalin. The slides were rehydrated in TBS-T wash buffer, permeablized with 0.1% triton X-100 for 10 minutes, and then endogenous peroxidase activity was blocked with 0.3%

hydrogen peroxide (Dako, Carpinteria, CA) for 8 minutes. Five minute incubation in 50% acetone and 50% methanol was used for antigen retrieval on a subset of slides. Protein block containing 0.25% casein and 0.1% Tween 20 was applied for 10 minutes. Slides were incubated over night at 4C with a 1:5 dilution of either clone E5-5 or P2B1 in a 0.3 M NaCl antibody diluent containing 1% BSA. Staining was developed using Mach2 HRP-labeled polymers (Biocare Medical, Concord, CA). The staining was visualized with 3,3'-diaminobenzidine (DAB, Dako) for 8 minutes, and the sections were counter-stained with hematoxylin (Dako) for 2 minutes. Concentration matched isotype control slides were run for each tissue sample (Jackson ImmunoResearch).

Table 3.1. Dux4 mRNA Expression in FSHD and Control Biopsies

Biopsy Code	Status	Haplotypes[A]	Dux4-fl	Dux4-s	3-prime
F 1998	FSHD	A161(10)/n.d.	**XS**	0	**XS**
F 0519	FSHD	A161(8)/A161	**XS**	0	**XS**
F-0515	FSHD	A161(5)/A161	**XS**	0	**XS**
F-0509	FSHD	A161(5)/A166H	0	**XS**	0
F-0531	FSHD	A161(7)/A161	0	**XS**	0
F-2306	FSHD	A161(4)/B163	0	X	0
F-2331	FSHD	A161(3)/B163	X	0	**XS**
F-2316	FSHD	A161(6)/A168	X	0	**XS**
F-2319	FSHD	A161(5)/B168	0	X	0
F-2315	FSHD	A161(5)/B168	0	0	0
C22	contr	A161/B163	0	X	0
C34	contr	A161/B163	0	**XS**	0
C40	contr	A161/B168	0	X	0
2333	contr	A161/A161	0	X	0
2397	contr	A161/B168	0	X	0
2401	contr	A161/B168	0	X	0
2398	contr	A161/B163	0	X	0
C33	contr	A166/B170	0	X	0
C39	contr	A166/B168	0	0	0
C10	contr	B168/B168	0	0	0
C11	contr	B168/B163	0	0	0
C31	contr	B163/B163	0	0	0
2318*	contr	B169/B166	0	0	0
C20	contr	B168/B168	0	0	0
C38	contr	B162/B163	0	0	0

[A] Chrosome 4 haplotype. For FSHD the number in paratheses indicates the number of D4Z4 units on the contracted allele; n.d., indicates that the haplotype of the second allele was not determined.

X, product present; **XS,** product sequenced; 0, product absent

* has contracted 10qA allele with 9 repeats

Table 3.2. Dux4 mRNA Expression in FSHD and Control Cell Lines

Cell Code	Status	Haplotypes	Cell type[A]	Dux4-fl	Dux4-s	3-prime
MB73	FSHD	A161(8)[B]/161[C]	MB	X	X	X
MB183	FSHD	A161(5)/B163	MB	X	O	X
MB148	FSHD	A161(3)/A161	MB	X	O	X
MB200	FSHD2	A161(14)/B168	MB	X	**XS**	X
MB216	FSHD	A161(6)/A168	MB	O	O	O
MB219	FSHD	A161(5)/B168	MB	O	O	O
MB230	contr	A161/163[A]	MB	O	X	O
MB196	contr	A161/163[A]	MB	O	X	O
MB226	contr	A161/A161	MB	O	X	O
MB135	contr	A161/B163	MB	O	**XS**	O
MB73	FSHD	A161/161	MT	X	X	**XS**
MB183	FSHD	A161/B163	MT	X	O	X
MB148	FSHD	A161/A161	MT	X	O	**XS**
MB200	FSHD2	A161/B168	MT	**XS**	O	**XS**
MB216	FSHD	A161/A168	MT	X	O	**XS**
MB219	FSHD	A161/B168	MT	X	O	X
MB230	contr	A161/163[A]	MT	O	X	O
MB196	contr	A161/163[A]	MT	O	X	O
MB226	contr	A161/A161	MT	O	X	O
MB135	contr	A161/B163	MT	O	X	O
M83-9	contr	A161/168[A]	fibro	O	**XS**	O
M83-9	contr		IPS	**XS**	O	X
M83-9	contr		EB	O	X	O
43-1	FSHD	A161(5)/A161	fibro	**XS**	O	O
43-1	FSHD		IPS	**XS**	O	X
43-1	FSHD		EB	X	O	**XS**
83-6	FSHD	A161(7)/B168	fibro	**XS**	**XS**	X
83-6	FSHD		IPS	X	O	X
83-6	FSHD		EB	X	O	**XS**

[A] MB, myoblasts; MT, myotubes; fibro, fibroblasts; IPS, induced pleuripotent stem cells,
EB, embryoid bodies

[B] Number of repeats on contracted allele, if known

[C] Assignment of the second 4q allele variant is incomplete

n.d., not tested; X, product present; **XS**, product sequenced; O, product absent

Table 3.3. Haplotype identification of Dux4 mRNA in Human Testis

Sample Code	%10A[A]	% 4A	% 4B	Total Number[B]
T1	22	78	0	9
T2	60	40	0	10
T3: 8606	22	67	11	9
T4: H12817	56	0	44	9
T7: N30	56	22	22	9
T6: N21	20	80	0	10
T5: N11	22	78	0	9
Total	38	53	9	64

[A] Percentage of the sequences containing SNPs of each haplotype

[B] Number of sequenced cDNA

Table 3.4. Diagnostic polymorphisms in exon 2

genotype	sequence
4A161, 168, 159	GTCTAGGCCCGGTGAGAGACTCCACACCGCG
4A166	GTCTAGGCCCGGTGAGAGACTCCACACAGCG
10A166	GCCTAGGCCCGGTGAGAGACTCCACACAGCG
4B163	GTCCAGGCCCGGTGAGAGACTCCACACCGCG

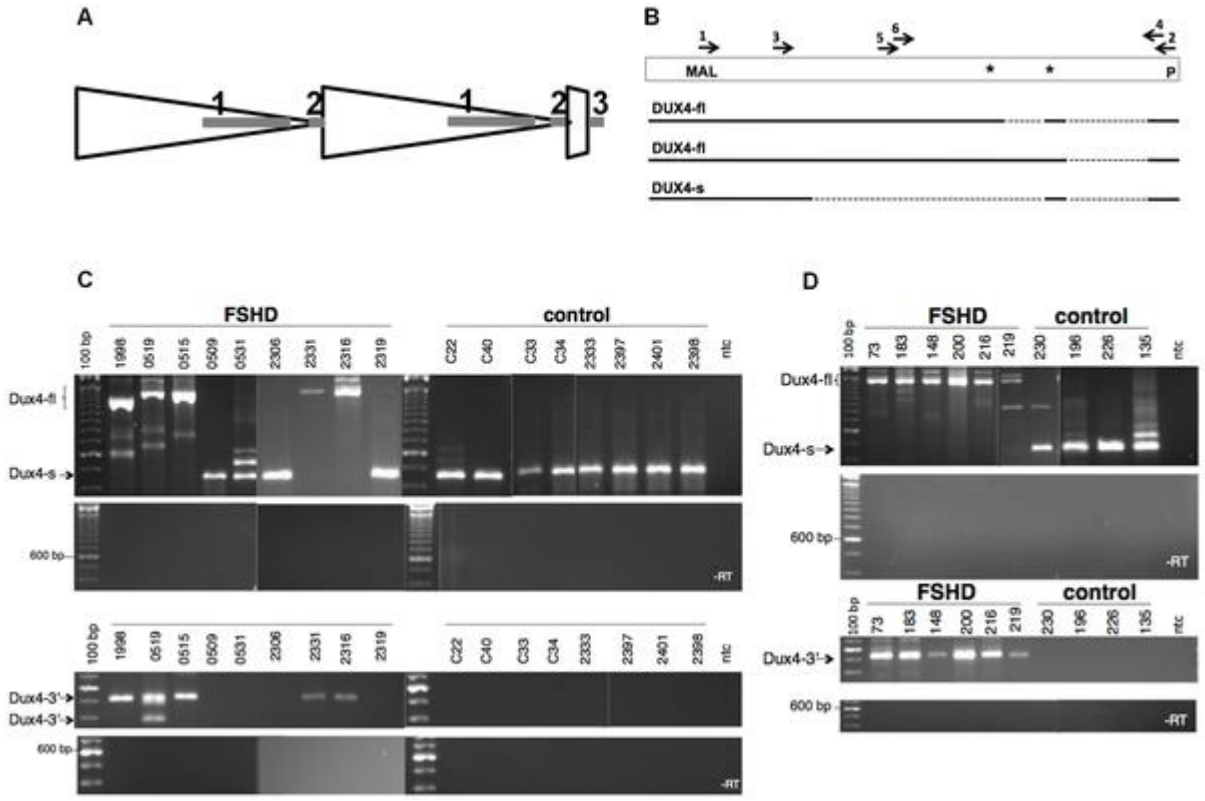


Figure 3.1. Expression of DUX4-fl and DUX4-s and D4Z4 in control and FSHD cells.
(A) Diagram of D4Z4 repeat array with two most telomeric full units (large triangles), the last partial repeat, and the adjacent pLAM sequence that contains exon 3. Exons shown as shaded rectangles, with exon 1 and 2 in the D4Z4 units and exon 3 in the pLAM region. (B) The open rectangle represents the region of D4Z4 and pLAM containing the DUX4 retrogene, and the solid and dashed lines represent the regions of exons and introns, respectively, in the short splice form (DUX4-s) and the transcript with the full-length DUX4 ORF (DUX4-fl), which has two isoforms with alternative splicing in the 3-prime untranslated region. First round PCR for DUX4-fl and DUX4-s was performed with primer sets 1 and 2 and second round PCR with nested primers 3 and 4. Nesting was used to ensure specificity and because of the very low abundance of DUX4 transcripts, both DUX4-fl and DUX4-s. MAL, represents location of initial amino-acid codons; *, Stop codons; P, polyadentylation site. (C) Composite image of representative PCR products from FSHD and control muscle biopsies for DUX4-fl, DUX4-s, and DUX4-fl3′. DUX4-fl and DUX4-s are indicated. Variation in size reflects alternative intron usage and the faint intermediate bands represent background non-DUX4 PCR amplicons frequently associated with repetitive sequence. (D) Representative PCR products of cultured FSHD and control muscle cells under differentiation conditions. ntc, no template control.

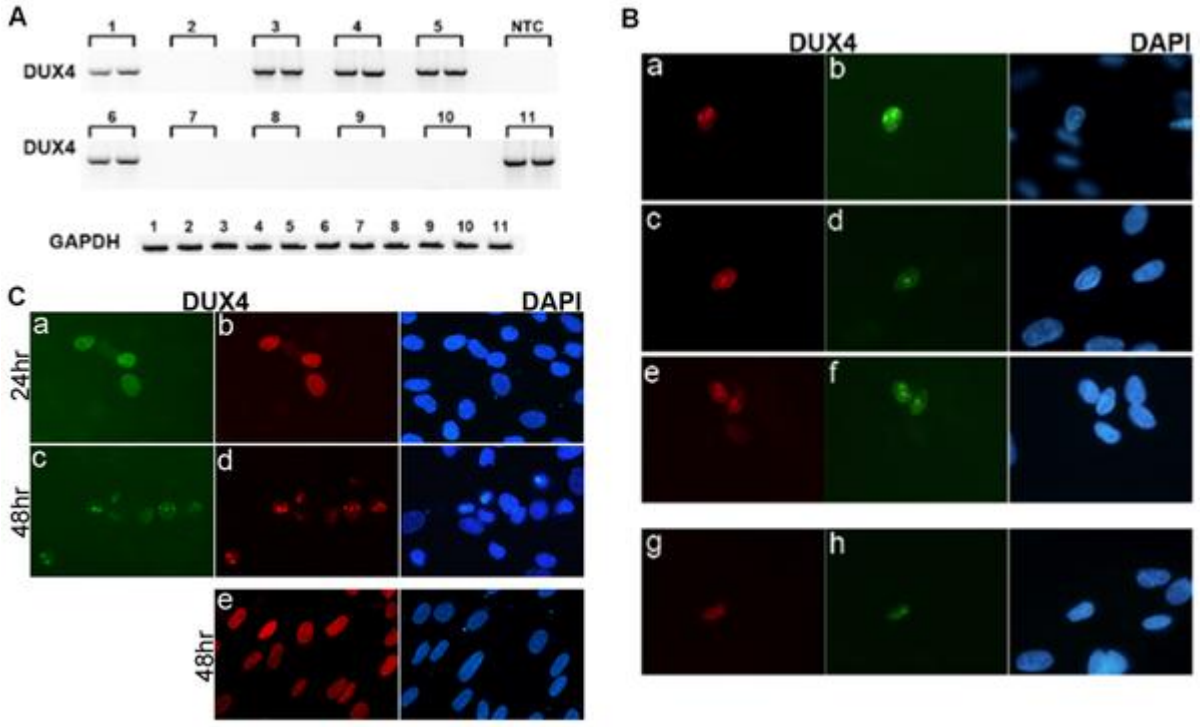


Figure 3.2. A small number of FSHD muscle cells express a relatively large amount of DUX4. (A) RT-PCR for full length DUX4 (DUX4-fl3′) was performed in duplicate on polyadenylated RNA isolated from ten pools of 600 cultured FSHD muscle cells (lanes 1–10) and a single pool of 10,000 cells (lane 11). The presence of DUX4 mRNA in one-half of the pools indicates that approximately one cell per 1000 is expressing DUX4 mRNA. NTC, no template control. (B) Cultured FSHD muscle cells were differentiated and immunostained with monoclonal antibodies to DUX4. Cells were co-stained with the E5-5 rabbit monoclonal antibody to the DUX4 C-terminal region (panels a, c, e) and the P2G4 mouse monoclonal antibody to the N-terminal region (panels b,d,f), or co-stained with the P4H2 mouse monoclonal antibody to the C-terminal region and the E14-3 rabbit monoclonal antibody to the N-terminal region. Approximately 1 cell per 1000 showed nuclear staining and the co-localization of both the n-terminal and c-terminal regions indicates that these cells are expressing the full-length DUX4 protein. No positive nuclei were apparent in the control muscle cultures. (C) Control human muscle cultures infected with a lenti-virus expressing full length DUX4 (a–d) or DUX4-s (e) co-stained with P4H2 (panels a, c) and E14-3 (panels b, d, e). In the cells expressing DUX4-fl, at 24 hrs there is a relatively homogenous nuclear distribution of DUX4 (a,b), whereas at 48 hrs as the cells are undergoing apoptosis DUX4 staining becomes more focal and punctate (c,d), similar to the punctate DUX4 staining in the FSHD muscle cells in panel B. Expression of DUX4-s does not induce apoptosis (data not shown) and the DUX4-s protein maintains a relatively homogenous nuclear distribution at 48 hrs (e).

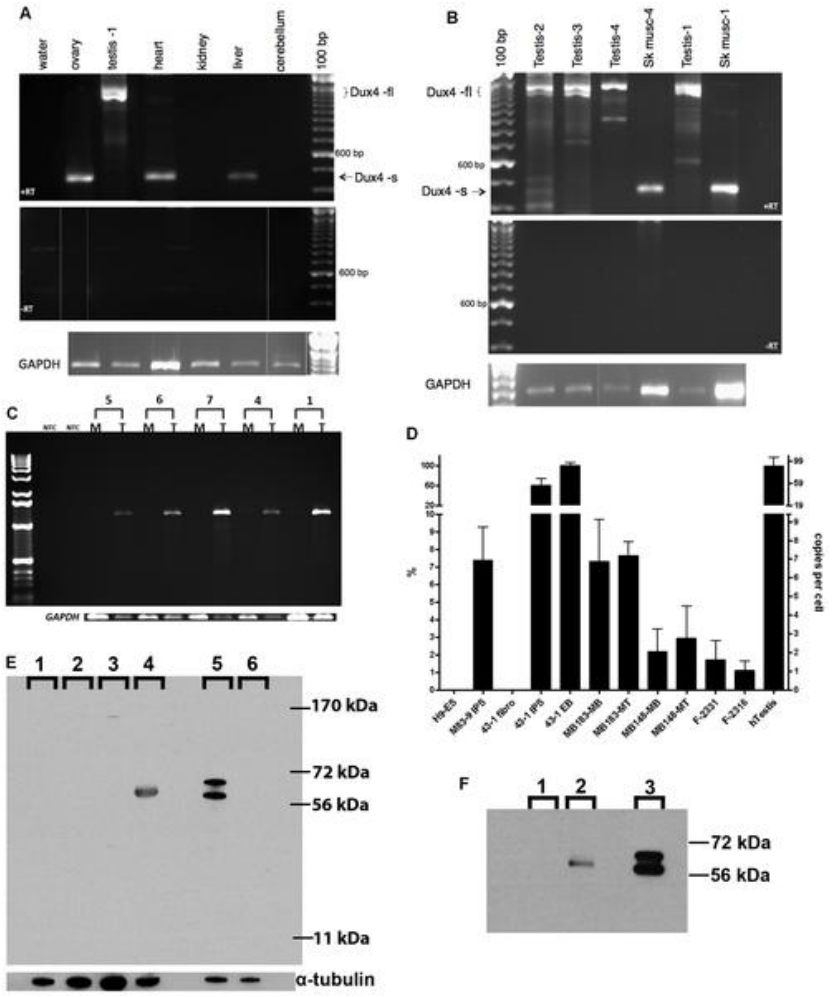


Figure 3.3. Expression of DUX4-fl in human tissues. (A) RT-PCR of RNA from human tissues showing DUX4-fl in the testis sample (Testis-1) and DUX4-s in ovary heart and liver. Note that each sample is from an unknown individual and their genotype is not known. (B) RT-PCR of three additional testis samples and matched skeletal muscle RNA from the Testis-1 and Testis-4 donors. (C) RT-PCR of the full-length DUX4 ORF in muscle (M) and testes (T) RNA from the same individuals, showing expression in testes and not in muscle. (D) RT-qPCR of DUX4-fl3' showing relative abundance in muscle cells, muscle biopsies, human testis and other indicated cells. The sample with the highest expression was set at 100% and the number of copies per cell roughly estimated based on titration of input DNA. H9-ES, human embryonic cell line H9; M83-9 iPS, iPS line made from control fibroblasts; 43-1 series are the fibroblast, iPS, and EB cells from an FSHD fibroblast line; MB183 and MB148, FSHD muscle cultures under growth (MB) and differentiation (MT) conditions; F-2331 and F-2316, FSHD muscle biopsy samples; hTestis, human testis. (E) Western detection of DUX4 protein in whole cell extracts from tissues and cell lines: 1-control muscle culture; 2-HCT116 cell line; 3-mouse testis; 4-human testis; 5-C2C12 cells transfected with human DUX4 expression vector; 6-C2C12 cells. (F) Immunoprecipitation of indicated protein extracts with the E14-3 rabbit monoclonal to the N-terminal region of DUX4 followed by western with the P4H2 mouse monoclonal to the C-terminal region of DUX4 demonstrating that the protein recognized by the rabbit anti-DUX4 is also recognized by an independent mouse monoclonal to DUX4. Lanes: 1-HCT116 cell line lysate; 2- Testis protein lysate; 3-C2C12 cells transfected with DUX4 expression vector.

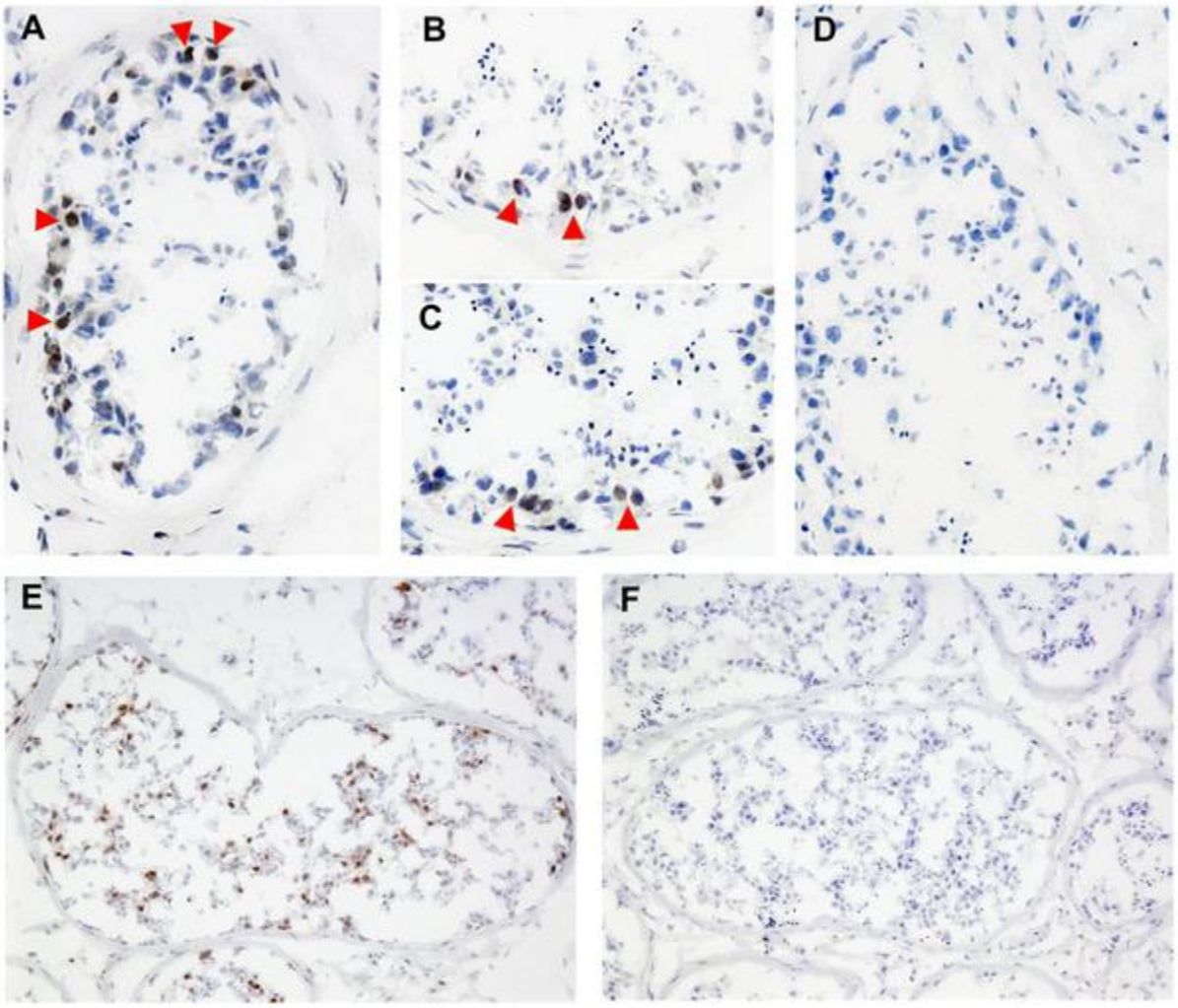
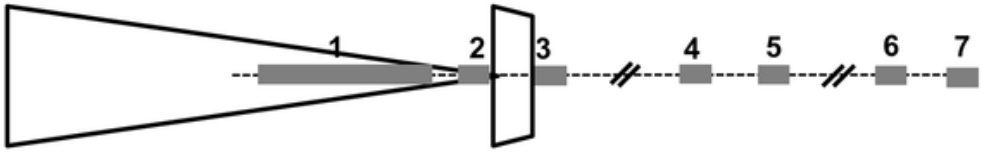


Figure 3.4. Expression of DUX4 in the testis. Frozen sections of human testis stained with anti-DUX4 monoclonal antibodies P2B1 and E5-5 together (A), or P2B1 (B), or E5-5 (C) individually, compared to an isotype control (D). Arrowheads indicate a few of the many nuclei that show a brown antibody-dependent precipitate superimposed on the blue hematoxylin stained nuclei. (E) Antigen retrieval reveals a larger number of more differentiated germ-line cells stained with the P2B1 antibody, compared to an isotype control (F). Because of the range of nuclear morphologies and relatively large numbers of positive cells, the DUX4 expressing cells in the germ line lineage at a range of differentiation stages.

A



B. Sequence of exon 4:
ctgtagGCAGAGGCTAAACAAGAGTTACATCACCTGGATTTTGTTTCCTGCAATAT
GTCACAATGGCGAGGgtgagg

C. Sequence of exon 5
gtatagGCTATTAAGACATGTTTGTCTTCAAAGAATGGCCTTGGTTTCTGTGGACA
GTTTCTCCTCATGGAAAGgtagtg

D. Sequence of exon 6
tcacagCAGCCTTTGTCGCTTCAAACACCGCAAGTGTTCTTTTAAAAGAATTATAT
CAACCTTTCAAGTGAAATGCAACATGTCTGAAACGTGGTATCTGGAGAGgtgaga

E. Sequence of exon 7
ttctagGAACAGTAAGAGGACCTTGTGAGTGAATAATTTGTTTCCACATTACAGAG
TGGGTAATAAGCAGATTAGTAAAAACAATTCTGCTTCACTTCAAtaacag

Figure 3.5. Alternative exon and polyadenylation site usage in germ-line and somatic tissues. (A) Schematic of last D4Z4 unit, last partial repeat, and distal exons. Exon 7 is approximately 6.5 kb from the polyadenylation site in exon 3. DUX4-fl from FSHD muscle contains exons 1-2-3. DUX4-s uses a non-consensus splice donor in the middle of exon 1 to create a short exon 1: 1s-2-3. Both are derived exclusively from chromosome 4A in muscle and other sampled somatic tissues. Germ line tissue expresses the 4A transcript with exons 1-2-3 but also expresses both 4A and 10A transcripts with exons 1-2-6-7. We have also identified a transcript in testis that shows a 1-2-4-5-6-7 splice usage. (B–E) Exon sequences (upper case letters) with flanking genomic sequence (lower case letters). B, Exon 4. C, Exon 5. D. Exon 6. E. Exon 7. Exon 6 and 7 sequences are from a cDNA assigned to chromosome 10, exon 4 and 5 sequences are from a cDNA assigned to 4A161.

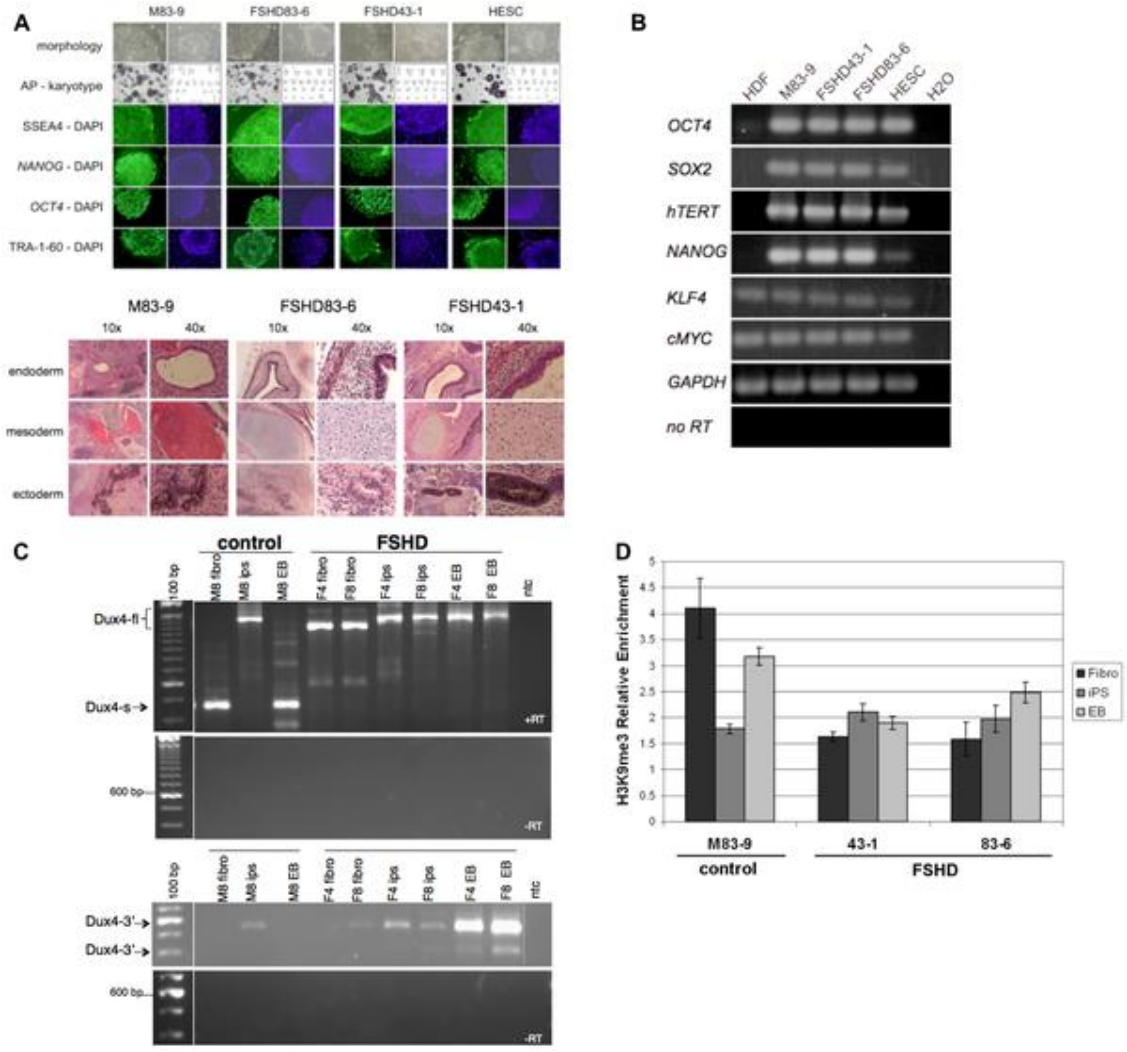


Figure 3.6. Expression of DUX4-fl and DUX4-s in pluripotent stem cells and differentiated tissues. (A) Top: iPS cells from unaffected and FSHD-affected individuals. Human iPS cell lines were generated from skin fibroblasts cultured from an unaffected (M83-9) and two FSHD-affected individuals (FSHD83-6 and FSHD43-1). Human ES cells are also shown. Each iPS cell line was generated by transduction of fibroblasts with murine retrovirus vectors encoding Human SOX2, OCT4, and KLF4. Colonies developed approximately 20 days after infection and had the characteristic growth morphology of an iPS cell.Cells contained tissue non-specific alkaline phosphatase activity (AP), had normal karyotypes, and were immunoreactive (green) for SSEA4, NANOG, OCT4, and TRA-1-60. DAPI staining for nuclei. Bottom: hematoxylin and eosin stained tissue sections of teratomas that developed in SCID-Beige mice after intramuscular injection of iPS cells generated from skin fibroblasts from a control or two different FSHD-affected individuals. Signatures of all germs layers were evident. (B) RT-PCR for OCT4, SOX2, and KLF4 using primers specific for non-vector encoded transcripts. Human ES cells were used as positive control. (-RT), no reverse transcriptase; GAPDH, internal standard. (C) RT-PCR of DUX4 in iPS cells derived from control or FSHD fibroblasts. M8 (M83-9), control fibroblast line; F4 (FSHD 43-1) and F8 (FSHD 83-6), FSHD fibroblast line. (D) ChIP qPCR analysis of H3K9me3 at the 5′-region of DUX4 in control and FSHD fibroblasts, iPS cells, and EB differentiated from the iPS. Bars represent relative enrichment. The H3K9me3 IP signals were normalized to control IgG IP and to input, presented as mean ± stdev.

Chapter 4: DUX4 induces germline genes in skeletal muscle of individuals with FSHD

This chapter is currently in press for publication:

Geng LN, Yao Z, Snider L, Fong AP, Cech JN, Young JM, van der Maarel SM, Ruzzo WL, Gentleman RC, Tawil R, Tapscott SJ. (2011) DUX4 activates germline genes, retroelements and immune-mediators: Implications for facioscapulohumeral dystrophy. *Developmental Cell,* in press.

Summary

Facioscapulohumeral dystrophy (FSHD) is one of the most common inherited muscular dystrophies. The causative gene remains controversial and the mechanism of pathophysiology unknown. Here we identify genes associated with germline and early stem cell development as targets of the DUX4 transcription factor, a leading candidate gene for FSHD. The genes regulated by DUX4 are reliably detected in FSHD muscle but not in controls, providing direct support for the model that misexpression of DUX4 is a causal factor for FSHD. Additionally, we show that DUX4 binds and activates LTR elements from a class of MaLR endogenous primate retrotransposons and suppresses the innate immune response to viral infection, at least in part through the activation of DEFB103, a human defensin that can inhibit muscle differentiation. These findings suggest specific mechanisms of FSHD pathology and identify candidate biomarkers for disease diagnosis and progression.

Introduction

Facioscapulohumeral dystrophy (FSHD) is the third most common muscular dystrophy. The mutation that causes FSHD was identified nearly 20 years ago (Wijmenga et al., 1992), yet the molecular mechanism(s) of the disease remains elusive. The most prevalent form of FSHD (FSHD1) is caused by the deletion of a subset of D4Z4 macrosatellite repeats in the subtelomeric region of chromosome 4q. Unaffected individuals have 11-100 of the 3.3kb D4Z4 repeat units, whereas FSHD1 individuals have 10 or fewer repeats. At least one repeat unit appears necessary for FSHD because no case has been identified with a complete deletion of D4Z4 repeats (Tupler et al., 1996). Each repeat unit contains a copy of the double homeobox retrogene DUX4 (Clapp et al., 2007; Gabriels et al., 1999; Lyle et al., 1995), and inappropriate expression of DUX4 was initially proposed as a possible cause of FSHD. This was supported by the observations that repeat contraction is associated with decreased repressive epigenetic marks in the remaining D4Z4 units (van Overveld et al., 2003; Zeng et al., 2009) and that overexpression of the DUX4 protein in a variety of cells, including skeletal muscle,

causes apoptotic cell death (Kowaljow et al., 2007; Wallace et al., 2011; Wuebbles et al., 2010). However, initial attempts to identify DUX4 mRNA transcripts in FSHD muscle were unsuccessful, leading to the suggestion that other genes in the region were causative for FSHD (Gabellini et al., 2002; Klooster et al., 2009; Laoudj-Chenivesse et al., 2005; Reed et al., 2007).

Recent progress has returned the focus to the DUX4 retrogene as a leading candidate for FSHD. First, a subset of individuals with clinical features of FSHD do not have contracted D4Z4 repeats on chromosome 4 but do have decreased repressive heterochromatin at the D4Z4 repeats (de Greef et al., 2009) (FSHD2), indicating that loss of repressive chromatin at D4Z4 is the primary cause of FSHD. Second, genetic studies identified polymorphisms that create a DUX4 polyadenylation site as necessary for a D4Z4 contraction to cause FSHD (Lemmers et al., 2010a). Third, high sensitivity RT-PCR assays detect DUX4 mRNA specifically in FSHD muscle (Dixit et al., 2007; Snider et al., 2010). Still, a major problem with the hypothesis that DUX4 expression causes FSHD has been the extremely low abundance of the mRNA and inability to reliably detect the protein in FSHD biopsy samples. Our prior work demonstrated that the low abundance of DUX4 in FSHD muscle cells represents a relatively high expression in a small subset of nuclei (Snider et al., 2010). However, it remained unclear whether the expression of DUX4 in FSHD muscle has a biological consequence that might drive the pathophysiology of FSHD.

The coding sequence of the DUX4 retrogene has been conserved in primates (Clapp et al., 2007), but whether this retrogene has a normal physiological function is unknown. Previously we found that DUX4 is normally expressed at high levels in germ cells of human testes and is epigenetically repressed in somatic tissues (Snider et al., 2010), whereas the epigenetic repression of the DUX4 locus in somatic tissues is less efficient in both FSHD1 and FSHD2, resulting in DUX4 expression in FSHD muscle cell nuclei. The germline-specific expression pattern of DUX4 is similar to that of other double homeodomain proteins (Booth and Holland, 2007; Wu et al., 2010). The function of this distinct family of DNA-binding proteins is unknown, but their shared tissue

expression pattern may indicate a possible role for double homeodomain transcription factors in reproductive biology.

Here we report that DUX4 regulates the expression of genes involved in germline and early stem cell development. These DUX4 target genes are aberrantly expressed in FSHD skeletal muscle but not in control muscle biopsies. Therefore, the low level of DUX4 expression in FSHD is sufficient to effect numerous downstream changes and activate genes of germ cell and early development in postmitotic skeletal muscle. Additionally, we show that DUX4 binds and activates LTR elements from a class of MaLR endogenous primate retrotransposons and at the same time suppresses the innate immune response to retroviral infection, at least in part through transcriptional activation of DEFB103, a human defensin that can inhibit muscle differentiation. These findings suggest specific mechanisms of FSHD pathology and identify candidate biomarkers for disease diagnosis and progression.

Results

Identification of genes regulated by DUX4 in human primary myoblasts

Previously, we identified two different DUX4 mRNA transcripts in human skeletal muscle, both at extremely low abundance: a full-length open reading frame mRNA (DUX4-fl) only detected in FSHD muscle and an internally spliced form of DUX4 mRNA (DUX4-s) that maintains the N-terminal double-homeobox domains but deletes the C-terminal domain and is detected in both control and FSHD muscle (Snider et al., 2010). Forced over-expression of DUX4-fl is toxic to cells, inducing apoptotic cell death (Kowaljow et al., 2007; Wallace et al., 2010), whereas forced over-expression of DUX4-s is not toxic to cultured human skeletal muscle cells (Geng et al., 2011). To determine whether gene expression is regulated by DUX4-fl and/or DUX4-s in human muscle cells, we transduced primary myoblasts from a control individual (unaffected by muscle disease) with a lentiviral vector expressing either DUX4-fl or DUX4-s and performed expression microarrays. At 24 hours after transduction, DUX4-fl increased the

expression of 1071 genes and decreased the expression of 837 genes compared to a control myoblast population similarly infected with a GFP expressing lentivirus (2-fold change and FDR<0.01); whereas DUX4-s increased the expression of 159 genes and decreased expression of 45 genes (Fig. 4.1a and see Table S4.1 for the complete list of genes regulated by DUX4-fl and/or DUX4-s). Using a slightly more stringent 3-fold criteria (> 1.584 $\log_2$-fold change and FDR<0.01), 466 genes were increased and 244 decreased by DUX4-fl; and 37 were increased and one decreased by DUX4-s. Only two annotated genes were increased 3-fold or more by both (CCNA1, MAP2), and none were decreased 3-fold or more by both. A representative sample of genes activated by DUX4-fl is shown in Table 4.1 and the full set of genes regulated by DUX4-fl or DUX4-s is in Table S4.1.

The Gene Ontology (GO) terms significantly enriched in 3-fold up-regulated genes by DUX4-fl included categories such as RNA polymerase II mediator complexes, RNA splicing and processing, and gamete/spermatogenesis (Table S4.2); whereas down-regulated genes were enriched in immune response pathways (Table S4.3). The up-regulation of a large number of transcription-related and RNA processing factors suggests that DUX4-fl might be a central component of a complex gene regulatory network, and the large number of germline associated genes suggests a possible role in reproductive biology.

In primary human myoblasts, DUX4-fl robustly induced a large number of genes not normally detected in skeletal muscle (see the contrail of genes along the Y-axis in the left panel of Fig. 4.1a). These genes are good candidate biomarkers of DUX4 activity in skeletal muscle, since there is little to no background expression in control muscle. GO analysis for these highly induced genes showed enrichment for gamete generation and spermatogenesis categories (Table S4.4). In many cases, DUX4-fl activated multiple members of gene families involved in germ cell biology and early development, including some primate-specific genes (Table 4.2). We validated the differential expression of 15 of the DUX4-fl regulated genes by RT-PCR (Fig. S4.1).

Identification of DUX4-fl binding sites and a consensus binding sequence motif

Double homeodomain proteins comprise a distinct group of DNA-binding proteins (Holland et al., 2007), but their consensus recognition sites and genomic targets are unknown. Therefore, we performed chromatin immunoprecipitation combined with high throughput sequencing (ChIP-Seq) to identify DUX4-binding sites in human muscle cells. We used two polyclonal rabbit antisera against DUX4 (Fig. S4.2) to immunoprecipitate DUX4-fl from human primary myoblasts 24 hours after transduction with lentiviral expressed DUX4-fl or control non-transduced primary myoblasts. Non-redundant reads unambiguously mapped to the human genome were computationally extended to a total length of 200 nucleotides and "peaks" were defined as regions where the number of reads was higher than a statistical threshold compared to the background (see Methods). Reads mapping to the X and Y chromosomes were excluded from our analysis.

A total of 62,028 and 39,737 peaks were identified at P-value thresholds of 10^{-10} and 10^{-15}, respectively, after subtracting background peaks in the control samples. DUX4-fl peaks were widely distributed both upstream and downstream of gene transcription start sites (TSSs) with higher numbers in introns and intergenic regions, but showing a relatively constant peak density in all genomic regions when normalized for the size of the genomic compartment (Fig. 4.1b, left). This pattern differs from that reported for many other transcription factors, such as MYOD (Cao et al., 2010), shown for comparison (Fig. 4.1b, right), that show higher average peak density in regions near TSSs.

A *de novo* motif analysis identified the sequence TAAYBBAATCA (IUPAC nomenclature) (Fig. 4.1c, top) near the center of greater than 90% of peaks. To our knowledge, this motif has not been described for any other transcription factor, but does contain two canonical homeodomain binding motifs (TAAT) arranged in tandem and separated by one nucleotide. Approximately 30% of sequences under the DUX4-fl peaks also contained a second larger motif that encompasses the primary DUX4-fl binding motif. This longer motif matches the long terminal repeat (LTR) of retrotransposons (Fig. 4.1c, bottom). Assessment of the representation of DUX4-fl

binding at different annotated repetitive elements in the genome shows a nearly 10-fold enrichment of DUX4-fl binding in the Mammalian apparent LTR-Retrotransposon (MaLR) family of retrotransposons and some enrichment in the related ERV family (Table S4.5). Note that the quantitative estimate of repeat-associated binding sites is conservative since reads mapping to more than one locus are excluded from our analysis.

MaLR family members expanded in the primate lineages (Smit, 1993). Thus, if DUX4-fl binding sites were carried throughout the genome during this expansion, these newer sites might have a different sequence motif compared to DUX4-fl binding sites located outside of MaLR repeats. To determine if the expansion of MaLR-associated binding sites might obscure the identification of a different DUX4 binding motif in non-repetitive elements, we performed separate motif analysis of MaLR-associated sites and sites not associated with repeats; both yielded nearly identical core motifs, TAAYYBAATCA and TAAYBYAATCA, respectively, but the repeat-associated motifs had slightly more flanking nucleotides preferences reflecting the LTR sequence (Fig. S4.3a). Electrophoretic mobility shift assay (EMSA) confirmed that DUX4-fl binds the core motif present in both MaLR-associated and non-repeat associated sites and that mutation of the core nucleotides abolishes binding (Fig. S4.3b-d), including sites from both repeat and non-repeat regions. Because the DUX4-s alternative splice form retains the N-terminal DNA-binding homeodomains, we hypothesized that it would bind to the same sites as DUX4-fl. EMSA confirmed that DUX4-s specifically binds the same core binding site as DUX4-fl *in vitro* (Fig. S4.3e). Thus, DUX4-s can bind the same sequences as DUX4-fl but does not activate transcription of the same genes, which supports the prior determination that the C-terminus contains a transactivation domain (Kawamura-Saito et al., 2006).

DUX4-fl is a transcriptional activator

The number of DUX4-fl binding locations exceeds the number of genes that robustly increase expression in muscle cells following transduction with DUX4-fl. A genome-wide analysis of peak height and regional gene expression shows only a weak

association of binding and gene expression for DUX4-fl (Fig. S4.4a). To determine whether DUX4-fl binding might function as a transcriptional activator at some of the identified binding sites, DUX4 binding sites from selected genes were cloned upstream of the SV40 promoter in the pGL3-promoter luciferase construct. Co-transfection with DUX4-fl in human rhabdomyosarcoma cell line RD significantly induced luciferase expression independent of orientation or position, and mutation of the DUX4 binding motif eliminated the induction (Fig. 4.2a and Fig. S4.4b). In contrast to DUX4-fl, DUX4-s did not activate expression despite demonstrating in vitro binding to this site (Fig. 4.2a and Fig. S4.3e).

To determine whether DUX4 binding might directly regulate transcription of select genes, we cloned the 1.9 kb enhancer and promoter region of the ZSCAN4 gene that includes four DUX4 binding sites. Co-transfection with DUX4-fl significantly induced expression of this reporter and mutation of three of the four DUX4 binding sites nearly abolished the induction (Fig 4.2b). DUX4-s interfered with the activity of DUX4-fl when the two were co-expressed (Fig. 4.2b, right), suggesting that DUX4-s acts as a dominant negative for DUX4-fl activity.

DUX4-fl also activated transcription through DUX4 sites in repetitive elements: DUX4-fl activated transcription of a luciferase reporter containing DUX4 binding sites cloned from LTRs at a MaLR THE1D element (Fig. 4.2c) and RT-PCR showed induction of endogenous MaLR transcripts in muscle cells transduced with DUX4-fl (Fig. 4.2d).

DUX4-fl directly regulates genes involved in germline development

To identify the set of genes that might reflect the function of DUX4-fl prior to the expansion of MaLRs in primates, we identified the subset of genes activated at least 3-fold by DUX4-fl that also contain a non-repeat associated binding site within six kilobases of the TSS and not separated from the TSS by a binding site for the insulator factor CTCF (Table S4.6). The 74 genes meeting these criteria are highly enriched for genes involved in stem and germ cell functions, RNA processing, and regulated components of the PolII complex, similar to the major GO categories identified for all of the genes regulated by DUX4-fl. Quantitative RT-PCR of six DUX4-regulated genes on

paired samples of testis mRNA and skeletal muscle mRNA from two control individuals found high expression of these targets in the testes and absent, or nearly absent, expression in skeletal muscle (Fig, 4.3). We also detected the expression of the related DUXA and DUX1 genes in healthy human testis (data not shown), further supporting the notion that this family of double homeodomain proteins has a role in germ cell biology.

DUX4-fl-regulated genes are expressed in FSHD muscle

To determine whether the low levels of endogenous DUX4-fl mRNA detected in FSHD skeletal muscle is sufficient to activate DUX4 target genes, we assessed the expression of these genes in a set of control and FSHD muscle. Cultured muscle cells from control biopsies showed low or absent expression of the six DUX4-fl regulated genes, whereas these genes were expressed at significantly higher levels in the FSHD muscle cultures (Fig. 4.4a), including those from both FSHD1 and FSHD2 individuals.

Similar to the expression of DUX4-fl regulated targets in cultured FSHD muscle, muscle biopsies from FSHD individuals had readily detectable mRNA of DUX4-fl regulated genes, although at varying levels in different biopsies (Fig. 4.4b). The DUX4-fl mRNA is at extremely low abundance in FSHD muscle and it is notable that some biopsy samples in which the DUX4-fl mRNA was not detected showed elevation of DUX4 regulated targets (Table S4.7 and Fig. S4.5a), indicating that the target mRNA is of significantly higher abundance and perhaps more stable than the DUX4 mRNA.

To determine whether the expression of these genes in FSHD muscle cells was directly due to DUX4, we transfected FSHD muscle cells with siRNA to the endogenous DUX4-fl mRNA. The siRNA sequences that decreased the DUX4-fl mRNA also resulted in decreased expression of the DUX4 target genes (Fig. 4.4c), confirming that endogenous DUX4 drives the expression of these genes in FSHD muscle cells. In addition, expression of the dominant negative DUX4-s (see Fig. 4.2b) also inhibited the endogenous expression of the target genes (Fig. S4.5b).

DUX4-induced DEFB103 inhibits the innate immune response and muscle differentiation

As noted above, genes enriched in the innate immunity pathways were expressed at lower levels in myoblasts transduced with lenti-DUX4-fl compared to the lenti-GFP or lenti-DUX4-s (see Table S4.3). When compared to non-transduced cells, it was evident that about 350 genes, most of which were in the innate immunity pathway, were unchanged in the lenti-DUX4-fl transduced myoblasts but increased in cells transduced with either control lenti-GFP or lenti-DUX4-s (Table S4.8). Therefore, lentiviral induction of the innate immune response in human muscle cells appeared to be inhibited by DUX4-fl. RT-qPCR validated that lenti-GFP, lenti-DUX4-s, and multiple other lentivirus constructs induced the innate immune response in myoblasts, whereas similar titers of lenti-DUX4-fl did not (Fig. 4.5a, left, and data not shown). Additionally, supernatant from DUX4-fl infected cells reduced the induction of these genes by lenti-GFP (Fig. 4.5b), indicating that a secreted factor induced by DUX4-fl could mediate this suppressive effect.

DUX4-fl robustly induced expression of DEFB103A and DEFB103B (β-defensin 3) (Fig. 4.5a, right, Table 4.1 and S4.1), which has been shown to inhibit the transcription of pro-inflammatory genes in TLR4-stimulated macrophages (Semple et al., 2011). Indeed, addition of DEFB103 peptide also inhibited the induction of the innate immune response to lenti-GFP when added to the muscle cells at the time of infection (Fig. 4.5b) but did not prevent viral entry and transduction as measured by copies of viral integrants in the genome and levels of GFP mRNA expressed (data not shown). Thus, DUX4 can prevent the innate immune response to viral infection in skeletal muscle cells, at least in part, through the transcriptional induction of DEFB103.

Like other DUX4-regulated genes, endogenous expression of DEFB103 was detected in FSHD cultured muscle cells, FSHD muscle biopsies, and in healthy testes, but little to none was seen in control skeletal muscle (Fig. 4.5c). DEFB103 has been previously shown to bind the CCR6, CCR2, and melanocortin receptors and to be an antagonist ligand for the CXCR4 receptor, which is important for muscle cell migration and differentiation (Candille et al., 2007; Feng et al., 2006; Jin et al., 2010; Yang et al., 1999). To determine whether DEFB103 could affect myoblasts or muscle differentiation, we treated cultured control human muscle cells with DEFB103 peptide at concentrations considered to be physiological (0.5-1.0 ug/ml) (Midorikawa et al.,

2003; Semple et al., 2011) and assessed changes with gene expression arrays. Based on a 2-fold change threshold, DEFB103 did not alter the expression of any genes in myoblasts, although it is of interest that myostatin was upregulated approximately 50% and RT-qPCR confirmed that DEFB103 increased the mRNA for myostatin in myoblasts (Fig. 4.5d). In contrast, exposing differentiating muscle cells to DEFB103 reduced the expression of 44 genes relative to the untreated control, the majority of which were genes associated with muscle differentiation (Table S4.9), and RT-qPCR on select genes (ACTA1, CKM, CASQ2, MYH2 and TNNT3) validated the array results (Fig. 4.5e). Therefore, DEFB103 activates the expression of myostatin in myoblasts and inhibits the expression of genes necessary for normal muscle differentiation, although it remains to be determined whether this activity is mediated by the CXCR4 receptor. Therefore, DUX4-mediated expression of DEFB103 in FSHD muscle can modulate the innate immune response to retroviral infection and can inhibit myogenic differentiation.

Discussion

Recent genetic and molecular studies indicated DUX4 as the likely candidate gene for FSHD (Dixit et al., 2007; Lemmers et al., 2010a; Snider et al., 2010). Although the abundance of DUX4-fl mRNA was extremely low in FSHD muscle, we previously showed that this represented relatively high expression of both DUX4-fl mRNA and protein in a small percentage of muscle nuclei at any time point, either because the gene was on transiently or the expressing nuclei were eliminated (Snider et al., 2010). Yet, it remained unclear whether DUX4-fl expression had a biological consequence in FSHD. In our current study, we identify genes regulated by DUX4-fl and show that they are expressed at readily detectable levels in FSHD skeletal muscle, both cell lines and muscle biopsies, but not in control tissues, providing direct support for the model that misexpression of DUX4-fl is a causal factor for FSHD. The genes regulated by DUX4-fl suggest several specific mechanisms for FSHD pathophysiology.

Many of the genes highly upregulated by DUX4-fl normally function in the germline and/or early stem cells and are not present in healthy adult skeletal muscle.

This supports a biological role for DUX4-fl in germ cell development and suggests potential disease mechanisms for FSHD. Activation of the gametogenic program might be incompatible with post-mitotic skeletal muscle, leading to apoptosis or cellular dysfunction. Also, the testis is an immune-privileged site and testis proteins misexpressed in cancers can induce an adaptive immune response (Simpson et al., 2005). In fact, some of the genes regulated by DUX4-fl, such as the PRAME family (Chang et al., 2011), are known cancer testis antigens, so it is reasonable to suggest that expression of these genes in skeletal muscle might also induce an adaptive immune response. An immune-mediated mechanism for FSHD is consistent with the focal inflammation and CD8+ T-cell infiltrates that characterize FSHD muscle biopsies (Frisullo et al., 2011; Molnar et al., 1991).

The induction of DEFB103 by DUX4 might influence both the adaptive and the innate immune response. DEFB103 can have a pro-inflammatory role in the adaptive immune response and can act as a chemo-attractant for monocytes, lymphocytes and dendritic cells (Lai and Gallo, 2009). In this regard, it might enhance an adaptive immune response to germline antigens expressed in FSHD muscle. Though traditionally known for its role in antimicrobial defense (Sass et al., 2010), DEFB103 has been shown to suppress the innate immune response to LPS and TLR4 stimulation in macrophages (Semple et al., 2011; Semple et al., 2010), and has also been shown to be an antagonistic ligand of the CXCR4 receptor (Feng et al., 2006), which is important for muscle migration, regeneration, and differentiation (Griffin et al., 2010; Melchionna et al., 2010). In this study we show that DEFB103 inhibited the innate immune response to lentiviral infection in skeletal muscle cells, modestly induced myostatin in myoblasts, and impaired muscle cell differentiation. Therefore, DEFB103 might contribute to FSHD pathology by modulating the adaptive and innate immune response, as well as through inhibiting muscle differentiation, although it remains to be determined whether these effects are mediated by the CXCR4 receptor.

Reactivation of retroelements can result in genomic instability (Belancio et al., 2010) and transcriptional deregulation (Schulz, 2006). Therefore, DUX4 activation of MaLR transcripts might directly contribute to FSHD pathophysiology. It is interesting

that DUX4 both activates retroelement transcription and suppresses the virally induced innate immune response. Although we have shown that DEFB103 can substitute for DUX4 to suppress the innate immune response, products of retroelements and endogenous retroviruses may do the same and, thus, the DUX4-mediated suppression might be multi-factorial. Since DEFB103 is also expressed in the testis, it is interesting to consider whether the role of DUX4 in the germline might include a simultaneous activation of retroelement transcription and suppression of the innate immune response to those transcripts.

DUX4 regulated targets also include genes involved in RNA splicing, developmentally regulated components of the Pol II transcription complex, and ubiquitin-mediated protein degradation pathways, all of which may have pathophysiological consequences. For example, DUX4 is known to induce apoptosis and inhibit myogenesis in muscle cells, but the specific pathways that mediate these responses remain unknown. As in many human diseases, a single mutation can effect multiple pathological pathways that collectively account for the complex disease phenotype. Our study of DUX4 regulated genes has identified several candidate pathways and future work will be necessary to determine their relative contributions to the disease phenotype.

In this regard, other genes have been identified as candidates for FSHD. For example, FRG1 expression has been reported to be elevated in FSHD muscle (Gabellini et al., 2002) and FRG1 transgenic mice display a muscular dystrophy phenotype (Gabellini et al., 2006). It is interesting the FRG1 is reported to alter RNA splicing in FSHD muscle (Gabellini et al., 2006) and that our study shows that DUX4-fl also alters the expression of many genes that regulate splicing and RNA processing. It will be important to determine the relative contributions of DUX4 and FRG1 to FSHD pathophysiology; however, the human genetics shows a convincing linkage to polymorphisms necessary for the polyadenylation of the DUX4 mRNA (Lemmers et al., 2010a), indicating that DUX4 mRNA is a necessary component of the disease. Therefore, one therapeutic avenue to pursue for FSHD is to reduce the activity of DUX4, either by

eliminating its expression in the muscle cells as we have done in vitro with an siRNA or by introducing a dominant negative, such as the DUX4-s splice form.

A previous study identified PITX1 as a DUX4 target gene expressed in FSHD skeletal muscle and in mouse cells transfected with DUX4 (Dixit et al., 2007). Others have expressed DUX4 in mouse muscle cells and identified repression of the glutathione redox pathway (Bosnakovski et al., 2008b). Both of these findings are consistent with our expression array data. However, since many of the DUX4 binding sites reside in primate-specific MaLRs and some of the DUX4 targets, including RFPL2 and DEFB103, are not conserved in mice, further studies are necessary to determine the conserved and primate-specific functions of DUX4, an important consideration for evaluating mouse models of FSHD.

In conclusion, our data support the model that inappropriate expression of DUX4 plays a causal role in FSHD skeletal muscle pathophysiology by activating germline gene expression, endogenous retrotransposons, and suppressors of differentiation in skeletal muscle. The set of genes robustly upregulated by DUX4 in FSHD skeletal muscle are candidate biomarkers because they are absent in control muscle and easily detected in FSHD1 and FSHD2 muscle. Furthermore, some target genes encode secreted proteins, which offer the potential for developing blood tests to diagnose FSHD or monitor response to interventions. Beyond their utilities as candidate biomarkers, the DUX4 targets identified in this study point to specific mechanisms of disease and may help guide the development of therapies for FSHD.

Materials and Methods

Cell culture

Primary human myoblasts were collected and cultured as previously described (Snider et al., 2010). Human RD cells were grown in DMEM in 10% bovine calf serum (Hyclone) and pencillin/streptomycin.

Microarray and GO Analysis

Quadruplicate total RNA samples were collected from control human primary myoblasts transduced with lentivirus carrying DUX4-fl, DUX4-s or GFP (MOI = 15) for 24 h. Samples were analyzed by Illumina HumanHT-12 v4 Expression BeadChip Whole Genome arrays. Probe intensities were corrected, normalized, and summarized by the Lumi package of Bioconductor(Du et al., 2008). Differentially expressed genes were identified by the LIMMA package of Bioconductor (Wettenhall and Smyth, 2004). Gene set enrichment analysis (GSEA) was performed using the Bioconductor GOstats package(Falcon and Gentleman, 2007).

Chromatin Immunoprecipitation and Ultra-High-Throughput Sequencing

ChIP was performed and ChIP DNA samples were prepared as previously described(Cao et al., 2010). Anti-DUX4 C-terminus rabbit polyclonal antibodies MO488 and MO489 were combined to immunoprecipitate DUX4-fl. The samples were sequenced with Illumina Genome Analyzer II.

Defining peaks

Sequences were extracted by Illumina package GApipeline and reads were aligned using BWA to the human genome (hg18). We only kept one of the duplicated sequences to minimize the artifacts of PCR amplification. Reads mapping to multiple locations in the genome were excluded from our analysis. Each read was extended in the sequencing orientation to a total of 200 bases to infer the coverage at each genomic position. We performed Peak calling by an in-house developed R package "peakSig"(pending submission to Bioconductor), which models background reads by a negative binomial distribution. The negative binomial distribution can be viewed as a continuous mixture of Poisson distribution where the mixing distribution of the Poisson rate is modeled as a Gamma prior. This prior distribution is used to capture the variation of background reads density across the genome. Model parameters were estimated by fitting the truncated distribution on the number of bases with low coverage (one to three), to avoid the problem of inferring effective genome size excluding the non-mappable regions, and to eliminate contamination of any foreground signals in the high coverage regions. We also fit a GC dependent mixture model so that

the significance of the peaks is determined not only by peak height, but also by the GC content of the neighboring genomic regions.

Motif analysis

We used an in-house developed R package "motifRG" (pending submission to Bioconductor) described previously for discriminative de-novo motif discovery(Palii et al., 2011). Briefly, it finds motifs that distinguish positive and negative sequence datasets, which in this study correspond to DUX4 binding sites and randomly sampled flanking regions of DUX4 binding sites. To generate a more accurate presentation of the DUX4 binding sites from the consensus pattern returned by the motifRG package, we further used a positional weight matrix (PWM) model, using the matches of the consensus pattern as the seed to initialize the iterative expectation-maximization (EM) refinement process similar to MEME. The motifs were extended iteratively as long as there was sequence preference in the flanking region, and refined in the same EM process.

Luciferase Assay

Transient DNA transfections of RD cells were performed using SuperFect (Qiagen) according to manufacturer specifications. Briefly, 3 x 10^5 cells were seeded per 35 mm plate the day prior to transfection. Cells were co-transfected with pCS2 expression vectors (2 ug/plate) carrying either β-galactosidase, DUX4-fl or DUX4-s and with pGL3-promoter luciferase reporter vectors (1 ug/plate) carrying various putative DUX4 binding sites or mutant sites upstream of the SV40 promoter or pGL3-basic reporter vector (1 μg/plate) carrying test promoter fragment upstream of the firefly luciferase gene. Cells were lysed 24 h post-transfection in Passive Lysis Buffer (Promega). Luciferase activities were quantified using reagents from the Dual-Luciferase Reporter Assay System (Promega) following manufacturer's instructions. Light emission was measured using BioTek Synergy2 luminometer. Luciferase data are given as the averages ± SD of at least triplicates.

Real-time PCR

One microgram of total RNA was reverse transcribed into first strand cDNA in a 20 uL reaction using SuperScript III (Invitrogen) and digested with 1U of RNase H

(Invitrogen) for 20 min at 37°C. cDNA was diluted and used for quantitative PCR with iTaq SYBR Green supermix with ROX (Bio-Rad). The relative expression levels of target genes were normalized to those of ribosomal protein L13A (RPL13A) by $2^{\Delta Ct}$. Undetermined values were equated to zero. Standard deviations from the mean of the ΔCt values were calculated from triplicates. PCR primers used for detecting the transcripts of the selected genes are listed in Supplementary methods. All primers amplify with similar and high (>90%) efficiencies.

Muscle biopsies and human RNA

Muscle biopsy samples were collected from the vastus lateralis muscle of clinically affected and control individuals as previously described (Snider et al., 2010). RNA from matched tissues from healthy donors were purchased from BioChain (Hayward, CA).

Statistical analyses

Statistical significance between two means was determined by unpaired one-tailed *t* tests with *P*-value <0.05. Statistics for the microarray and ChIP-Seq experiments are described separately.

RT-PCR

Total RNA was treated with DNase using TURBO DNA-free kit (Ambion) according to manufacturer's protocol. One μg of DNase-treated RNA was reverse transcribed to first strand cDNA with SuperScript III and anchored oligo dT (Invitrogen) at 52°C for 1 h. Residual RNA was digested with RNase H at 37°C for 20 min. cDNA was used in various PCR and real-time PCR reactions with primers listed below.

Microarray Gene Targets Validation

cDNA from DUX4-fl transduced or untransduced primary myoblasts was diluted 1:5 and used in PCR reactions with Platinium *Taq* polymerase (Invitrogen) with conditions of 55°C annealing temperature and 35 cycles. Primers were designed to span exon-exon junctions where possible. Primers from select genes were also used in real-time PCR reactions to examine endogenous expression of targets in FSHD versus control samples described separately.

Gene name	Forward primer sequence	Reverse primer sequence
TRIM43	ACCCATCACTGGACTGGTGT	CACATCCTCAAAGAGCCTGA
PRAMEF1	GCTGGAACACCTTCAGTTGC	AGTTCTCCAAGGGGTTCTGG
RFPL4B	GAGACGTAGGCTTCGGATCTT	GGCTGAATTCAAGTGGGTCT
ZSCAN4	TGGAAATCAAGTGGCAAAAA	CTGCATGTGGACGTGGAC
KHDC1	ACCAATGGTGTTTCACATGG	TGAATAAGGGTGTGGCTGTG
RFPL2	CCCACATCAAGGAACTGGAG	TGTTGGCATCCAAGGTCATA
CXCR4	CGTGGAACGTTTTTCCTGTT	GGTGCTGAAATCAACCCACT
WDR33	GGTCCCACCTATAGGAATGTTG	GACCAAGCGTCTTCCTTCTG
MBD3L2	GCGTTCACCTCTTTTCCAAG	GCCATGTGGATTTCTCGTTT
CCNA1	TGAAGCAGATCCATTCTTGAAA	ACCCTGTAAATGCAGCAAGG
TRIM48	TGAATGTGGAAACCACCAGA	GTTGAGCCTGTCCCTCAGTC
PRAMEF2	ACCTTCTTCAGTGGGCACCT	TGGGAACTGGGAGAGACACT
IFI27	CCATAGCAGCCAAGATGATG	GAACTTGGTCAATCCGGAGA
TESK2	GCAGGAGAGGGATAGGAAGC	CTTGTGGGGGATCTTGTCAT
PELI1	CTAAGGCAAATGGGGTGAAG	TCTGGGCCCGAGATAAAGTA
FRG2B	GTCCAGCTCATATCGGGAAA	GCTGCACTCCTTTTCTGGAC
HSPA2	CTTCTGCCGTGATTGTGAGG	CCAGGGGGTCTAGGTAGGAG
RPL13A	AACCTCCTCCTTTTCCAAGC	GCAGTACCTGTTTAGCCACGA

MaLR Expression Analysis

Real-time PCR was performed as described in main methods. Water and minus RT controls were checked to ensure there was no amplification of these repetitive elements from residual or contaminating genomic DNA. Primer sequences were: THE1 forward, 5’ – ACCCCTCATGGAGAACCTCT – 3’ and THE1 reverse, 5’ – ACCCTCTTCTCACAGCTCCA – 3’.

Antibody development and characterization

Custom anti-DUX4 polyclonal antibodies MO488 and MO489 were developed through Covance. Rabbits were immunogenized with GST-DUX4 C-terminus fusion protein as antigen(Geng et al., 2011). Human myoblasts were transduced with lentivirus-DUX4-fl

and used for testing the antibodies on immunofluorescence as previously described(Geng et al., 2011). Briefly, cells were fixed in 2% paraformaldehyde and incubated overnight with antibodies at a 1:1000 dilution. Cells were counterstained with 4',6-diamidino-2-phenylindole (DAPI) for nuclei. Immunoprecipitation of lysates were performed with rabbit polyclonals bound to a 1:1 mixture of Protein A and Protein G Dynabeads (Invitrogen, CA) following manufacturer's instructions. DUX4 protein was immunoprecipitated overnight at 4°C. Precipitated material was eluted directly in Laemmli buffer and boiled for western blot. Samples were run on 4-12% gradient bis-tris polyacrylamide gel, transferred to 0.45 μm nitrocellulose membranes and probed with a custom mouse monoclonal antibody against DUX4 called P4H2. Anti-mouse kappa light chain (SouthernBiotech, AL) was used a secondary antibody to minimize cross reactivity against denatured rabbit IgG heavy chain.

Electromobility Shift Assay

EMSA was performed with ^{32}P-labeled 31-bp oligonucleotides from endogenous genomic sequences containing the putative DUX4 binding site as probes (sequences below; only forward shown). Radiolabeled probes were incubated with in vitro translated protein generated from pCS2-DUX4-fl or pCS2-DUX4-s(Geng et al., 2011) vectors using the TNT SP6 Coupled Wheat Germ Extract System (Promega) according to manufacturer's instructions. To obtain supershift of protein-DNA complexes, 0.1 μg of E14-3 anti-DUX4 rabbit monoclonal antibody was added to the mixtures. For competition experiments, excess unlabeled probes of either wild-type or mutant sequences were included in the binding reaction. The gels were prepared and run as previously described(Knoepfler et al., 1999).

Probe	*Forward oligo sequence*
TRIM48	AGGAGTGATGATAATTTAATCAGCCGTGCAA
TRIM48mut	AGGAGTGATGATACTTTTATGAGCCGTGCAA
THED1	CCTGTGGGAGGTAATCCAATCATGGAGGCAG
THE1Dmut	CCTGTGGGAGGTACTCCTATGATGGAGGCAG

CSF1R	CCAGGTGGAGATAATTGAATCATGGGGGCAG
CSF1Rmut	CCAGGTGGAGATACTTGTATGATGGGGGCAG

Association of binding and expression

We associated a peak to its closest TSS within the region flanked by CTCF binding sites, which were identified in a ChIP-seq experiment on human CD4+ T cells (GEO accession number GSE12889/GSM325895).

Enhancer activity reporter test

The DUX4 binding site in the ZSCAN4 pGL3-promoter construct was either reversed in orientation or moved downstream of the reporter gene. Cells were co-transfected with pCS2 expression vectors (1 ug/plate) carrying either β-galactosidase or DUX4-fl and with pGL3-promoter luciferase reporter vectors (1 ug/plate).Transfections and luciferase assays were done as in main methods. Data are given as the averages ± SD of triplicates.

Real-time PCR of targets in matched testis and skeletal muscle

Real-time PCR was performed as described in main methods. Primer sequences for muscle markers are listed below.

Gene name	Forward primer sequence	Reverse primer sequence
MYH2	TTCTCAGGCTTCAAGATTTGG	CTGGAGCTTGCGGAATTTAG
CKM	CACCCCAAGTTCGAGGAGAT	AGCGTTGGACACGTCAAATA

DUX4-fl PCR

Nested DUX4-fl3’ PCR on primary myoblast and muscle biopsies were performed as previously described(Snider et al., 2010). Primers for were 182 forward (5’ – CACTCCCCTGCGGCCTGCTGCTGGATGA – 3’) and 183 reverse (5’ – CCAGGAGATGTAACTCTAATCCAGGTTTGC – 3’) nested with 1A forward (5’ – GAG CTC CTG GCG AGC CCG GAG TTT CTG – 3’) and 184 reverse (5’ – GTAACTCTAATCCAGGTTTGCCTAGACAGC – 3’).

siRNA Knockdown of DUX4

siRNAs (Dharmacon) targeted to the mature mRNA of DUX4 and a control siRNA against luciferase were designed and provided to us by Isis Pharmaceuticals, Carlsbad, CA. Sequences of siRNAs were as follows:

Control siRNA	5'-r(CUUACGCUGAGUACUUCGA)d(TT)-3'
DUX4 siRNA 1	5'-r(GAGCCUGCUUUGAGCGGAA)d(TT)-3'
DUX4 siRNA 2	5'-r(GCGCAACCUCUCCUAGAAA)d(TT)-3'
DUX4 siRNA 3	5'-r(CAAACCUGGAUUAGAGUUA)d(TT)-3'
DUX4 siRNA 4	5'-r(GAUGAUUAGUUCAGAGAUA)d(TT)-3'

Cultured FSHD myoblasts were transfected in 35mm dishes with 30pmol of siRNA using RNAiMAX according to manufacturer's recommendations. Following overnight incubation with siRNA complexes, cells were washed and allowed to recover for 12-24 hours in fresh growth media (F10, 20% FBS, 1 μM dexamethasone, 0.01 μg/ml FGF). When confluent, cultures were changed to differentiation media (F10, 1% horse serum, 10μ/ml each insulin and transferrin,) for 48 hours. RNA was isolated using RNeasy Miniprep Kit (Qiagen), RT and PCR protocols were performed as described in Snider et al. (2010) using primers 1A and 183 (see above for sequences).

Dominant Negative Inhibition of DUX4-fl

Cultured FSHD myoblasts were grown to confluence and switched to differentiation media as described in main methods. Simultaneously, cells were transduced by lentivirus carrying DUX4-s or GFP along with 8 μg/mL polybrene. Cells were washed and changed to plain differentiation media after 24 hours. Cells were harvested for RNA after 48 hours of differentiation. Untransduced cells were used to assess baseline expression of DUX4-fl target genes.

Beta-defensin 3 and innate immune response

Cultured control myoblasts were grown to 80% confluence and infected with equivalent titers of lenti-GFP, lenti-DUX4-s and lenti-DUX4-fl in growth media supplemented with 8 μg/mL polybrene. Expression of innate immune responsive genes including IFIH1 (MDA5), ISG20 and DEFB103 were assessed by real-time qPCR as previously described at 24 hours (primer sequences below). For conditioned media, cells were infected with lenti-DUX4-fl for 12 hours, thoroughly washed 3 times with PBS and switched to fresh growth media to condition for 12 hours. Control conditioned media was produced from cells not infected with any lentivirus. Myoblasts were infected with lenti-GFP in either control conditioned media, lenti-DUX4-fl conditioned media or regular growth media supplemented with 1 μM human β-defensin 3 (Peptides International, Louisville, KY). Expression of innate immune responsive genes were examined after 24 hours.

Gene name	Forward primer sequence	Reverse primer sequence
IFIH1	CTAGCCTGTTCTGGGGAAGA	AGTCGGCACACTTCTTTTGC
ISG20	GAGCGCCTCCTACACAAGAG	CGGATTCTCTGGGAGATTTG
DEFB103	TGTTTGCTTTGCTCTTCCTG	CGCCTCTGACTCTGCAATAA

Beta-defensin 3 and muscle differentiation

Cultured control myoblasts were grown at 50% confluence and treated with 1 μM human β-defensin 3. Equivalent volume of vehicle (water) was added to myoblasts for the control condition. Quadruplicate samples of control- and DEFB103-treated myoblasts were analyzed for global expression changes on HumanHT-12 v4 Expression BeadChip Whole Genome arrays and analyzed as described in main methods. Differential expression of myostatin was confirmed by real-time qPCR.

Gene name	Forward primer sequence	Reverse primer sequence
MSTN	CTGTAACCTTCCCAGGACCA	TCCCTTCTGGATCTTTTTGG

Cultured control myoblasts were grown to confluence and switched to differentiation media[1]. 24 hours later, media was refreshed and either supplemented with 1 μM human

β-defensin 3 or equivalent volume of water. Media was refreshed again at 48 hours. Cells were differentiated for a total of 72 hours. Quadruplicate samples were analyzed by expression microarrays as described above. Differential expression of various markers of muscle differentiation were confirmed by real-time qPCR.

Gene name	Forward primer sequence	Reverse primer sequence
ACTA1	GTACCCTGGGATCGCTGAC	CCGATCCACACCGAGTATTT
CKM	CACCCCAAGTTCGAGGAGAT	AGCGTTGGACACGTCAAATA
CASQ2	AGATTGGGGTGGTGAATGTC	TCCTCAATCCAGTCCTCCAG
MYH2	TTCTCAGGCTTCAAGATTTGG	CTGGAGCTTGCGGAATTTAG
TNNT3	CAAGTTCGAGTTTGGGGAGA	AGCCTTCTTGCTGTGCTTCT
MYF6	GCCAAGTGTTTCCGATCATT	CACGATGGAAGAAAGGCATC
DESMIN	GATCAATCTCCCCATCCAGA	TGGCAGAGGGTCTCTGTCTT

Table 4.1. Representative genes induced by DUX4-fl

Category	Log_2 DUX4-fl Fc*	Log_2 DUX4-s Fc*	Comments
Germline and Stem Cells			
ZSCAN4	8.3	0.0	Genome stability, telomere length
PRAMEF1	8.1	0.1	Melanoma antigen family
SPRYD5	8.0	-0.1	Expressed in oocyte
KHDC1L	8.0	-0.1	KH RNA binding domain
MBD3L2	7.6	0.0	Methyl-CpG-binding protein
ZNF705A	6.8	-0.1	Zinc finger protein
TRIM43	5.8	0.0	Preimplantation embryo
TPRX1	4.5	-0.1	Homeobox protein
ZNF217	4.1	-0.3	Expressed in cancer stem cells
HSPA2	3.7	-0.3	Chaperone, heat shock 70 kd
JUP	3.2	-0.1	expressed in germline and testicular cancers
FGFR3	3.1	0.0	Expressed in spermatogonia
CD24	2.6	-0.4	Stem cell marker
SLC2A14	2.4	0.2	Spermatogenesis
ID2	2.3	0.3	Negative regulator of cell differentiation
PVRL3	2.2	0.4	Spermatid-sertoli junction
HOXB2	2.2	0.0	Anterior-posterior axis development
ZSCAN2	2.2	-0.2	Spermatogenesis and embryonic development
RNA Processing			
SFRS2B	4.2	-0.3	Splicing
THOC4	4.0	-0.2	Splicing, RNA transport
ZNHIT6	3.5	0.3	sno-RNA processing
DBR1	3.4	0.2	RNA lariat debranching enzyme
TFIP11	3.2	0.1	Splicesome assembly
CWC15	2.6	0.1	Spliceosome-associated
ARS2	2.6	-0.2	miRNA processing
PABPN1	2.6	-0.3	PolyA binding
SFRS17A	2.5	0.2	Spliceosome-associated
RMRP	2.3	0.1	Mitochondrial RNA processing
SNIP1	2.1	-0.2	miRNA biogenesis
RPPH1	2.0	0.2	tRNA processing
RNGTT	2.0	-0.6	mRNA processing
Ubiquitin Pathway			
SIAH1	3.7	-0.1	Targets TRF2 telomere maintenance
FBXO33	3.2	0.2	E3 ubiquitin-protein ligase complex
PELI1	2.9	0.1	E3 ligases involved in innate immunity
USP29	2.6	-0.1	Ubiquitin-specific peptidase
ARIH1	2.2	0.8	Ubiquitin-conjugating enzyme E2 binding protein
TRIM23	2.2	0.6	E3 ubiquitin ligase involved in immunity
Immunity and Innate Defense			
DEFB103B	6.4	0.1	Innate defense
IFRD1	3.0	-0.2	Interferon-related developmental regulator
CXADR	2.5	-0.1	Leukocyte migration
CBARA1	2.1	-0.2	T-helper 1-mediated autoreactivity
SON	2.1	-0.3	Viral response
CXCR4	2.0	-0.1	Chemotaxis

Table 4.1 continued

General Transcription			
GTF2F1	3.2	0.3	General transcription factor IIF
MED26	2.1	0.1	RNA Pol II mediator complex
RRN3	2.1	0.1	RNA Pol I preinitiation complex
Cancer Expressed			
CSAG3	5.9	0.1	Chondrosarcoma-associated gene
SLC34A2	5.5	0.0	Breast cancer biomarker
PNMA6B	3.6	-0.2	Paraneoplastic antigen
CSE1L	2.9	0.1	Cellular apoptosis susceptibility protein
AMACR	2.7	0.1	Prostate cancer biomarker
Other			
FLJ45337	3.7	-0.2	Endogenous retrovirus
HNRNPCL1	3.5	-0.1	Nucleosome assembly
SPTY2D1	3.3	-0.3	Suppressor of ty retrotransposons in yeast
MGC10997	2.4	-0.3	Endogenous retrotransposon

*Fc = fold change

Table 4.2. DUX4 highly activates gene families involved in germ cell and early development

Gene Family	Members	Biological Context	Fc range
Preferentially expressed in melanoma family	PRAMEF1 PRAMEF2 PRAMEF4-15 PRAMEF17 PRAMEF20	Cancer-testis antigen (Chang et al., 2011)	9 - 269
Tripartite motif-containing	TRIM43 TRIM48 TRIM49 TRIM53 TRIM64	Testis-expressed, preimplantation embryos (Stanghellini et al., 2009)	27-235
Methyl-binding protein-like	MBD3L2 MBD3L3 MBD3L5	Spermatids & germ cell tumors (Jiang et al., 2002; Jin et al., 2008)	197-310
Zinc finger and SCAN domain containing	ZSCAN4 ZSCAN5B ZSCAN5D	Telomere maintenance in embryonic stem cells (Zalzman et al., 2010)	13-320
Ret-finger Protein-like	RFPL1 RFPL1S RFPL2 RFPL3 RFPL4A RFPL4B	Primate neocortex development (Bonnefont et al., 2008)	20-336
KH homology domain containing	KHDC1 KHDC1L	Oocyte- and embryo-expressed (Pierre et al., 2007)	108-258
Family with sequence similarity 90	FAM90A1 FAM90A2P FAM90A6P FAM90A7	Primate-specific gene family with unknown function (Bosch et al., 2007)	9-19

Table S4.1. Expression Array Analysis of DUX4-fl and DUX4-s in cultured human skeletal muscle

Symbol	Entrez ID	Full.fc	Short.fc	Full.fdr	Short.fdr
RFPL1S	10740	8.40	0.11	0.00	0.42
LOC643263	643263	8.35	-0.13	0.00	0.34
RFPL4B	442247	8.34	-0.10	0.00	0.42
LOC390031	390031	8.33	-0.11	0.00	0.38
ZSCAN4	201516	8.32	0.04	0.00	0.75
LOC340970	340970	8.32	0.03	0.00	0.82
LOC136157	136157	8.30	0.05	0.00	0.73
LOC643445	643445	8.25	0.08	0.00	0.52
LOC729458	729458	8.25	0.01	0.00	0.97
LOC653192	653192	8.23	-0.04	0.00	0.81
LOC645669	645669	8.20	0.09	0.00	0.53
LOC391769	391769	8.19	0.15	0.00	0.25
LOC196120	196120	8.18	0.05	0.00	0.73
LOC651308	651308	8.17	0.04	0.00	0.86
RFPL3	10738	8.14	-0.06	0.00	0.62
PRAMEF1	65121	8.07	0.07	0.00	0.64
LOC100134199	100134199	8.05	0.03	0.00	0.83
SPRYD5	84767	8.04	-0.08	0.00	0.55
LOC284428	284428	8.02	-0.10	0.00	0.51
LOC642362	642362	8.02	-0.01	0.00	0.93
KHDC1L	100129128	8.01	-0.07	0.00	0.60
LOC653656	653656	7.90	-0.14	0.00	0.21
TRIM48	79097	7.88	-0.08	0.00	0.63
LOC653657	653657	7.86	0.19	0.00	0.12
PRAMEF12	390999	7.80	0.13	0.00	0.43
LOC441584	441584	7.78	-0.14	0.00	0.26
LOC730974	730974	7.72	0.02	0.00	0.90
PRAMEF7	441871	7.63	0.11	0.00	0.38
MBD3L2	125997	7.62	0.03	0.00	0.88
LOC440040	440040	7.53	0.08	0.00	0.53
CCNA1	8900	7.53	1.88	0.00	0.00
PRAMEF13	400736	7.42	0.12	0.00	0.33
LOC342900	342900	7.39	0.13	0.00	0.21
LOC340096	340096	7.38	-0.08	0.00	0.66
PRAMEF5	343068	7.35	0.11	0.00	0.62
RFPL2	10739	7.29	0.03	0.00	0.87
PRAMEF9	343070	7.13	-0.07	0.00	0.67
LOC100134006	100134006	7.09	-0.02	0.00	0.89
PRAMEF4	400735	7.06	0.04	0.00	0.83
PRAMEF15	653619	7.00	-0.06	0.00	0.69

Table S4.1 continued

LOC100131392	100131392	6.98	-0.01	0.00	0.97
NP	4860	6.96	0.23	0.00	0.03
LOC399939	399939	6.93	-0.07	0.00	0.61
LOC642148	642148	6.85	0.14	0.00	0.37
LOC729384	729384	6.83	-0.07	0.00	0.55
ZNF705A	440077	6.83	-0.06	0.00	0.62
C6orf148	80759	6.76	-0.10	0.00	0.50
TRIM49	57093	6.55	-0.02	0.00	0.91
DEFB103B	414325	6.44	0.07	0.00	0.64
PRAMEF2	65122	6.44	-0.05	0.00	0.74
RFPL1	5988	6.26	0.34	0.00	0.01
LOC100133984	100133984	6.20	-0.03	0.00	0.84
LOC642127	642127	6.11	0.04	0.00	0.81
CA2	760	6.09	0.10	0.00	0.50
PRAMEF10	343071	6.06	0.01	0.00	0.97
LOC646698	646698	6.01	0.11	0.00	0.48
LOC729516	729516	5.95	0.00	0.00	0.98
PRAMEF11	440560	5.94	0.12	0.00	0.37
CSAG3	389903	5.87	0.09	0.00	0.55
PRAMEF6	440561	5.83	-0.06	0.00	0.66
LOC391764	391764	5.82	0.12	0.00	0.34
TRIM43	129868	5.81	0.02	0.00	0.93
LOC391742	391742	5.73	0.16	0.00	0.14
LOC391766	391766	5.72	-0.06	0.00	0.68
ZNF296	162979	5.54	0.18	0.00	0.13
SLC34A2	10568	5.51	-0.03	0.00	0.88
LOC391767	391767	5.49	0.06	0.00	0.80
LOC729368	729368	5.42	-0.06	0.00	0.69
LOC440563	440563	5.31	0.07	0.00	0.69
LOC646754	646754	5.11	-0.15	0.00	0.32
LOC654101	654101	5.03	0.09	0.00	0.62
LOC729731	729731	5.01	0.10	0.00	0.45
HIST2H3A	333932	4.95	-0.05	0.00	0.79
TRIM64	120146	4.94	-0.17	0.00	0.15
LOC402207	402207	4.90	-0.02	0.00	0.92
LOC729700	729700	4.82	-0.18	0.00	0.10
LOC645558	645558	4.80	-0.05	0.00	0.76
LOC642219	642219	4.80	0.00	0.00	1.00
PRAMEF20	645425	4.80	0.00	0.00	0.98
HBA1	3039	4.79	0.08	0.00	0.56
TRIM53	642569	4.78	0.08	0.00	0.59
LOC399940	399940	4.73	-0.04	0.00	0.83

Table S4.1 continued

HBA2	3040	4.72	0.09	0.00	0.42
LOC646103	646103	4.66	-0.14	0.00	0.39
LOC732393	732393	4.64	0.04	0.00	0.84
LOC100133446	100133446	4.63	0.12	0.00	0.35
LOC100131539	100131539	4.63	0.07	0.00	0.72
C12orf50	160419	4.52	-0.02	0.00	0.89
OR2T34	127068	4.52	-0.05	0.00	0.69
TPRX1	284355	4.48	-0.05	0.00	0.69
LOC402199	402199	4.39	0.08	0.00	0.61
LOC646066	646066	4.39	-0.03	0.00	0.88
ART3	419	4.36	0.09	0.00	0.50
RFPL4A	342931	4.35	-0.09	0.00	0.52
LOC401860	401860	4.27	-0.13	0.00	0.34
NXF1	10482	4.23	-0.35	0.00	0.00
LOC729706	729706	4.23	-0.02	0.00	0.93
PRAMEF17	649345	4.22	0.10	0.00	0.56
SFRS2B	10929	4.22	-0.33	0.00	0.00
RN5S9	100169760	4.19	0.76	0.00	0.00
PPP2R2B	5521	4.13	-0.11	0.00	0.41
ZNF217	7764	4.11	-0.33	0.00	0.01
ENTPD8	377841	4.07	0.04	0.00	0.82
LOC647827	647827	4.05	-0.05	0.00	0.78
THOC4	10189	4.03	-0.19	0.00	0.11
LOC729694	729694	4.03	-0.17	0.00	0.32
LOC440053	440053	3.92	-0.10	0.00	0.46
LOC440041	440041	3.89	0.02	0.00	0.91
HBEGF	1839	3.87	-0.08	0.00	0.61
NEUROG2	63973	3.86	0.00	0.00	0.98
PANX2	56666	3.83	0.20	0.00	0.11
ZNF280A	129025	3.80	0.06	0.00	0.69
LOC647366	647366	3.78	0.00	0.00	0.99
LOC285697	285697	3.78	-0.07	0.00	0.71
LOC441081	441081	3.76	0.06	0.00	0.72
LOC342933	342933	3.75	-0.10	0.00	0.43
EGR1	1958	3.75	0.03	0.00	0.86
DYNC2H1	79659	3.72	0.05	0.00	0.75
LOC100128202	100128202	3.71	-0.12	0.00	0.44
PRAMEF8	391002	3.70	0.17	0.00	0.25
SIAH1	6477	3.70	-0.08	0.00	0.52
FLJ45337	400754	3.69	-0.25	0.00	0.03
HSPA2	3306	3.67	-0.25	0.00	0.05
ODC1	4953	3.66	-0.20	0.00	0.07

Table S4.1 continued

LOC730167	730167	3.66	1.26	0.00	0.00
FAM90A1	55138	3.64	0.03	0.00	0.88
LOC653194	653194	3.63	-0.02	0.00	0.94
PNMA6B	728513	3.60	-0.22	0.00	0.17
LOC100132564	100132564	3.59	0.40	0.00	0.01
PRR4	11272	3.56	0.16	0.00	0.12
LOC653978	653978	3.55	0.19	0.00	0.16
HSPA1A	3303	3.54	-0.02	0.00	0.93
LOC729698	729698	3.54	-0.07	0.00	0.69
ZNHIT6	54680	3.52	0.33	0.00	0.00
NT5C1B	93034	3.50	-0.04	0.00	0.83
HNRNPCL1	343069	3.48	-0.10	0.00	0.50
CTGLF7	728127	3.44	-0.29	0.00	0.01
HSPA1B	3304	3.40	0.36	0.00	0.01
SLC2A3	6515	3.39	0.10	0.00	0.42
DBR1	51163	3.38	0.22	0.00	0.11
KLHL15	80311	3.37	0.77	0.00	0.00
LOC650167	650167	3.36	0.05	0.00	0.72
LOC100130652	100130652	3.33	0.11	0.00	0.47
SPTY2D1	144108	3.29	-0.29	0.00	0.01
SDHALP1	255812	3.28	0.12	0.00	0.28
FBXO33	254170	3.22	0.22	0.00	0.08
GTF2F1	2962	3.21	0.29	0.00	0.01
FAM90A7	441317	3.20	-0.21	0.00	0.20
TFIP11	24144	3.19	0.11	0.00	0.30
PRAMEF14	729528	3.19	-0.17	0.00	0.29
JUP	3728	3.17	-0.11	0.00	0.47
RAB6B	51560	3.17	0.01	0.00	0.98
CLDN14	23562	3.15	-0.30	0.00	0.00
LOC653111	653111	3.14	-0.10	0.00	0.54
FGFR3	2261	3.11	0.04	0.00	0.77
LOC642446	642446	3.10	0.19	0.00	0.17
LOC649330	649330	3.10	-0.27	0.00	0.07
SOX9	6662	3.10	-0.31	0.00	0.00
KBTBD8	84541	3.08	0.01	0.00	0.96
LOC727828	727828	3.05	-0.01	0.00	0.97
PPP1R14C	81706	3.05	-0.01	0.00	0.95
LOC729724	729724	3.04	0.11	0.00	0.35
LOC652433	652433	3.02	-0.03	0.00	0.87
LOC391761	391761	3.01	-0.06	0.00	0.73
IFRD1	3475	2.99	-0.19	0.00	0.13
LOC342934	342934	2.99	-0.06	0.00	0.77

Table S4.1 continued

DBNDD2	55861	2.99	0.23	0.00	0.04
MGC61598	441478	2.97	0.10	0.00	0.48
CSE1L	1434	2.94	0.07	0.00	0.59
NEFM	4741	2.93	0.85	0.00	0.00
LOC650236	650236	2.92	-0.06	0.00	0.73
LOC100130311	100130311	2.91	0.14	0.00	0.25
EOMES	8320	2.90	0.06	0.00	0.60
LOC645373	645373	2.89	-0.07	0.00	0.74
PELI1	57162	2.89	0.15	0.00	0.27
LOC285299	285299	2.84	-0.03	0.00	0.86
LOC652349	652349	2.84	0.02	0.00	0.93
LOC400464	400464	2.83	-0.08	0.00	0.58
LOC391747	391747	2.79	-0.10	0.00	0.34
BAMBI	25805	2.78	0.11	0.00	0.37
PELI2	57161	2.78	-0.37	0.00	0.00
T1560	387335	2.77	0.02	0.00	0.92
KCNH4	23415	2.75	-0.07	0.00	0.67
AMACR	23600	2.74	0.08	0.00	0.55
SLC3A1	6519	2.74	0.00	0.00	0.98
DYNLL2	140735	2.74	0.65	0.00	0.00
LOC642843	642843	2.73	-0.06	0.00	0.74
LOC100129053	100129053	2.72	0.10	0.00	0.50
CCNJ	54619	2.69	0.08	0.00	0.56
BZW2	28969	2.68	-0.40	0.00	0.00
CWC15	51503	2.64	0.09	0.00	0.48
CD24	100133941	2.62	-0.36	0.00	0.00
C9orf61	9413	2.61	-0.26	0.00	0.07
DENND2C	163259	2.61	-0.27	0.00	0.03
ARS2	51593	2.60	-0.20	0.00	0.09
YRDC	79693	2.60	0.00	0.00	1.00
USP29	57663	2.60	-0.09	0.00	0.49
EYA3	2140	2.57	-0.02	0.00	0.92
LOC646914	646914	2.57	-0.02	0.00	0.88
PABPN1	8106	2.57	-0.26	0.00	0.02
MGC40489	146880	2.54	-0.46	0.00	0.00
C11orf82	220042	2.53	-0.27	0.00	0.03
C14orf102	55051	2.53	0.18	0.00	0.12
FAM107B	83641	2.53	-0.18	0.00	0.17
CYCSL1	157317	2.51	0.37	0.00	0.00
DEFB103A	55894	2.50	-0.04	0.00	0.87
LOC646508	646508	2.50	-0.04	0.00	0.80
CSRNP3	80034	2.50	-0.08	0.00	0.63

Table S4.1 continued

CXADR	1525	2.49	-0.07	0.00	0.69
C13orf34	79866	2.49	0.00	0.00	0.99
LOC100134322	100134322	2.49	0.20	0.00	0.10
PRAMEF21	391001	2.49	-0.07	0.00	0.64
LOC728450	728450	2.47	0.13	0.00	0.28
SFRS17A	8227	2.46	0.24	0.00	0.05
FLJ45139	400867	2.45	0.10	0.00	0.54
C6orf117	112609	2.45	-0.07	0.00	0.61
NOLC1	9221	2.44	0.50	0.00	0.00
SYNJ1	8867	2.43	-0.08	0.00	0.64
MGC10997	84741	2.43	-0.29	0.00	0.01
LOC649563	649563	2.42	-0.06	0.00	0.70
KPNA2	3838	2.42	-0.51	0.00	0.00
MIR2278	100313780	2.41	0.10	0.00	0.58
ZNF622	90441	2.41	0.01	0.00	0.93
CTR9	9646	2.40	-0.02	0.00	0.89
NCRNA00092	100188953	2.39	-0.11	0.00	0.46
FAM46C	54855	2.39	-0.13	0.00	0.52
SLC2A14	144195	2.37	0.25	0.00	0.07
PRRG4	79056	2.37	0.21	0.00	0.14
SLIT2	9353	2.37	0.85	0.00	0.00
CRY1	1407	2.34	-0.13	0.00	0.29
ID2	3398	2.34	0.28	0.00	0.06
PRAMEF19	645414	2.34	-0.04	0.00	0.78
NFYA	4800	2.33	0.92	0.00	0.00
LOC732416	732416	2.33	0.06	0.00	0.70
NUP50	10762	2.32	1.03	0.00	0.00
LOC645137	645137	2.32	-0.15	0.00	0.23
LOC651709	651709	2.30	0.09	0.00	0.45
TMEM185A	84548	2.29	0.23	0.00	0.04
LOC648533	648533	2.28	0.04	0.00	0.89
WDR47	22911	2.28	0.14	0.00	0.31
RMRP	6023	2.27	0.07	0.00	0.66
C8orf33	65265	2.25	-0.06	0.00	0.66
DUSP12	11266	2.25	0.10	0.00	0.38
FAM90A9	441327	2.25	0.02	0.00	0.92
ARIH1	25820	2.24	0.78	0.00	0.00
TRIM23	373	2.23	0.63	0.00	0.00
ADPGK	83440	2.23	0.09	0.00	0.45
PVRL3	25945	2.22	0.45	0.00	0.00
ZNF214	7761	2.22	0.17	0.00	0.32
HSPH1	10808	2.22	-0.49	0.00	0.00

Table S4.1 continued

PIM1	5292	2.21	-0.68	0.00	0.00
PSPN	5623	2.20	-0.04	0.00	0.83
HOXB2	3212	2.20	-0.04	0.00	0.78
LOC100133588	100133588	2.20	0.04	0.00	0.85
C1orf63	57035	2.19	-0.14	0.00	0.54
STK3	6788	2.19	-0.21	0.00	0.13
HEY1	23462	2.18	-0.04	0.00	0.78
LOC728429	728429	2.18	-0.03	0.00	0.88
HNRPDL	9987	2.18	0.02	0.00	0.92
LOC727846	727846	2.18	0.00	0.00	0.99
LOC391045	391045	2.17	0.35	0.00	0.02
UBL3	5412	2.16	-0.81	0.00	0.00
ZSCAN2	54993	2.15	-0.20	0.00	0.08
PNO1	56902	2.15	0.25	0.00	0.02
GPR37	2861	2.12	0.22	0.00	0.15
TSPAN13	27075	2.12	-0.15	0.00	0.28
SNIP1	79753	2.10	-0.17	0.00	0.11
MED26	9441	2.10	0.14	0.00	0.29
C6orf191	253582	2.09	-0.06	0.00	0.72
LOC645381	645381	2.09	-0.13	0.00	0.29
PPP1R15A	23645	2.08	0.57	0.00	0.00
RRN3	54700	2.07	0.11	0.00	0.39
CBARA1	10367	2.06	-0.15	0.00	0.26
NGDN	25983	2.06	0.41	0.00	0.00
MED31	51003	2.05	0.31	0.00	0.01
SON	6651	2.05	-0.28	0.00	0.03
STX6	10228	2.05	-0.47	0.00	0.00
C1orf55	163859	2.05	-0.36	0.00	0.01
SGK	6446	2.04	-0.11	0.00	0.38
RPPH1	85495	2.04	0.23	0.00	0.07
CEP78	84131	2.04	-0.14	0.00	0.22
CASP6	839	2.04	0.23	0.00	0.07
ARID3B	10620	2.02	-0.07	0.00	0.57
AVPI1	60370	2.02	0.19	0.00	0.08
RNGTT	8732	2.02	-0.59	0.00	0.00
KIAA0020	9933	2.02	0.02	0.00	0.92
SLC25A44	9673	2.02	-0.08	0.00	0.54
RBM12	10137	2.01	-0.37	0.00	0.00
CXCR4	7852	2.00	-0.05	0.00	0.77
PDSS1	23590	2.00	-0.07	0.00	0.60
ISOC1	51015	2.00	-0.62	0.00	0.00
SERTAD1	29950	2.00	0.41	0.00	0.00

Table S4.1 continued

CCDC58	131076	2.00	0.08	0.00	0.61
DNAJC25	548645	2.00	-0.51	0.00	0.00
LSG1	55341	1.99	-0.06	0.00	0.64
HSPA6	3310	1.99	0.13	0.00	0.26
FRG2B	441581	1.99	0.13	0.00	0.30
CD9	928	1.99	0.00	0.00	0.98
LOC652080	652080	1.99	0.17	0.00	0.21
RAB11FIP1	80223	1.99	-0.48	0.00	0.00
RBBP6	5930	1.99	-0.13	0.00	0.44
INO80C	125476	1.98	-0.11	0.00	0.48
TRA2A	29896	1.98	-0.11	0.00	0.41
LYAR	55646	1.98	0.00	0.00	1.00
C1QTNF3	114899	1.97	-0.02	0.00	0.88
KLC1	3831	1.97	-0.07	0.00	0.64
LOC399988	399988	1.97	-0.41	0.00	0.00
MED10	84246	1.97	0.60	0.00	0.00
LOC391763	391763	1.96	0.09	0.00	0.62
RGS4	5999	1.95	-0.48	0.00	0.00
POLR3K	51728	1.95	0.06	0.00	0.70
OSBPL8	114882	1.95	1.05	0.00	0.00
PNN	5411	1.95	0.04	0.00	0.75
TUBB2C	10383	1.95	-0.54	0.00	0.00
SNAI1	6615	1.94	-0.04	0.00	0.83
LOC651390	651390	1.94	-0.08	0.00	0.64
EXOSC10	5394	1.94	0.28	0.00	0.16
GPBAR1	151306	1.94	-0.13	0.00	0.35
NEFH	4744	1.94	0.10	0.00	0.50
PEG10	23089	1.94	-0.10	0.00	0.61
LOC643336	643336	1.93	1.02	0.00	0.00
HNRPA1L-2	664709	1.92	-0.90	0.00	0.00
FRAT2	23401	1.92	-0.05	0.00	0.70
HSPB3	8988	1.91	-0.79	0.00	0.00
CRLF3	51379	1.91	-0.14	0.00	0.18
MIR503	574506	1.91	-0.14	0.00	0.38
POLR1B	84172	1.90	-0.13	0.00	0.29
LOC643731	643731	1.90	0.00	0.00	1.00
CCDC59	29080	1.89	0.25	0.00	0.04
ETNK1	55500	1.89	0.06	0.00	0.74
NOP58	51602	1.89	0.22	0.00	0.05
BRIX1	55299	1.89	0.30	0.00	0.02
SNRNP70	6625	1.88	0.02	0.00	0.91
ELOF1	84337	1.88	-0.30	0.00	0.01

Table S4.1 continued

CCNT2	905	1.88	-0.32	0.00	0.00
NANS	54187	1.87	-0.01	0.00	0.97
LOC100129630	100129630	1.87	-0.10	0.00	0.47
TAF4B	6875	1.87	-0.04	0.00	0.81
SRP19	6728	1.87	0.23	0.00	0.07
SGCG	6445	1.86	0.13	0.00	0.33
CCNE1	898	1.86	-0.13	0.00	0.45
SLC40A1	30061	1.86	-0.42	0.00	0.00
PKIB	5570	1.86	-0.17	0.00	0.21
LARP1B	55132	1.86	-0.26	0.00	0.01
SNORD56	26793	1.85	-0.17	0.00	0.25
SIRT1	23411	1.84	0.21	0.00	0.17
KIAA0114	57291	1.84	-0.32	0.00	0.02
LOC399937	399937	1.84	0.00	0.00	1.00
CLK1	1195	1.84	0.08	0.00	0.72
LOC391092	391092	1.84	0.11	0.00	0.48
SEC61A2	55176	1.84	0.21	0.00	0.07
KIF21A	55605	1.84	0.02	0.00	0.93
LOC651816	651816	1.84	-0.54	0.00	0.00
KCNA1	3736	1.83	0.03	0.00	0.89
PPM1B	5495	1.83	-0.03	0.00	0.84
MYLIP	29116	1.82	-0.30	0.00	0.01
KATNA1	11104	1.82	0.11	0.00	0.39
MBD2	8932	1.82	-0.08	0.00	0.58
MED13	9969	1.82	0.03	0.00	0.86
SOX4	6659	1.81	-0.22	0.00	0.11
SERPINI1	5274	1.81	0.04	0.00	0.82
LOC389633	389633	1.81	-0.09	0.00	0.54
CHORDC1	26973	1.80	0.13	0.00	0.30
ARC	23237	1.80	-0.19	0.00	0.10
INSM1	3642	1.80	0.23	0.00	0.07
RBM39	9584	1.80	0.26	0.00	0.04
LOC642538	642538	1.79	0.13	0.00	0.36
TESK2	10420	1.79	0.13	0.00	0.30
PDRG1	81572	1.78	0.06	0.00	0.69
KLF17	128209	1.78	0.08	0.00	0.60
FAM90A12	441332	1.77	0.05	0.00	0.78
LOC388275	388275	1.77	-0.94	0.00	0.00
ZNF365	22891	1.77	-0.37	0.00	0.00
RNF122	79845	1.77	0.19	0.00	0.14
KDM5B	10765	1.76	-0.11	0.00	0.40
HOXB6	3216	1.76	0.06	0.00	0.68

Table S4.1 continued

C21orf91	54149	1.76	0.08	0.00	0.60
FAM90A5	441315	1.76	0.10	0.00	0.37
FAM133B	257415	1.76	0.09	0.00	0.53
NIPSNAP3A	25934	1.75	-0.07	0.00	0.63
RNF152	220441	1.75	0.23	0.00	0.11
C13orf31	144811	1.75	0.01	0.00	0.93
ELOVL4	6785	1.75	-0.32	0.00	0.00
TP53BP2	7159	1.75	-0.04	0.00	0.80
RGMB	285704	1.74	-0.08	0.00	0.58
SNORD57	26792	1.74	-0.14	0.00	0.30
B3GNT2	10678	1.74	0.25	0.00	0.03
RHPN2	85415	1.74	0.16	0.00	0.15
YARS2	51067	1.73	0.14	0.00	0.26
SHISA2	387914	1.73	-0.29	0.00	0.02
IRX5	10265	1.73	0.19	0.00	0.14
ALG13	79868	1.73	0.27	0.00	0.02
STAU1	6780	1.73	0.22	0.00	0.06
EAF1	85403	1.72	0.01	0.00	0.96
LOC440258	440258	1.72	0.22	0.00	0.08
HNRPA1P4	389674	1.72	-1.13	0.00	0.00
LOC730081	730081	1.72	0.29	0.00	0.02
CDKN2AIP	55602	1.72	0.33	0.00	0.01
LOC440061	440061	1.71	-0.25	0.00	0.11
C16orf80	29105	1.71	-0.18	0.00	0.13
CTH	1491	1.71	0.27	0.00	0.02
DDX47	51202	1.71	0.25	0.00	0.02
TFB2M	64216	1.71	-0.01	0.00	0.94
C1orf52	148423	1.70	-0.22	0.00	0.03
C14orf138	79609	1.70	0.30	0.00	0.01
AURKAPS1	6791	1.70	0.07	0.00	0.58
ARPP-21	10777	1.70	0.01	0.00	0.95
PRPF18	8559	1.70	0.34	0.00	0.01
WDR43	23160	1.69	0.98	0.00	0.00
SLC25A4	291	1.69	0.23	0.00	0.09
EIF4A3	9775	1.68	-0.02	0.00	0.88
SNORD68	606500	1.68	0.10	0.00	0.44
LOC729423	729423	1.68	-0.81	0.00	0.00
MAP2	4133	1.68	1.96	0.00	0.00
TUBB4Q	56604	1.68	-0.29	0.00	0.03
MAD2L1BP	9587	1.67	0.49	0.00	0.00
NUP98	4928	1.67	-0.17	0.00	0.12
TDG	6996	1.67	0.17	0.00	0.18

Table S4.1 continued

TCEB3	6924	1.67	0.53	0.00	0.00
PTP4A1	7803	1.67	-0.13	0.00	0.31
HSPA8	3312	1.67	-0.35	0.00	0.00
UBL5	59286	1.67	-0.11	0.00	0.46
C2orf56	55471	1.66	0.28	0.00	0.02
MAST4	375449	1.66	0.03	0.00	0.83
BCL2L12	83596	1.66	-0.34	0.00	0.00
INVS	27130	1.65	0.43	0.00	0.00
CDC42	998	1.65	0.56	0.00	0.00
ATF3	467	1.65	1.16	0.00	0.00
HEY2	23493	1.64	0.10	0.00	0.55
ZSWIM6	57688	1.64	-0.48	0.00	0.00
C3orf58	205428	1.64	0.02	0.00	0.94
LOC401097	401097	1.64	-0.12	0.00	0.40
EGLN1	54583	1.64	1.21	0.00	0.00
CUGBP1	10658	1.63	0.30	0.00	0.01
LOC642678	642678	1.63	0.26	0.00	0.02
C15orf60	283677	1.63	-0.03	0.00	0.86
PHAX	51808	1.63	0.50	0.00	0.00
MED6	10001	1.63	0.23	0.00	0.03
CTNNAL1	8727	1.63	-0.18	0.00	0.14
PDK3	5165	1.63	0.77	0.00	0.00
PLEKHB2	55041	1.62	0.05	0.00	0.70
NUDT11	55190	1.62	-0.58	0.00	0.00
BDNF	627	1.62	0.04	0.00	0.87
NGLY1	55768	1.61	0.30	0.00	0.00
ZNF705D	728957	1.61	-0.03	0.00	0.89
MED15	51586	1.61	-0.07	0.00	0.64
ZIK1	284307	1.61	0.46	0.00	0.00
PRPF40A	55660	1.61	0.44	0.00	0.00
PMAIP1	5366	1.61	0.64	0.00	0.00
KRTAP2-1	81872	1.61	-0.01	0.00	0.97
SNORA67	26781	1.60	-0.02	0.00	0.91
NKIRAS1	28512	1.60	-0.22	0.00	0.04
SAMD8	142891	1.60	0.67	0.00	0.00
LOC728408	728408	1.60	0.20	0.00	0.12
MAPKAP1	79109	1.60	0.29	0.00	0.01
ECD	11319	1.60	0.44	0.00	0.00
TOPORS	10210	1.60	0.42	0.00	0.00
RTN4	57142	1.59	0.26	0.00	0.04
ARHGAP19	84986	1.59	-0.26	0.00	0.03
SH3GL2	6456	1.59	0.07	0.00	0.62

Table S4.1 continued

MYBPH	4608	1.59	-0.45	0.00	0.01
NT5DC3	51559	1.59	-0.47	0.00	0.00
SNRPN	6638	1.59	-0.18	0.00	0.16
GJA1	2697	1.59	-0.18	0.00	0.22
LOC400013	400013	1.59	0.03	0.00	0.82
HNRNPM	4670	1.59	-0.15	0.00	0.18
STX3	6809	1.58	0.36	0.00	0.01
LOC644914	644914	1.58	-0.07	0.00	0.81
LOC100133836	100133836	1.58	0.11	0.00	0.49
GTF2B	2959	1.57	0.35	0.00	0.00
SCML1	6322	1.57	0.19	0.00	0.17
TBPL1	9519	1.57	0.21	0.00	0.09
ZNF551	90233	1.57	0.00	0.00	1.00
SFRS10	6434	1.57	-0.75	0.00	0.00
OXR1	55074	1.56	-0.72	0.00	0.00
BHLHE22	27319	1.56	0.10	0.00	0.47
OR6X1	390260	1.56	-0.13	0.00	0.35
LMO4	8543	1.56	0.35	0.00	0.00
LOC645166	645166	1.55	0.19	0.00	0.08
STIL	6491	1.55	-0.48	0.00	0.00
LOC100133760	100133760	1.55	0.16	0.00	0.20
FZD7	8324	1.55	-0.13	0.00	0.33
BTAF1	9044	1.55	0.07	0.00	0.65
CCNC	892	1.55	0.10	0.00	0.42
DNAJB1	3337	1.55	-0.33	0.00	0.00
ASB7	140460	1.55	0.27	0.00	0.01
DDX39	10212	1.55	0.18	0.00	0.17
LOC85389	85389	1.55	0.01	0.00	0.97
APIP	51074	1.54	-0.01	0.00	0.97
ZNF330	27309	1.54	-0.07	0.00	0.61
ABL1	25	1.54	0.15	0.00	0.28
WDR45L	56270	1.54	0.36	0.00	0.00
CDR2	1039	1.54	-0.31	0.00	0.01
LOC648040	648040	1.54	0.05	0.00	0.74
TMSB15A	11013	1.54	-1.13	0.00	0.00
DDX3X	1654	1.53	-0.17	0.00	0.22
TAF5	6877	1.53	-0.21	0.00	0.08
SNORA80	677846	1.53	0.09	0.00	0.50
LOC731049	731049	1.53	-0.71	0.00	0.00
RBM14	10432	1.53	-0.22	0.00	0.04
CNNM4	26504	1.52	0.02	0.00	0.90
GLS	2744	1.52	-0.81	0.00	0.00

Table S4.1 continued

METTL7B	196410	1.52	-0.55	0.00	0.00
RSRC2	65117	1.51	-0.12	0.00	0.31
FAM90A17	728746	1.51	-0.10	0.00	0.47
SLU7	10569	1.51	0.72	0.00	0.00
LOC654256	654256	1.51	0.00	0.00	0.99
LOC730820	730820	1.51	0.26	0.00	0.02
GLMN	11146	1.51	-0.16	0.00	0.24
LOC728640	728640	1.51	-0.16	0.00	0.17
NUP54	53371	1.50	0.03	0.00	0.86
AHR	196	1.50	1.64	0.00	0.00
H2AFZ	3015	1.50	-0.39	0.00	0.00
C1orf128	57095	1.50	-0.07	0.00	0.58
RNF4	6047	1.50	-0.25	0.00	0.03
TFAP2C	7022	1.50	-0.19	0.00	0.14
C1orf185	284546	1.50	0.06	0.00	0.77
BRD2	6046	1.50	0.08	0.00	0.48
RAB3IP	117177	1.49	0.12	0.00	0.37
PITX1	5307	1.49	0.25	0.00	0.05
ALG11	440138	1.49	0.49	0.00	0.00
TCP1	6950	1.49	-0.06	0.00	0.75
DHX9	1660	1.49	-0.02	0.00	0.92
NOL11	25926	1.49	-0.04	0.00	0.77
TPM3	7170	1.49	0.58	0.00	0.00
LOC283116	283116	1.49	0.12	0.00	0.40
RYBP	23429	1.49	-0.35	0.00	0.00
SNORD43	26807	1.49	0.04	0.00	0.85
RHOBTB1	9886	1.49	-0.45	0.00	0.00
LOC641802	641802	1.49	0.38	0.00	0.00
HTRA4	203100	1.48	0.22	0.00	0.06
ZNF263	10127	1.48	-0.14	0.00	0.22
LOC645232	645232	1.48	0.00	0.00	1.00
DIO3	1735	1.48	-0.07	0.00	0.69
SRFBP1	153443	1.48	0.45	0.00	0.00
DNAJA1	3301	1.47	0.64	0.00	0.00
PPFIBP2	8495	1.47	0.00	0.00	0.99
NDEL1	81565	1.47	-0.02	0.00	0.88
RRP15	51018	1.47	0.17	0.00	0.11
SUPT6H	6830	1.46	0.26	0.00	0.08
EIF1	10209	1.46	-0.22	0.00	0.08
C1orf187	374946	1.46	-0.02	0.00	0.89
SLC35F3	148641	1.46	-0.09	0.00	0.61
LOC732387	732387	1.46	-0.17	0.00	0.23

Table S4.1 continued

DPPA3	359787	1.46	0.00	0.00	0.99
BCCIP	56647	1.45	0.11	0.00	0.38
FBXW7	55294	1.45	0.14	0.00	0.29
LOC732360	732360	1.45	0.16	0.00	0.20
ILF2	3608	1.45	-0.22	0.00	0.04
TAF7	6879	1.45	0.26	0.00	0.05
FBXO28	23219	1.45	0.33	0.00	0.00
LOC648390	648390	1.45	-0.17	0.00	0.13
LOC100134083	100134083	1.44	0.07	0.00	0.74
CDS1	1040	1.44	-0.04	0.00	0.86
TMPO	7112	1.44	-0.16	0.00	0.27
SLC10A4	201780	1.44	0.13	0.00	0.33
SFRS2	6427	1.44	-0.55	0.00	0.00
LOC100129267	100129267	1.44	-0.10	0.00	0.46
CSRP2	1466	1.44	-0.83	0.00	0.00
HNRNPA2B1	3181	1.44	0.13	0.00	0.30
CLP1	10978	1.43	0.35	0.00	0.00
MTF2	22823	1.43	0.24	0.00	0.04
C13orf27	93081	1.43	0.05	0.00	0.79
MEX3C	51320	1.43	0.09	0.00	0.48
CBX4	8535	1.43	0.31	0.00	0.01
DDX21	9188	1.43	-0.17	0.00	0.12
SFRS15	57466	1.42	0.17	0.00	0.15
AHCTF1	25909	1.42	0.29	0.00	0.01
STARD7	56910	1.42	0.01	0.00	0.96
LOC347376	347376	1.42	-0.32	0.00	0.10
C6orf66	29078	1.42	0.12	0.00	0.34
LOC644330	644330	1.41	0.66	0.00	0.00
ABCG1	9619	1.41	-0.03	0.00	0.87
NOV	4856	1.41	-0.05	0.00	0.78
GFM1	85476	1.41	-0.80	0.00	0.00
LOC652595	652595	1.41	-0.10	0.00	0.48
ZNF281	23528	1.40	0.00	0.00	0.99
ARID4B	51742	1.40	0.57	0.00	0.00
LOC645233	645233	1.40	0.10	0.00	0.46
MYOG	4656	1.40	-0.86	0.00	0.00
OSR2	116039	1.40	0.14	0.00	0.24
LOC727758	727758	1.40	0.34	0.00	0.01
HIC2	23119	1.40	-0.08	0.00	0.50
PHLPP2	23035	1.39	0.01	0.00	0.94
LOC728153	728153	1.39	-0.12	0.00	0.36
PHLPP1	23239	1.39	-0.11	0.00	0.34

Table S4.1 continued

SBNO1	55206	1.39	0.91	0.00	0.00
ZNF574	64763	1.39	-0.03	0.00	0.87
BAGE5	85316	1.39	0.31	0.00	0.01
EML4	27436	1.39	-0.89	0.00	0.00
SHFM1	7979	1.39	-0.03	0.00	0.86
SLC12A2	6558	1.39	-0.76	0.00	0.00
RND3	390	1.39	0.93	0.00	0.00
MGAT4C	25834	1.39	-0.08	0.00	0.58
ERN1	2081	1.39	-0.08	0.00	0.59
C16orf87	388272	1.38	-0.05	0.00	0.73
MBIP	51562	1.38	0.28	0.00	0.01
SUV420H1	51111	1.38	0.32	0.00	0.01
MFSD4	148808	1.38	-0.12	0.00	0.39
TSC22D2	9819	1.38	-0.27	0.00	0.07
FNBP1L	54874	1.38	-0.89	0.00	0.00
SLC25A13	10165	1.38	-0.09	0.00	0.50
GAB2	9846	1.38	-0.23	0.00	0.05
BMP2K	55589	1.38	0.16	0.00	0.25
CCK	885	1.38	-0.05	0.00	0.72
GCC1	79571	1.38	0.35	0.00	0.00
DOHH	83475	1.38	0.06	0.00	0.65
ZNF721	170960	1.37	-0.08	0.00	0.56
MGC39900	286527	1.37	-0.61	0.00	0.00
KCTD5	54442	1.37	-0.25	0.00	0.01
CDO1	1036	1.37	0.12	0.00	0.34
SNHG1	23642	1.37	-0.25	0.00	0.02
RPF1	80135	1.37	0.22	0.00	0.06
ZNF408	79797	1.37	0.06	0.00	0.64
PFKFB3	5209	1.37	0.25	0.00	0.03
C8orf79	57604	1.37	0.05	0.00	0.76
ZNF256	10172	1.37	0.09	0.00	0.50
VGLL2	245806	1.37	0.13	0.00	0.33
CCDC49	54883	1.36	0.14	0.00	0.26
AMD1	262	1.36	-0.76	0.00	0.00
RAPGEF2	9693	1.36	0.06	0.00	0.79
SNORD36A	26815	1.36	0.19	0.00	0.32
BUD31	8896	1.36	0.19	0.00	0.09
FBXL12	54850	1.36	0.17	0.00	0.24
SNORD55	26811	1.36	0.15	0.00	0.20
KIAA1429	25962	1.36	0.12	0.00	0.32
LOC729200	729200	1.35	0.17	0.00	0.16
DLEU1	10301	1.35	-0.35	0.00	0.00

Table S4.1 continued

BAZ1A	11177	1.35	0.40	0.00	0.00
TXNDC12	51060	1.35	-0.22	0.00	0.06
SDC2	6383	1.35	-0.09	0.00	0.52
ROCK1	6093	1.35	0.15	0.00	0.26
IER2	9592	1.35	0.06	0.00	0.74
MRPL44	65080	1.35	0.62	0.00	0.00
HIST2H2BE	8349	1.35	0.27	0.00	0.01
SRP14P1	390284	1.35	0.17	0.00	0.17
C13orf15	28984	1.35	-0.26	0.00	0.03
CRHBP	1393	1.34	-0.12	0.00	0.39
LOC642333	642333	1.34	0.03	0.00	0.89
KIAA0922	23240	1.34	0.04	0.00	0.78
CAMK2G	818	1.34	-0.17	0.00	0.18
TNFRSF10D	8793	1.34	0.14	0.00	0.32
TC2N	123036	1.34	-0.12	0.00	0.41
FBXO44	93611	1.33	-0.03	0.00	0.87
C1orf182	128229	1.33	0.21	0.00	0.10
NRBF2	29982	1.33	0.05	0.00	0.75
TMEM119	338773	1.33	-0.54	0.00	0.00
TFAM	7019	1.33	-0.20	0.00	0.09
ADNP2	22850	1.33	-0.11	0.00	0.32
DUX4	22947	1.33	-0.13	0.00	0.35
LOC100132418	100132418	1.33	0.07	0.00	0.56
FAM89A	375061	1.32	0.49	0.00	0.00
DOPEY1	23033	1.32	0.00	0.00	0.98
RPS7	6201	1.32	0.02	0.00	0.89
LOC285407	285407	1.32	-0.17	0.00	0.30
TRIM36	55521	1.32	0.01	0.00	0.97
C5orf27	202299	1.32	0.15	0.00	0.42
FAM53C	51307	1.32	-0.18	0.00	0.17
ACAP2	23527	1.32	-0.40	0.00	0.00
LOC653080	653080	1.32	-0.01	0.00	0.95
NEDD4	4734	1.32	-0.04	0.00	0.83
RBM7	10179	1.32	0.06	0.00	0.73
HIST1H2BK	85236	1.32	1.00	0.00	0.00
LOC728779	728779	1.31	-0.23	0.00	0.14
SF3B4	10262	1.31	0.02	0.00	0.89
SNORD95	619570	1.31	0.20	0.00	0.07
HIST2H2AC	8338	1.31	0.96	0.00	0.00
NCOA7	135112	1.31	-0.06	0.00	0.68
SNORD35A	26816	1.31	-0.12	0.00	0.45
HIST2H2AA4	723790	1.31	1.07	0.00	0.00

Table S4.1 continued

POLB	5423	1.31	-0.11	0.00	0.31
USPL1	10208	1.31	0.11	0.00	0.44
PPP2R5E	5529	1.31	-0.22	0.00	0.04
APOE	348	1.30	-0.22	0.00	0.34
YY1	7528	1.30	0.50	0.00	0.00
TNNC1	7134	1.30	-0.34	0.00	0.06
MCCC1	56922	1.30	-0.12	0.00	0.28
TAF13	6884	1.30	0.92	0.00	0.00
TCEAL6	158931	1.30	0.09	0.00	0.53
CCNT1	904	1.30	0.32	0.00	0.01
THOC5	8563	1.29	0.20	0.00	0.09
DCLK2	166614	1.29	0.00	0.00	1.00
HIST2H2AA3	8337	1.29	1.07	0.00	0.00
LOC732432	732432	1.29	0.36	0.00	0.00
LOC642414	642414	1.29	0.08	0.00	0.56
LOC100128086	100128086	1.29	-0.18	0.00	0.09
LOC650029	650029	1.29	-0.16	0.00	0.22
KLHL11	55175	1.28	0.48	0.00	0.00
CACYBP	27101	1.28	0.10	0.00	0.47
ZNF207	7756	1.28	-0.35	0.00	0.01
MYH3	4621	1.28	0.04	0.00	0.78
LOC100129186	100129186	1.28	-0.06	0.00	0.68
RAB8B	51762	1.27	-0.25	0.00	0.07
UTP6	55813	1.27	0.08	0.00	0.51
LOC727759	727759	1.27	-0.09	0.00	0.61
C10orf137	26098	1.27	-0.22	0.00	0.12
LHX3	8022	1.27	0.18	0.00	0.31
C17orf85	55421	1.27	-0.35	0.00	0.00
FAM90A6P	389618	1.27	-0.12	0.00	0.42
IVNS1ABP	10625	1.27	-0.47	0.00	0.00
KIAA0831	22863	1.27	-0.17	0.00	0.14
PLS1	5357	1.27	-0.17	0.00	0.18
RBM15	64783	1.27	0.29	0.00	0.02
ABCE1	6059	1.26	-0.27	0.00	0.02
TMEM126A	84233	1.26	-0.06	0.00	0.66
ITPR1	3708	1.26	-0.01	0.00	0.93
RLF	6018	1.26	0.29	0.00	0.01
UTP14A	10813	1.26	0.10	0.00	0.43
LGMN	5641	1.26	1.11	0.00	0.00
RBBP5	5929	1.26	0.09	0.00	0.50
CCDC109A	90550	1.25	-0.17	0.00	0.16
LOC644863	644863	1.25	0.16	0.00	0.19

Table S4.1 continued

EEF1B2	1933	1.25	-0.13	0.00	0.25
PTGR1	22949	1.25	-0.07	0.00	0.55
SLC4A7	9497	1.25	-0.74	0.00	0.00
LOC651441	651441	1.25	0.13	0.00	0.38
LOC653544	653544	1.25	8.13	0.00	0.00
WDSOF1	25879	1.25	0.34	0.00	0.01
LOC730740	730740	1.25	-0.51	0.00	0.00
MED30	90390	1.24	-0.05	0.00	0.73
FASTKD5	60493	1.24	-0.04	0.00	0.75
NFKBIB	4793	1.24	0.16	0.00	0.23
TDRD3	81550	1.24	0.16	0.00	0.32
EAF2	55840	1.24	0.01	0.00	0.97
RBM24	221662	1.24	-0.35	0.00	0.00
LOC653884	653884	1.24	0.51	0.00	0.00
ZNF644	84146	1.24	0.19	0.00	0.08
LOC146517	146517	1.23	0.13	0.00	0.24
LTV1	84946	1.23	0.17	0.00	0.14
SNRPB	6628	1.23	0.06	0.00	0.70
C17orf98	388381	1.23	-0.17	0.00	0.41
TUFT1	7286	1.23	-0.49	0.00	0.00
LOC728889	728889	1.23	-0.31	0.00	0.03
C10orf2	56652	1.23	0.21	0.00	0.09
TPMT	7172	1.23	-0.83	0.00	0.00
VGF	7425	1.23	0.14	0.00	0.26
LOC654121	654121	1.23	0.57	0.00	0.00
USP33	23032	1.23	0.36	0.00	0.01
NLF2	388125	1.23	-0.01	0.00	0.96
EIF1AX	1964	1.22	0.40	0.00	0.04
DNAJB9	4189	1.22	0.30	0.00	0.01
SCYL2	55681	1.22	0.47	0.00	0.00
TMED7	51014	1.22	-0.05	0.00	0.73
DSP	1832	1.22	0.43	0.00	0.00
TERF2IP	54386	1.22	-0.24	0.00	0.02
U2AF2	11338	1.22	0.02	0.00	0.91
U2AF1	7307	1.22	-0.04	0.00	0.78
SNORA63	6043	1.22	-0.07	0.00	0.65
HMGB2	3148	1.22	-0.53	0.00	0.01
DHX8	1659	1.22	0.21	0.00	0.07
DOCK10	55619	1.22	-0.04	0.00	0.75
LOC144438	144438	1.22	0.13	0.00	0.28
WBP11	51729	1.22	0.40	0.00	0.00
PUM1	9698	1.22	0.10	0.00	0.40

Table S4.1 continued

PSMD12	5718	1.21	0.80	0.00	0.00
CCDC148	130940	1.21	0.04	0.00	0.80
ZCCHC8	55596	1.21	0.14	0.00	0.24
DCP1A	55802	1.21	0.50	0.00	0.00
ARL5B	221079	1.21	-0.09	0.00	0.49
IMMP2L	83943	1.21	0.12	0.00	0.31
RFWD3	55159	1.21	-0.12	0.00	0.30
C20orf7	79133	1.21	-0.02	0.00	0.90
SCML2	10389	1.21	-0.07	0.00	0.62
FBXL20	84961	1.21	0.45	0.00	0.00
LOC653541	653541	1.21	7.88	0.00	0.00
DCP2	167227	1.20	-0.35	0.00	0.02
LOC388796	388796	1.20	0.04	0.00	0.82
LOC652051	652051	1.20	-0.07	0.00	0.70
CYB5R4	51167	1.20	0.35	0.00	0.02
NOMO3	408050	1.20	0.05	0.00	0.72
FRG2	448831	1.19	0.02	0.00	0.92
BCYRN1	618	1.19	-0.22	0.00	0.07
LOC728732	728732	1.19	-0.38	0.00	0.01
RPS24	6229	1.19	-0.02	0.00	0.87
LOC729101	729101	1.19	0.08	0.00	0.54
C16orf91	283951	1.19	0.22	0.00	0.04
FAM32A	26017	1.19	-0.36	0.00	0.00
ARGLU1	55082	1.19	-0.44	0.00	0.00
PDE12	201626	1.19	-0.25	0.00	0.03
NUDT1	4521	1.19	-0.61	0.00	0.00
ZNF197	10168	1.19	0.07	0.00	0.76
RWDD1	51389	1.19	0.21	0.00	0.10
GABPB2	2553	1.19	0.62	0.00	0.00
PPTC7	160760	1.19	0.38	0.00	0.00
C16orf33	79622	1.18	-0.29	0.00	0.00
EZH2	2146	1.18	-0.10	0.00	0.52
CLK3	1198	1.18	0.14	0.00	0.30
C6orf211	79624	1.18	-0.12	0.00	0.33
PPHLN1	51535	1.18	-0.09	0.00	0.55
CCT2	10576	1.18	-0.04	0.00	0.78
RYK	6259	1.18	0.03	0.00	0.82
RNF38	152006	1.18	-0.32	0.00	0.00
FOSB	2354	1.18	0.07	0.00	0.67
40605.00	115123	1.18	-0.15	0.00	0.21
F13A1	2162	1.18	-0.23	0.00	0.04
UTP23	84294	1.18	0.17	0.00	0.21

Table S4.1 continued

ADAT3	113179	1.18	0.16	0.00	0.13
IMPA1	3612	1.18	-0.08	0.00	0.65
RAE1	8480	1.18	0.09	0.00	0.50
FKBP14	55033	1.18	0.03	0.00	0.82
LOC653555	653555	1.18	-0.10	0.00	0.44
CHD2	1106	1.18	0.54	0.00	0.00
LBH	81606	1.18	0.15	0.00	0.30
LPAR2	9170	1.18	-0.20	0.00	0.18
UTX	7403	1.18	0.37	0.00	0.00
C21orf66	94104	1.17	0.23	0.00	0.04
PKN2	5586	1.17	0.38	0.00	0.01
NCRNA00120	55389	1.17	0.15	0.00	0.23
PAMR1	25891	1.17	-0.54	0.00	0.00
HNRNPF	3185	1.17	0.05	0.00	0.73
SESTD1	91404	1.17	0.56	0.00	0.00
KIAA1370	56204	1.17	0.22	0.00	0.27
DUX5	26581	1.17	8.35	0.00	0.00
H2AFX	3014	1.17	-0.41	0.00	0.00
LOC151162	151162	1.17	-0.52	0.00	0.00
FLNA	2316	1.17	-0.16	0.00	0.44
KLC2	64837	1.17	-0.05	0.00	0.73
CXorf40B	541578	1.17	-0.20	0.00	0.05
LOC649679	649679	1.16	-0.11	0.00	0.43
YTHDF3	253943	1.16	-0.54	0.00	0.00
WDR1	9948	1.16	0.31	0.00	0.00
GPR137C	283554	1.16	-0.15	0.00	0.22
ZNF280C	55609	1.16	-0.31	0.00	0.01
SNRPA1	6627	1.16	-0.10	0.00	0.36
LOC731314	731314	1.16	-0.59	0.00	0.00
WDR33	55339	1.16	0.03	0.00	0.87
KIAA0174	9798	1.16	0.09	0.00	0.46
CCNYL1	151195	1.16	0.32	0.00	0.01
ZSCAN5A	79149	1.16	0.10	0.00	0.53
RQCD1	9125	1.16	0.28	0.00	0.02
SYT14	255928	1.16	-0.05	0.00	0.79
ZFP37	7539	1.15	-0.16	0.00	0.16
LOC100132715	100132715	1.15	-0.60	0.00	0.00
ARPC5L	81873	1.15	0.19	0.00	0.09
EIF2C3	192669	1.15	0.09	0.00	0.59
LOC648218	648218	1.15	0.12	0.00	0.53
ERMP1	79956	1.15	-0.34	0.00	0.01
SLC25A26	115286	1.15	0.11	0.00	0.43

Table S4.1 continued

LOC100134229	100134229	1.15	-0.07	0.00	0.61
PDZD8	118987	1.15	0.69	0.00	0.00
CAMSAP1	157922	1.15	-0.29	0.00	0.02
GRPEL2	134266	1.15	-0.35	0.00	0.00
ATXN1L	342371	1.15	0.16	0.00	0.18
JARID1A	5927	1.15	0.66	0.00	0.00
HACL1	26061	1.15	0.17	0.00	0.08
LOC652864	652864	1.15	0.25	0.00	0.02
DNAJA2	10294	1.15	-0.33	0.00	0.00
NARG2	79664	1.15	0.18	0.00	0.22
MRPL4	51073	1.15	-0.02	0.00	0.90
LOC402112	402112	1.14	-0.31	0.00	0.05
RBM25	58517	1.14	-0.75	0.00	0.00
LOC440957	440957	1.14	-0.11	0.00	0.35
RBM4	5936	1.14	-0.13	0.00	0.26
DOCK7	85440	1.14	0.12	0.00	0.34
MTAP	4507	1.14	-0.42	0.00	0.00
SPRY2	10253	1.14	0.12	0.00	0.35
SGSH	6448	1.14	-0.08	0.00	0.58
PSME4	23198	1.14	-0.14	0.00	0.23
SNRPD3	6634	1.14	-0.40	0.00	0.00
CDC20	991	1.14	-1.06	0.00	0.00
TAPT1	202018	1.14	-0.30	0.00	0.01
IL34	146433	1.14	-0.15	0.00	0.18
KIF5C	3800	1.14	0.28	0.00	0.01
SNORD46	94161	1.14	0.13	0.00	0.35
LOC650681	650681	1.14	0.25	0.00	0.21
SDHD	6392	1.14	0.17	0.00	0.18
UTP3	57050	1.14	-0.02	0.00	0.94
SFRS13A	10772	1.13	0.35	0.00	0.00
RAPGEF6	51735	1.13	-0.20	0.00	0.14
BNIP2	663	1.13	0.06	0.00	0.69
LOC100129585	100129585	1.13	-0.34	0.00	0.01
ACTR6	64431	1.13	0.21	0.00	0.13
SNORD3D	780854	1.13	-0.15	0.00	0.35
PCDH7	5099	1.13	-0.12	0.00	0.34
ZFAND2A	90637	1.13	0.42	0.00	0.00
C18orf19	125228	1.13	0.58	0.00	0.00
CD55	1604	1.13	0.10	0.00	0.63
ORC6L	23594	1.12	-0.26	0.00	0.05
C9orf72	203228	1.12	-0.06	0.00	0.73
GART	2618	1.12	0.00	0.00	0.99

Table S4.1 continued

C2orf25	27249	1.12	0.20	0.00	0.08
DNAJC12	56521	1.12	-0.04	0.00	0.83
USP38	84640	1.12	-0.22	0.00	0.05
C12orf43	64897	1.12	0.09	0.00	0.52
KIAA1553	57673	1.12	-0.16	0.00	0.24
CCNE2	9134	1.12	-0.54	0.00	0.00
LOC440013	440013	1.12	-0.11	0.00	0.49
HSPC111	51491	1.12	-0.08	0.00	0.54
CROP	51747	1.12	-0.02	0.00	0.92
LOC650659	650659	1.11	-0.02	0.00	0.92
TOP1P2	7152	1.11	0.46	0.00	0.01
INA	9118	1.11	0.37	0.00	0.00
SNORD96A	619571	1.11	0.17	0.00	0.17
CTGF	1490	1.11	0.05	0.00	0.76
PELO	53918	1.11	0.45	0.00	0.00
FAM13B	51306	1.11	-0.40	0.00	0.00
SECISBP2L	9728	1.11	-0.22	0.00	0.08
PTRH2	51651	1.11	0.05	0.00	0.76
ZNF326	284695	1.11	0.09	0.00	0.60
MRPS22	56945	1.10	0.14	0.00	0.20
ETFA	2108	1.10	-0.11	0.00	0.30
UBE2C	11065	1.10	-0.66	0.00	0.00
CPEB4	80315	1.10	-0.05	0.00	0.78
LRIG1	26018	1.10	-0.39	0.00	0.00
MTERFD1	51001	1.10	0.20	0.00	0.06
RNF11	26994	1.10	0.80	0.00	0.00
API5	8539	1.10	-0.36	0.00	0.00
LOC100008589	100008589	1.10	0.01	0.00	0.93
TMEM41B	440026	1.10	-0.06	0.00	0.60
DISC1	27185	1.10	0.01	0.00	0.97
POLR2C	5432	1.10	0.14	0.00	0.22
MELK	9833	1.10	0.01	0.00	0.97
CSPP1	79848	1.10	-0.29	0.00	0.01
ZFAND6	54469	1.10	0.12	0.00	0.49
LOC645159	645159	1.10	0.34	0.00	0.02
NUP35	129401	1.10	0.24	0.00	0.02
C4orf32	132720	1.10	-0.14	0.00	0.25
TIPIN	54962	1.10	0.34	0.00	0.00
MTMR14	64419	1.10	-0.20	0.00	0.30
AHSA1	10598	1.10	0.11	0.00	0.40
FAM91A1	157769	1.10	1.25	0.00	0.00
MTX3	345778	1.10	-0.07	0.00	0.62

Table S4.1 continued

DYSF	8291	1.10	-0.70	0.00	0.00
SDCBP	6386	1.09	0.75	0.00	0.00
GOLGB1	2804	1.09	0.05	0.00	0.74
TTC14	151613	1.09	0.12	0.00	0.48
LOC651959	651959	1.09	-0.09	0.00	0.48
DNTT	1791	1.09	0.00	0.00	0.98
MATR3	9782	1.09	0.31	0.00	0.03
FAM108B1	51104	1.09	0.28	0.00	0.05
BMP4	652	1.09	0.16	0.00	0.18
RBP1	5947	1.09	-0.51	0.00	0.00
PDCL3	79031	1.09	0.02	0.00	0.92
CBLL1	79872	1.09	0.83	0.00	0.00
LOC100130856	100130856	1.09	0.07	0.00	0.58
ALKBH1	8846	1.09	0.17	0.00	0.11
LOC728643	728643	1.09	-0.63	0.00	0.00
FLRT3	23767	1.09	-0.51	0.00	0.00
PPP2CA	5515	1.09	-0.84	0.00	0.00
KITLG	4254	1.09	-0.49	0.00	0.00
LOC729608	729608	1.08	0.43	0.00	0.00
NHP2L1	4809	1.08	0.28	0.00	0.02
HIATL1	84641	1.08	0.13	0.00	0.26
NRAS	4893	1.08	0.51	0.00	0.00
LCOR	84458	1.08	0.15	0.00	0.29
STIM2	57620	1.08	0.14	0.00	0.31
C20orf4	25980	1.08	0.12	0.00	0.27
TNNT2	7139	1.08	-1.30	0.00	0.00
CDKN1A	1026	1.08	0.61	0.00	0.00
FOS	2353	1.08	0.47	0.00	0.01
TRAPPC6B	122553	1.08	0.62	0.00	0.00
DIP2B	57609	1.08	-0.22	0.00	0.12
LOC100128266	100128266	1.08	0.09	0.00	0.49
UBXN7	26043	1.08	0.30	0.00	0.01
LOC649137	649137	1.07	0.09	0.00	0.48
PLAGL2	5326	1.07	-0.11	0.00	0.36
ENC1	8507	1.07	-0.28	0.00	0.02
CS	1431	1.07	-0.32	0.00	0.00
TSC1	7248	1.07	0.13	0.00	0.32
SNHG12	85028	1.07	0.17	0.00	0.23
MAPRE3	22924	1.07	0.13	0.00	0.31
ZNF509	166793	1.07	0.04	0.00	0.78
NAF1	92345	1.07	0.15	0.00	0.23
BRPF3	27154	1.07	-0.04	0.00	0.75

Table S4.1 continued

HDAC2	3066	1.07	0.01	0.00	0.94
LOC652874	652874	1.07	0.01	0.00	0.98
ZMYM5	9205	1.07	0.07	0.00	0.62
GALM	130589	1.06	0.00	0.00	1.00
UPF2	26019	1.06	0.17	0.00	0.21
ZFP91	80829	1.06	0.53	0.00	0.00
NUP153	9972	1.06	-0.08	0.00	0.58
PRPF3	9129	1.06	0.12	0.00	0.29
LOC646786	646786	1.06	0.09	0.00	0.56
PRPF38A	84950	1.06	-0.17	0.00	0.22
LOC647081	647081	1.06	0.09	0.00	0.53
LOC643509	643509	1.06	0.00	0.00	0.99
RN7SK	125050	1.06	0.47	0.00	0.01
LOC641844	641844	1.06	0.03	0.00	0.81
LOC729120	729120	1.05	-0.02	0.00	0.91
KHSRP	8570	1.05	0.18	0.00	0.17
LOC645691	645691	1.05	-0.44	0.00	0.01
ZFYVE1	53349	1.05	0.13	0.00	0.26
SYAP1	94056	1.05	0.50	0.00	0.00
LOC644877	644877	1.05	0.09	0.00	0.47
LOC647037	647037	1.05	0.12	0.00	0.35
FAM103A1	83640	1.05	0.55	0.00	0.00
FLRT2	23768	1.05	-0.52	0.00	0.00
NUP155	9631	1.05	0.07	0.00	0.61
SLC25A38	54977	1.05	-0.02	0.00	0.92
SNRPB2	6629	1.05	-0.01	0.00	0.95
AZIN1	51582	1.05	0.28	0.00	0.04
LOC729920	729920	1.04	0.18	0.00	0.07
FAM176A	84141	1.04	0.68	0.00	0.00
BTBD7	55727	1.04	0.08	0.00	0.61
DIRC2	84925	1.04	0.34	0.00	0.00
LOC642268	642268	1.04	0.09	0.00	0.64
PRICKLE1	144165	1.04	-0.33	0.00	0.01
RSL1D1	26156	1.04	0.20	0.00	0.07
LOC647150	647150	1.04	0.40	0.00	0.00
HK2	3099	1.04	0.76	0.00	0.00
ZNF280B	140883	1.04	-0.34	0.00	0.00
LOC648210	648210	1.04	-0.54	0.00	0.00
RECQL	5965	1.04	0.68	0.00	0.00
C7orf40	285958	1.04	0.25	0.00	0.10
PABPC4L	132430	1.04	-0.28	0.00	0.02
RIF1	55183	1.04	0.26	0.00	0.04

Table S4.1 continued

MYC	4609	1.04	0.37	0.00	0.00
LOC652481	652481	1.04	0.11	0.00	0.45
ARMC5	79798	1.04	0.08	0.00	0.58
C4orf39	152756	1.04	-0.01	0.00	0.93
LRRC42	115353	1.04	-0.04	0.00	0.78
SLC25A25	114789	1.04	-0.43	0.00	0.00
CIRH1A	84916	1.03	0.16	0.00	0.13
NOMO1	23420	1.03	0.02	0.00	0.92
LOC649167	649167	1.03	0.00	0.00	1.00
UTP11L	51118	1.03	0.13	0.00	0.30
FAM126B	285172	1.03	0.21	0.00	0.05
OTUD6B	51633	1.03	-0.25	0.00	0.04
OCIAD2	132299	1.03	0.27	0.00	0.04
OVOL1	5017	1.03	-0.05	0.00	0.71
MAK10	60560	1.03	0.26	0.00	0.02
C12orf35	55196	1.03	0.14	0.00	0.33
TROVE2	6738	1.03	0.17	0.00	0.20
LOC648638	648638	1.03	0.22	0.00	0.04
LOC642812	642812	1.03	0.26	0.00	0.04
HNRNPAB	3182	1.03	-0.24	0.00	0.03
LOC643167	643167	1.03	0.23	0.00	0.03
HIST1H4H	8365	1.03	1.23	0.00	0.00
YTHDC1	91746	1.03	0.18	0.00	0.13
CD2AP	23607	1.03	-0.01	0.00	0.94
C16orf61	56942	1.02	-0.26	0.00	0.01
ZMAT2	153527	1.02	0.17	0.00	0.08
C12orf31	84298	1.02	-0.08	0.00	0.62
GTF2A2	2958	1.02	-0.20	0.00	0.09
LOC100132528	100132528	1.02	-0.74	0.00	0.00
TOMM40	10452	1.02	-0.08	0.00	0.50
YWHAE	7531	1.02	-0.09	0.00	0.52
SNAPC2	6618	1.02	-0.05	0.00	0.74
TRO	7216	1.02	-0.23	0.00	0.03
STIP1	10963	1.02	-0.12	0.00	0.26
GOLPH4	27333	1.02	0.72	0.00	0.00
URB2	9816	1.02	-0.29	0.00	0.00
TRK1	7206	1.02	0.07	0.00	0.67
ZDHHC14	79683	1.02	0.08	0.00	0.63
FAM90A18	441326	1.02	0.07	0.00	0.69
LOC100130550	100130550	1.02	0.53	0.00	0.00
LOC730153	730153	1.02	-0.01	0.00	0.98
HNRNPA1	3178	1.01	-0.42	0.00	0.00

Table S4.1 continued

COX10	1352	1.01	-0.05	0.00	0.69
LOC728022	728022	1.01	-0.20	0.00	0.18
CENPN	55839	1.01	0.00	0.00	0.99
CTDSPL2	51496	1.01	0.35	0.00	0.00
TRMT11	60487	1.01	-0.06	0.00	0.61
RPF2	84154	1.01	0.49	0.00	0.00
SLC25A36	55186	1.01	0.31	0.00	0.02
DPM1	8813	1.01	0.01	0.00	0.94
RNU6-15	100302741	1.01	0.71	0.00	0.00
LOC100133950	100133950	1.01	-0.14	0.00	0.39
U2AF1L2	8233	1.01	0.14	0.00	0.21
LOC651864	651864	1.01	0.01	0.00	0.95
SSR2	6746	1.01	0.01	0.00	0.96
SFT2D2	375035	1.01	0.37	0.00	0.00
BCL2L11	10018	1.01	0.16	0.00	0.27
ILF3	3609	1.00	-0.20	0.00	0.07
FAM90A15	389630	1.00	0.01	0.00	0.96
GADD45B	4616	1.00	0.35	0.00	0.00
PSPC1	55269	1.00	-0.24	0.00	0.11
IDI1	3422	1.00	-0.21	0.00	0.10
TAF9	6880	1.00	0.31	0.00	0.02
DUSP6	1848	1.00	0.27	0.00	0.03
RBM28	55131	1.00	0.37	0.00	0.00
ACTR5	79913	1.00	-0.18	0.00	0.14
STX11	8676	0.93	1.16	0.00	0.00
RORA	6095	0.89	1.89	0.00	0.00
YME1L1	10730	0.89	1.01	0.00	0.00
DNAJC3	5611	0.87	1.60	0.00	0.00
RDH5	5959	0.86	-1.20	0.00	0.00
DUX3	26582	0.83	7.83	0.00	0.00
MGC87042	256227	0.82	1.30	0.00	0.00
AKIRIN1	79647	0.82	1.03	0.00	0.00
ACTG2	72	0.61	-1.03	0.00	0.00
OGFRL1	79627	0.60	1.17	0.00	0.00
NRP1	8829	0.58	1.05	0.00	0.00
CDKN1C	1028	0.57	-1.05	0.00	0.00
STEAP1	26872	0.55	1.03	0.00	0.00
UBE2N	7334	0.52	-1.08	0.00	0.00
ACO1	48	0.52	-1.16	0.00	0.00
CLDN5	7122	0.50	-1.37	0.00	0.00
CGGBP1	8545	0.46	1.05	0.00	0.00
LOC441455	441455	0.44	1.00	0.00	0.00

Table S4.1 continued

LOC651861	651861	0.39	6.60	0.00	0.00
MCL1	4170	0.38	1.67	0.00	0.00
HNRNPA0	10949	0.35	-1.25	0.00	0.00
LOC203547	203547	0.34	-1.09	0.00	0.00
AK3L1	205	0.33	1.03	0.00	0.00
ABL2	27	0.32	1.07	0.00	0.00
HN1	51155	0.31	-1.04	0.00	0.00
SFRS6	6431	0.31	-1.01	0.00	0.00
HIST1H2AC	8334	0.26	1.07	0.01	0.00
SOX8	30812	0.17	-1.26	0.09	0.00
ARFGEF2	10564	0.13	1.34	0.31	0.00
TIMM23B	653252	0.12	1.88	0.46	0.00
LOC100133997	100133997	0.11	2.78	0.31	0.00
CCL20	6364	0.11	1.63	0.38	0.00
ELK4	2005	0.11	1.31	0.40	0.00
WDR36	134430	0.10	-1.07	0.31	0.00
SPCS3	60559	0.10	1.04	0.32	0.00
ABHD2	11057	0.09	1.05	0.49	0.00
ALOX5AP	241	0.09	-1.03	0.45	0.00
FGFR4	2264	0.08	-1.05	0.50	0.00
PAFAH1B2	5049	0.06	1.07	0.58	0.00
SHD	56961	0.06	-1.12	0.58	0.00
AIM2	9447	0.04	1.07	0.83	0.00
CD274	29126	0.03	1.36	0.80	0.00
FBXO6	26270	0.03	1.02	0.80	0.00
DUX2	26583	0.03	1.03	0.78	0.00
RET	5979	0.02	1.11	0.84	0.00
CUL4B	8450	0.02	1.10	0.86	0.00
C10orf6	55719	0.00	-1.17	1.00	0.00
LOC652641	652641	-0.01	1.08	0.96	0.00
UBE2L3	7332	-0.01	-1.09	0.90	0.00
RALA	5898	-0.04	-1.21	0.73	0.00
OLFML2B	25903	-0.05	-1.28	0.72	0.00
SFRS2IP	9169	-0.06	-1.07	0.61	0.00
STK17B	9262	-0.06	1.31	0.63	0.00
CYP2J2	1573	-0.07	1.13	0.60	0.00
UBR1	197131	-0.09	1.11	0.43	0.00
USP13	8975	-0.09	-1.04	0.51	0.00
NES	10763	-0.09	-1.12	0.38	0.00
SERPINB8	5271	-0.10	1.02	0.37	0.00
PANX1	24145	-0.11	1.23	0.31	0.00
PPM1K	152926	-0.15	1.44	0.09	0.00

Table S4.1 continued

HSPA13	6782	-0.18	1.34	0.11	0.00
LOC649425	649425	-0.21	2.54	0.07	0.00
UHRF1	29128	-0.21	-1.03	0.08	0.00
RUNX1	861	-0.21	-1.05	0.10	0.00
LOC730996	730996	-0.22	1.05	0.02	0.00
STRN	6801	-0.23	1.09	0.08	0.00
AASDHPPT	60496	-0.24	-1.25	0.09	0.00
SERBP1	26135	-0.24	-1.18	0.02	0.00
SLC25A24	29957	-0.28	1.10	0.02	0.00
HIPK3	10114	-0.30	1.55	0.01	0.00
COL1A1	1277	-0.30	-1.26	0.01	0.00
ZBP1	81030	-0.33	1.19	0.01	0.00
SLC30A1	7779	-0.35	1.24	0.01	0.00
HBB	3043	-0.35	1.20	0.00	0.00
REEP5	7905	-0.36	-1.14	0.00	0.00
PAFAH1B1	5048	-0.36	-1.24	0.00	0.00
GPD2	2820	-0.37	1.20	0.00	0.00
ALPK2	115701	-0.38	-1.08	0.00	0.00
IFNB1	3456	-0.39	1.04	0.01	0.00
GPM6B	2824	-0.40	-1.12	0.00	0.00
C1orf58	148362	-0.41	1.05	0.00	0.00
CENPF	1063	-0.41	-1.03	0.00	0.00
RBMS2P	643427	-0.45	1.31	0.00	0.00
GMPR	2766	-0.45	1.36	0.00	0.00
SLC8A1	6546	-0.46	1.05	0.00	0.00
SCAMP1	9522	-0.47	1.37	0.00	0.00
LAMC1	3915	-0.51	-1.06	0.00	0.00
SLC29A1	2030	-0.51	-1.03	0.00	0.00
RIOK3	8780	-0.52	1.13	0.00	0.00
C3orf38	285237	-0.55	1.03	0.00	0.00
FGD4	121512	-0.56	1.21	0.00	0.00
FEM1B	10116	-0.57	1.03	0.00	0.00
APOOL	139322	-0.58	1.37	0.00	0.00
NAMPT	10135	-0.58	1.21	0.00	0.00
MDM2	4193	-0.59	1.08	0.00	0.00
LITAF	9516	-0.59	1.12	0.00	0.00
AHI1	54806	-0.59	-1.05	0.00	0.00
RCAN2	10231	-0.60	-1.00	0.00	0.00
RBMS1	5937	-0.62	1.10	0.00	0.00
PDGFRL	5157	-0.62	1.20	0.00	0.00
MBTPS2	51360	-0.65	1.03	0.00	0.00
NT5C3	51251	-0.66	1.88	0.00	0.00

Table S4.1 continued

DCBLD1	285761	-0.67	1.59	0.00	0.00
OSMR	9180	-0.74	1.12	0.00	0.00
SPPL2A	84888	-0.75	1.12	0.00	0.00
IL4I1	259307	-0.76	1.77	0.00	0.00
LOC400759	400759	-0.76	1.03	0.00	0.00
RBMS2	5939	-0.77	1.45	0.00	0.00
FAM62B	57488	-0.79	-1.39	0.00	0.00
IL6	3569	-0.80	1.20	0.00	0.00
ZC3HAV1	56829	-0.82	1.47	0.00	0.00
ARSK	153642	-0.84	1.03	0.00	0.00
LGALS9	3965	-0.85	1.71	0.00	0.00
C15orf48	84419	-0.87	1.65	0.00	0.00
CXCL10	3627	-0.88	3.23	0.00	0.00
CMPK2	129607	-0.88	1.35	0.00	0.00
UGCG	7357	-0.89	1.45	0.00	0.00
GBP5	115362	-0.91	1.31	0.00	0.00
EEA1	8411	-0.96	1.03	0.00	0.00
CLIC4	25932	-1.00	1.23	0.00	0.00
TRIM78P	117852	-1.00	1.70	0.00	0.00
LEPR	3953	-1.00	-0.55	0.00	0.00
C3	718	-1.00	0.25	0.00	0.10
ANXA1	301	-1.00	0.09	0.00	0.47
STARD13	90627	-1.00	-0.90	0.00	0.00
TRIM69	140691	-1.00	0.26	0.00	0.04
TAGLN2	8407	-1.00	0.05	0.00	0.72
TIGD5	84948	-1.00	-0.22	0.00	0.09
MZF1	7593	-1.00	0.06	0.00	0.76
LOC338758	338758	-1.00	-0.07	0.00	0.59
HOXA5	3202	-1.00	-0.18	0.00	0.14
DIS3L	115752	-1.01	-0.27	0.00	0.03
CKAP4	10970	-1.01	0.04	0.00	0.80
ATRIP	84126	-1.01	0.00	0.00	0.99
EBPL	84650	-1.01	-0.09	0.00	0.43
REC8	9985	-1.01	-0.24	0.00	0.08
PHYH	5264	-1.01	-0.58	0.00	0.00
TMEM9	252839	-1.01	-0.21	0.00	0.07
CLSTN1	22883	-1.01	-0.28	0.00	0.01
HSCB	150274	-1.01	0.17	0.00	0.11
RRM2B	50484	-1.01	0.62	0.00	0.00
ACVRL1	94	-1.01	0.06	0.00	0.66
SDSL	113675	-1.01	-0.09	0.00	0.64
NFKB2	4791	-1.01	0.22	0.00	0.03

Table S4.1 continued

ALDH2	217	-1.01	0.21	0.00	0.10
ANTXR1	84168	-1.01	0.36	0.00	0.01
IL15	3600	-1.01	-0.10	0.00	0.39
TTC15	51112	-1.01	-0.12	0.00	0.30
ILVBL	10994	-1.02	-0.28	0.00	0.01
PPM1M	132160	-1.02	-0.24	0.00	0.03
ACSL1	2180	-1.02	0.15	0.00	0.37
C15orf52	388115	-1.02	-0.30	0.00	0.02
ZNF650	130507	-1.02	0.69	0.00	0.00
ATG7	10533	-1.02	0.21	0.00	0.16
TGFBR2	7048	-1.02	0.17	0.00	0.17
VAMP4	8674	-1.02	0.47	0.00	0.00
BEXL1	56271	-1.02	-0.15	0.00	0.15
RUNX2	860	-1.02	0.06	0.00	0.68
VAMP8	8673	-1.02	0.00	0.00	0.99
LOC653583	653583	-1.02	0.10	0.00	0.44
TERF2	7014	-1.02	0.33	0.00	0.01
PTPRA	5786	-1.02	0.04	0.00	0.77
TRPV2	51393	-1.02	0.31	0.00	0.01
CXorf12	8269	-1.02	0.03	0.00	0.85
RNF146	81847	-1.02	0.15	0.00	0.35
GAMT	2593	-1.02	-0.35	0.00	0.01
GLT8D1	55830	-1.02	-0.27	0.00	0.06
WIPI1	55062	-1.02	0.08	0.00	0.53
NTN4	59277	-1.02	0.04	0.00	0.86
ZMYM3	9203	-1.02	-0.01	0.00	0.95
FAM3A	60343	-1.02	-0.18	0.00	0.15
RBM23	55147	-1.02	-0.08	0.00	0.55
SMARCD3	6604	-1.02	-0.34	0.00	0.00
C1orf131	128061	-1.02	-0.09	0.00	0.54
RNF213	57674	-1.02	0.63	0.00	0.00
DNAJB4	11080	-1.02	0.41	0.00	0.01
ULBP1	80329	-1.02	0.62	0.00	0.00
IDS	3423	-1.02	0.67	0.00	0.00
SLC16A5	9121	-1.03	-0.04	0.00	0.83
TINF2	26277	-1.03	0.39	0.00	0.00
GAS6	2621	-1.03	-0.32	0.00	0.00
SLC43A3	29015	-1.03	0.44	0.00	0.00
FAM156A	29057	-1.03	-0.14	0.00	0.28
ASAP2	8853	-1.03	-0.34	0.00	0.00
GJD3	125111	-1.03	0.86	0.00	0.00
C2orf28	51374	-1.03	-0.08	0.00	0.53

Table S4.1 continued

EFEMP1	2202	-1.03	-0.21	0.00	0.08
GSTO1	9446	-1.03	0.31	0.00	0.00
D2HGDH	728294	-1.03	-0.40	0.00	0.01
PTEN	5728	-1.03	-0.30	0.00	0.01
RAB13	5872	-1.03	0.35	0.00	0.02
APBA3	9546	-1.03	0.00	0.00	0.99
MGST1	4257	-1.03	0.23	0.00	0.04
MGLL	11343	-1.03	0.00	0.00	0.99
MOSC1	64757	-1.03	-0.11	0.00	0.58
ZNF672	79894	-1.03	-0.15	0.00	0.28
CDC42EP4	23580	-1.03	0.07	0.00	0.61
NUDT18	79873	-1.03	0.17	0.00	0.17
MXD4	10608	-1.04	-0.92	0.00	0.00
TMEM189-UBE2V1	387522	-1.04	0.06	0.00	0.69
SLC9A3R1	9368	-1.04	-0.06	0.00	0.63
DENND5A	23258	-1.04	0.08	0.00	0.48
HOXB7	3217	-1.04	-0.06	0.00	0.64
PACSIN2	11252	-1.04	-0.33	0.00	0.00
ACCS	84680	-1.04	-0.16	0.00	0.20
XPO1	7514	-1.04	-0.23	0.00	0.06
C20orf72	92667	-1.04	-0.05	0.00	0.70
ITGA3	3675	-1.04	0.13	0.00	0.29
IRF2BP2	359948	-1.04	-0.81	0.00	0.00
SRC	6714	-1.04	0.06	0.00	0.67
LOC285296	285296	-1.04	0.82	0.00	0.00
MYF6	4618	-1.04	-0.02	0.00	0.89
ZFP36	7538	-1.04	0.23	0.00	0.04
C4orf34	201895	-1.04	0.61	0.00	0.00
FAM57A	79850	-1.04	0.00	0.00	0.99
LSS	4047	-1.04	-0.01	0.00	0.97
MME	4311	-1.05	-0.18	0.00	0.11
SULF2	55959	-1.05	-0.05	0.00	0.72
ATP6AP1L	92270	-1.05	0.03	0.00	0.86
RCOR2	283248	-1.05	-0.19	0.00	0.17
FAM160B1	57700	-1.05	0.63	0.00	0.00
RFX5	5993	-1.05	-0.06	0.00	0.63
FAM65A	79567	-1.05	-0.13	0.00	0.32
TRIM55	84675	-1.05	0.20	0.00	0.07
MYPOP	339344	-1.05	-0.04	0.00	0.80
TCFL5	10732	-1.05	0.03	0.00	0.79
GPR1	2825	-1.05	0.22	0.00	0.05

Table S4.1 continued

IGDCC4	57722	-1.05	-0.52	0.00	0.00
ARHGEF2	9181	-1.06	-0.07	0.00	0.59
CCND2	894	-1.06	-0.53	0.00	0.00
EEF2K	29904	-1.06	-0.53	0.00	0.00
CASP4	837	-1.06	0.69	0.00	0.00
RNASE4	6038	-1.06	-0.11	0.00	0.41
AUH	549	-1.06	0.16	0.00	0.15
TDRD7	23424	-1.06	1.07	0.00	0.00
RCN3	57333	-1.06	-0.20	0.00	0.06
JUNB	3726	-1.06	0.52	0.00	0.00
HAS1	3036	-1.06	0.22	0.00	0.16
C14orf93	60686	-1.06	-0.11	0.00	0.37
PPP2R2C	5522	-1.06	-0.23	0.00	0.08
ANGPTL4	51129	-1.06	1.27	0.00	0.00
FXYD5	53827	-1.07	0.09	0.00	0.44
HOM-TES-103	25900	-1.07	-0.32	0.00	0.01
FLT3LG	2323	-1.07	0.39	0.00	0.00
GBA	2629	-1.07	0.29	0.00	0.01
ETS1	2113	-1.07	-0.18	0.00	0.27
THY1	7070	-1.07	-0.64	0.00	0.00
AARS	16	-1.07	0.28	0.00	0.02
NDST1	3340	-1.07	-0.69	0.00	0.00
ALOX15B	247	-1.07	0.45	0.00	0.00
TAX1BP3	30851	-1.07	-0.34	0.00	0.00
RILPL1	353116	-1.07	0.62	0.00	0.00
FEZ1	9638	-1.07	0.19	0.00	0.10
KTELC1	56983	-1.07	0.16	0.00	0.14
RHOT1	55288	-1.07	0.33	0.00	0.00
MAP3K6	9064	-1.07	-0.23	0.00	0.05
TBC1D2	55357	-1.07	0.29	0.00	0.03
FAM189B	10712	-1.08	0.04	0.00	0.80
NUP37	79023	-1.08	-0.13	0.00	0.22
PSAT1	29968	-1.08	0.74	0.00	0.00
PLEKHA2	59339	-1.08	0.06	0.00	0.65
BID	637	-1.08	-0.17	0.00	0.12
HS1BP3	64342	-1.08	0.05	0.00	0.71
ARSD	414	-1.08	-0.21	0.00	0.09
GSTK1	373156	-1.08	0.06	0.00	0.62
ACPL2	92370	-1.08	0.15	0.00	0.23
RNF31	55072	-1.08	0.05	0.00	0.74
TBK1	29110	-1.08	0.36	0.00	0.00
NOTCH1	4851	-1.08	-0.67	0.00	0.00

Table S4.1 continued

C20orf117	140710	-1.08	-0.53	0.00	0.00
HLA-A	3105	-1.08	0.56	0.00	0.00
CYFIP2	26999	-1.08	-0.63	0.00	0.00
ATP2B1	490	-1.08	-0.70	0.00	0.00
ZKSCAN1	7586	-1.08	0.38	0.00	0.02
C8orf55	51337	-1.08	-0.39	0.00	0.00
CLIC1	1192	-1.08	0.31	0.00	0.00
ARHGAP22	58504	-1.08	0.08	0.00	0.55
NADSYN1	55191	-1.09	0.12	0.00	0.32
DRAM1	55332	-1.09	-0.36	0.00	0.00
SHMT2	6472	-1.09	0.26	0.00	0.03
PTGFR	5737	-1.09	0.15	0.00	0.26
CXCL16	58191	-1.09	1.18	0.00	0.00
CD68	968	-1.09	1.18	0.00	0.00
NRSN2	80023	-1.09	0.07	0.00	0.55
LOC730278	730278	-1.09	0.19	0.00	0.18
THBS1	7057	-1.09	-0.57	0.00	0.00
NFE2L2	4780	-1.09	0.33	0.00	0.00
LOC728060	728060	-1.09	0.26	0.00	0.04
AP1S1	1174	-1.10	-0.14	0.00	0.35
CAV1	857	-1.10	-0.19	0.00	0.10
BEX1	55859	-1.10	-0.23	0.00	0.04
PRKD1	5587	-1.10	-0.61	0.00	0.00
WEE1	7465	-1.10	0.33	0.00	0.01
PRKAG2	51422	-1.10	0.75	0.00	0.00
HCP5	10866	-1.10	0.14	0.00	0.21
SLC24A6	80024	-1.10	0.00	0.00	1.00
C2orf32	25927	-1.10	0.08	0.00	0.59
DPYSL2	1808	-1.10	0.23	0.00	0.04
ETV6	2120	-1.10	-0.06	0.00	0.72
ODF3B	440836	-1.10	0.43	0.00	0.00
PEX16	9409	-1.10	0.02	0.00	0.90
SH3PXD2B	285590	-1.10	-0.02	0.00	0.92
BEST1	7439	-1.10	0.27	0.00	0.01
LRP10	26020	-1.10	0.42	0.00	0.00
PMP22	5376	-1.10	0.27	0.00	0.03
RIPK1	8737	-1.10	-0.24	0.00	0.09
WNT5A	7474	-1.10	-0.40	0.00	0.00
LOC390530	390530	-1.10	0.11	0.00	0.51
TXNRD2	10587	-1.10	-0.17	0.00	0.20
ENG	2022	-1.10	0.49	0.00	0.00
MOBKL2C	148932	-1.11	0.13	0.00	0.41

Table S4.1 continued

PPAP2A	8611	-1.11	0.54	0.00	0.00
PFKP	5214	-1.11	0.78	0.00	0.00
HOXC13	3229	-1.11	0.44	0.00	0.00
IL10RB	3588	-1.11	-0.05	0.00	0.69
APOBEC3C	27350	-1.11	0.58	0.00	0.00
FES	2242	-1.11	-0.33	0.00	0.02
P4HB	5034	-1.11	-0.19	0.00	0.10
HLA-E	3133	-1.11	0.81	0.00	0.00
BCL9L	283149	-1.11	0.73	0.00	0.00
ZBTB25	7597	-1.11	0.35	0.00	0.00
FAM156B	727866	-1.11	-0.17	0.00	0.14
SPG7	6687	-1.11	-0.28	0.00	0.03
MYOD1	4654	-1.11	-0.46	0.00	0.00
NPEPL1	79716	-1.11	-0.26	0.00	0.04
OPTN	10133	-1.11	-0.57	0.00	0.00
TRIM5	85363	-1.11	0.50	0.00	0.00
GRAMD3	65983	-1.12	0.38	0.00	0.00
RFNG	5986	-1.12	-0.22	0.00	0.04
PDE7B	27115	-1.12	-0.57	0.00	0.00
ARMCX1	51309	-1.12	-0.06	0.00	0.70
FRMD3	257019	-1.12	-0.29	0.00	0.00
C1orf57	84284	-1.12	0.35	0.00	0.00
PRMT2	3275	-1.12	0.07	0.00	0.60
ZNF319	57567	-1.12	-0.08	0.00	0.51
SLC2A10	81031	-1.12	-0.32	0.00	0.01
CDR2L	30850	-1.12	-0.20	0.00	0.04
CORO6	84940	-1.12	-0.30	0.00	0.03
LBA1	9881	-1.12	0.50	0.00	0.00
CERCAM	51148	-1.12	-0.18	0.00	0.15
CTDSP1	58190	-1.12	0.04	0.00	0.78
PCBP3	54039	-1.12	0.21	0.00	0.12
TIAF1	9220	-1.12	-0.53	0.00	0.00
LPXN	9404	-1.12	-0.02	0.00	0.91
TYMS	7298	-1.13	-0.09	0.00	0.44
FKBP11	51303	-1.13	0.30	0.00	0.01
PSRC1	84722	-1.13	-0.49	0.00	0.00
FAM129B	64855	-1.13	0.44	0.00	0.00
LUM	4060	-1.13	0.39	0.00	0.00
LYRM1	57149	-1.13	0.01	0.00	0.91
TNFRSF10A	8797	-1.13	0.84	0.00	0.00
ACP2	53	-1.13	0.22	0.00	0.09
SYNC1	81493	-1.13	-0.25	0.00	0.11

Table S4.1 continued

USP41	373856	-1.13	0.52	0.00	0.00
TNFSF10	8743	-1.13	2.97	0.00	0.00
LOC645638	645638	-1.13	-0.07	0.00	0.58
PID1	55022	-1.13	0.31	0.00	0.01
TOM1	10043	-1.13	0.00	0.00	1.00
GBE1	2632	-1.13	0.48	0.00	0.00
PNPO	55163	-1.13	-0.03	0.00	0.87
LHFPL2	10184	-1.13	-0.14	0.00	0.39
CST3	1471	-1.13	0.16	0.00	0.16
SLC44A1	23446	-1.13	0.78	0.00	0.00
SLC35E3	55508	-1.13	0.46	0.00	0.00
RXRA	6256	-1.13	-0.32	0.00	0.00
CREB1	1385	-1.13	-0.09	0.00	0.46
FAS	355	-1.13	0.15	0.00	0.26
SLC22A4	6583	-1.13	0.38	0.00	0.01
NOD2	64127	-1.13	0.43	0.00	0.00
ASB5	140458	-1.14	-0.46	0.00	0.00
IFI27L2	83982	-1.14	-0.03	0.00	0.88
CDC42EP5	148170	-1.14	0.14	0.00	0.37
TP53INP1	94241	-1.14	-0.57	0.00	0.00
PSME1	5720	-1.14	0.28	0.00	0.01
FAT1	2195	-1.14	0.41	0.00	0.00
CRTC3	64784	-1.14	-0.12	0.00	0.33
LOC650215	650215	-1.14	0.64	0.00	0.00
EMP3	2014	-1.14	-0.21	0.00	0.09
TRIM56	81844	-1.14	0.31	0.00	0.01
LOC728809	728809	-1.14	0.02	0.00	0.90
C11orf68	83638	-1.14	0.36	0.00	0.00
PEX11B	8799	-1.14	-0.22	0.00	0.08
PDE4B	5142	-1.14	0.68	0.00	0.00
LOC653506	653506	-1.14	0.02	0.00	0.91
CBLN3	643866	-1.14	0.15	0.00	0.24
OSBPL7	114881	-1.14	-0.37	0.00	0.00
HIF1A	3091	-1.15	0.42	0.00	0.04
MFGE8	4240	-1.15	0.03	0.00	0.83
LASP1	3927	-1.15	-0.46	0.00	0.00
ATP6V0A1	535	-1.15	-0.04	0.00	0.76
KCNK2	3776	-1.15	0.21	0.00	0.05
ZC3H5	85451	-1.15	0.10	0.00	0.51
LTBR	4055	-1.15	0.41	0.00	0.00
TXLNA	200081	-1.16	-0.16	0.00	0.20
SMARCAL1	50485	-1.16	0.02	0.00	0.92

Table S4.1 continued

TOP2A	7153	-1.16	-0.61	0.00	0.00
IRS1	3667	-1.16	0.28	0.00	0.01
SLC44A2	57153	-1.16	-0.35	0.00	0.00
CYP4V2	285440	-1.16	-0.50	0.00	0.00
MTSS1	9788	-1.16	-0.41	0.00	0.01
LOC100129165	100129165	-1.16	0.22	0.00	0.04
PARD6G	84552	-1.16	-0.01	0.00	0.96
KDELR3	11015	-1.16	-0.44	0.00	0.00
ADAMTS1	9510	-1.16	0.18	0.00	0.23
ZNF302	55900	-1.16	-0.01	0.00	0.96
PSPH	5723	-1.16	0.01	0.00	0.97
FRMD4A	55691	-1.16	-0.55	0.00	0.00
LOC644739	644739	-1.17	-0.07	0.00	0.58
ZMIZ1	57178	-1.17	-0.04	0.00	0.80
NPR2	4882	-1.17	-0.05	0.00	0.78
ID1	3397	-1.17	0.73	0.00	0.00
NLRX1	79671	-1.17	-0.14	0.00	0.25
C14orf173	64423	-1.17	-0.27	0.00	0.01
HECW2	57520	-1.17	0.06	0.00	0.65
RAB40C	57799	-1.17	-0.07	0.00	0.68
UGP2	7360	-1.17	0.11	0.00	0.40
CASP1	834	-1.17	2.17	0.00	0.00
CPSF4	10898	-1.17	-0.27	0.00	0.01
HIPK2	28996	-1.17	0.36	0.00	0.00
MTHFD1L	25902	-1.17	0.40	0.00	0.01
ABLIM3	22885	-1.17	0.01	0.00	0.93
CDH11	1009	-1.17	0.24	0.00	0.06
TSPAN4	7106	-1.17	-0.12	0.00	0.36
JSRP1	126306	-1.18	-0.47	0.00	0.00
FBXO32	114907	-1.18	0.77	0.00	0.00
NCOR2	9612	-1.18	-0.22	0.00	0.09
MVP	9961	-1.18	-0.12	0.00	0.45
RAB7L1	8934	-1.18	0.43	0.00	0.00
PLOD1	5351	-1.18	-0.15	0.00	0.22
PDE4C	5143	-1.18	0.40	0.00	0.00
HEG1	57493	-1.18	-0.25	0.00	0.02
MT1F	4494	-1.18	-0.54	0.00	0.00
CREB3L2	64764	-1.18	-0.25	0.00	0.03
CAPN5	726	-1.18	-0.54	0.00	0.00
ABCB6	10058	-1.18	-0.27	0.00	0.02
SYT7	9066	-1.18	-0.65	0.00	0.00
PTGER2	5732	-1.18	-0.20	0.00	0.07

Table S4.1 continued

BCAR3	8412	-1.18	-0.15	0.00	0.22
SMYD4	114826	-1.19	0.21	0.00	0.08
TRIM8	81603	-1.19	-0.35	0.00	0.00
ZMYM6	9204	-1.19	-0.01	0.00	0.96
PLEKHF1	79156	-1.19	0.20	0.00	0.21
TMED10P	286102	-1.19	0.45	0.00	0.00
FAM174B	400451	-1.19	-0.57	0.00	0.00
C1S	716	-1.19	0.42	0.00	0.00
LOC728855	728855	-1.19	0.42	0.00	0.00
NEXN	91624	-1.19	-0.26	0.00	0.03
STC2	8614	-1.19	-0.36	0.00	0.01
AMZ2	51321	-1.19	0.02	0.00	0.92
CAV2	858	-1.19	0.43	0.00	0.00
SLC9A1	6548	-1.19	0.20	0.00	0.10
FYN	2534	-1.19	0.31	0.00	0.03
POLR3GL	84265	-1.19	-0.10	0.00	0.41
LOC374395	374395	-1.20	0.36	0.00	0.00
IKBKE	9641	-1.20	0.39	0.00	0.00
TMEM62	80021	-1.20	1.22	0.00	0.00
C1orf66	51093	-1.20	-0.08	0.00	0.52
SPG21	51324	-1.20	-0.01	0.00	0.97
C7orf10	79783	-1.20	-0.32	0.00	0.00
CFD	1675	-1.20	0.25	0.00	0.04
SAMD4A	23034	-1.20	1.12	0.00	0.00
APOBEC3F	200316	-1.20	0.60	0.00	0.00
SIDT2	51092	-1.21	-0.30	0.00	0.00
PHF21A	51317	-1.21	0.06	0.00	0.64
RPS6KA2	6196	-1.21	-0.37	0.00	0.00
RRAS2	22800	-1.21	0.41	0.00	0.00
ABCA1	19	-1.21	-0.32	0.00	0.01
ASPSCR1	79058	-1.21	-0.31	0.00	0.01
SLC39A8	64116	-1.22	0.90	0.00	0.00
BLOC1S2	282991	-1.22	0.45	0.00	0.00
C7orf68	29923	-1.22	0.32	0.00	0.01
IFNGR2	3460	-1.22	0.02	0.00	0.88
SAA2	6289	-1.22	0.28	0.00	0.11
LOC642567	642567	-1.22	0.56	0.00	0.01
SPG11	80208	-1.22	0.30	0.00	0.01
ALDH3B1	221	-1.22	-0.15	0.00	0.27
KIAA0240	23506	-1.23	-0.25	0.00	0.02
BCL3	602	-1.23	-0.04	0.00	0.76
ARHGAP23	57636	-1.23	0.00	0.00	1.00

Table S4.1 continued

EML3	256364	-1.23	-0.15	0.00	0.17
ZNF37A	7587	-1.23	0.21	0.00	0.09
ANGPT1	284	-1.23	0.22	0.00	0.09
C5orf62	85027	-1.23	-0.06	0.00	0.69
TCEAL4	79921	-1.24	-0.34	0.00	0.00
CHES1	1112	-1.24	-0.43	0.00	0.00
TWIST1	7291	-1.24	0.47	0.00	0.00
ATP2B4	493	-1.24	-0.42	0.00	0.00
EDN1	1906	-1.24	0.55	0.00	0.00
NFE2L3	9603	-1.24	1.28	0.00	0.00
RIN2	54453	-1.24	-0.13	0.00	0.34
BAX	581	-1.24	-0.09	0.00	0.57
UBL4A	8266	-1.24	-0.05	0.00	0.69
TMEM42	131616	-1.25	-0.41	0.00	0.00
TAP2	6891	-1.25	1.35	0.00	0.00
IL18BP	10068	-1.25	1.40	0.00	0.00
KIAA1751	85452	-1.25	0.72	0.00	0.00
PDGFRB	5159	-1.25	-0.58	0.00	0.00
ZNF362	149076	-1.25	-0.42	0.00	0.00
FTSJD2	23070	-1.25	0.46	0.00	0.00
CNN2	1265	-1.25	-0.63	0.00	0.00
KAT2B	8850	-1.25	-0.25	0.00	0.03
PARP8	79668	-1.25	0.57	0.00	0.00
DAAM2	23500	-1.26	-0.06	0.00	0.72
MAP4K2	5871	-1.26	-0.25	0.00	0.02
MTMR11	10903	-1.26	0.45	0.00	0.00
BCL6	604	-1.26	0.36	0.00	0.00
AK3	50808	-1.26	-0.43	0.00	0.00
TTC39B	158219	-1.26	1.23	0.00	0.00
KLF9	687	-1.26	-0.05	0.00	0.67
COPS8	10920	-1.26	0.14	0.00	0.31
ASAM	79827	-1.26	0.02	0.00	0.89
INHBE	83729	-1.26	-0.49	0.00	0.00
RFTN1	23180	-1.26	-0.40	0.00	0.00
GPX1	2876	-1.26	0.32	0.00	0.01
ANKRA2	57763	-1.26	0.34	0.00	0.00
CLDN23	137075	-1.26	1.18	0.00	0.00
KAT2A	2648	-1.26	0.06	0.00	0.62
SLC3A2	6520	-1.27	0.65	0.00	0.00
MT1G	4495	-1.27	0.25	0.00	0.13
LRDD	55367	-1.27	0.01	0.00	0.98
CCDC92	80212	-1.27	-0.31	0.00	0.01

Table S4.1 continued

MMP7	4316	-1.27	0.21	0.00	0.10
HOXA10	3206	-1.27	-0.07	0.00	0.57
LOC100129034	100129034	-1.27	-0.34	0.00	0.01
TNFAIP8	25816	-1.27	0.79	0.00	0.00
NNMT	4837	-1.27	0.16	0.00	0.13
TLR3	7098	-1.27	1.67	0.00	0.00
ABI3BP	25890	-1.27	-0.41	0.00	0.00
RNF216	54476	-1.27	-0.28	0.00	0.01
NDUFA4L2	56901	-1.27	0.19	0.00	0.20
ADARB1	104	-1.27	-0.36	0.00	0.00
UBAP2L	9898	-1.27	-0.15	0.00	0.31
GLIPR2	152007	-1.27	-0.71	0.00	0.00
MIR1978	100302173	-1.28	0.09	0.00	0.48
MR1	3140	-1.28	0.68	0.00	0.00
LMO3	55885	-1.28	-1.06	0.00	0.00
FLJ41484	650669	-1.28	-0.09	0.00	0.43
OXTR	5021	-1.28	-0.81	0.00	0.00
BATF2	116071	-1.28	1.38	0.00	0.00
CPT1A	1374	-1.28	1.63	0.00	0.00
YPEL3	83719	-1.28	-0.27	0.00	0.05
ALDH3A2	224	-1.28	-0.20	0.00	0.07
SSH2	85464	-1.28	-0.32	0.00	0.01
SLC2A6	11182	-1.28	0.38	0.00	0.00
ECH1	1891	-1.29	-0.35	0.00	0.01
HSD3B7	80270	-1.29	0.10	0.00	0.46
DYRK4	8798	-1.29	-0.05	0.00	0.75
SIRT5	23408	-1.29	0.22	0.00	0.12
SIL1	64374	-1.29	-0.09	0.00	0.61
SSBP2	23635	-1.29	0.35	0.00	0.00
SLC26A6	65010	-1.29	0.14	0.00	0.34
BCL2L13	23786	-1.29	0.67	0.00	0.00
BTN3A1	11119	-1.30	0.54	0.00	0.00
WDR81	124997	-1.30	0.01	0.00	0.98
NRCAM	4897	-1.30	-0.11	0.00	0.36
FTL	2512	-1.30	0.00	0.00	1.00
AXL	558	-1.30	0.10	0.00	0.54
HTATIP2	10553	-1.30	0.52	0.00	0.00
SUSD1	64420	-1.30	0.51	0.00	0.00
CIDEC	63924	-1.31	-0.03	0.00	0.85
C6orf138	442213	-1.31	0.98	0.00	0.00
ZNF564	163050	-1.31	-0.10	0.00	0.38
TRIB3	57761	-1.31	0.58	0.00	0.00

Table S4.1 continued

CENTG2	116987	-1.31	0.52	0.00	0.00
KYNU	8942	-1.31	1.46	0.00	0.00
SIX2	10736	-1.31	-0.15	0.00	0.17
EBF3	253738	-1.31	-0.56	0.00	0.00
HOXC8	3224	-1.32	-0.25	0.00	0.02
HOXC4	3221	-1.32	-0.29	0.00	0.01
SIX1	6495	-1.32	0.41	0.00	0.00
TRAM2	9697	-1.32	-0.18	0.00	0.18
KBTBD11	9920	-1.32	-0.53	0.00	0.00
TMEM219	124446	-1.33	0.55	0.00	0.00
PDXK	8566	-1.33	-0.66	0.00	0.00
FTHL2	2497	-1.33	0.37	0.00	0.03
OASL	8638	-1.33	1.74	0.00	0.00
NEURL1B	54492	-1.33	-0.32	0.00	0.02
COMT	1312	-1.33	0.20	0.00	0.14
PYCARD	29108	-1.34	0.42	0.00	0.00
OSBPL5	114879	-1.34	-0.07	0.00	0.67
MICAL2	9645	-1.34	0.01	0.00	0.97
ASTN2	23245	-1.34	-0.22	0.00	0.06
STAT6	6778	-1.34	0.54	0.00	0.00
ST3GAL4	6484	-1.34	-0.33	0.00	0.01
ACSM5	54988	-1.34	0.05	0.00	0.75
LOC100133866	100133866	-1.34	-0.13	0.00	0.30
TGFBR3	7049	-1.34	0.46	0.00	0.00
CXCL2	2920	-1.34	0.00	0.00	1.00
IL1R1	3554	-1.34	0.30	0.00	0.01
RTN1	6252	-1.35	-0.06	0.00	0.65
MUSK	4593	-1.35	0.62	0.00	0.00
TANC2	26115	-1.35	0.15	0.00	0.19
A4GALT	53947	-1.35	0.25	0.00	0.02
LOC644423	644423	-1.35	0.00	0.00	0.99
PTPRM	5797	-1.35	0.24	0.00	0.04
LNPEP	4012	-1.35	2.08	0.00	0.00
SCO2	9997	-1.35	-0.06	0.00	0.67
SLC7A11	23657	-1.36	1.35	0.00	0.00
TMBIM1	64114	-1.36	0.03	0.00	0.84
TCEAL3	85012	-1.36	-0.50	0.00	0.00
TRIM47	91107	-1.36	0.11	0.00	0.42
WISP1	8840	-1.36	-0.42	0.00	0.00
TRIM21	6737	-1.36	0.98	0.00	0.00
HLA-G	3135	-1.36	0.12	0.00	0.36
GMPPA	29926	-1.36	-0.09	0.00	0.52

Table S4.1 continued

SERPINE2	5270	-1.37	0.20	0.00	0.11
TLCD1	116238	-1.37	-0.41	0.00	0.00
WASF2	10163	-1.37	0.17	0.00	0.27
ABR	29	-1.37	-0.09	0.00	0.44
APCDD1L	164284	-1.38	0.28	0.00	0.01
DFNA5	1687	-1.38	-0.03	0.00	0.83
PSTPIP2	9050	-1.38	0.48	0.00	0.00
BIRC3	330	-1.38	0.76	0.00	0.00
SGCD	6444	-1.38	0.13	0.00	0.40
EGFR	1956	-1.39	1.21	0.00	0.00
NPTX2	4885	-1.39	0.12	0.00	0.30
DUSP10	11221	-1.39	-0.37	0.00	0.00
RELB	5971	-1.39	0.29	0.00	0.01
ZNF395	55893	-1.39	0.26	0.00	0.04
ANKFY1	51479	-1.40	0.09	0.00	0.49
MCEE	84693	-1.40	-0.34	0.00	0.00
CYGB	114757	-1.40	0.16	0.00	0.21
SUSD2	56241	-1.40	-2.02	0.00	0.00
ARHGEF19	128272	-1.40	0.11	0.00	0.37
STEAP3	55240	-1.40	-0.29	0.00	0.01
RRAS	6237	-1.41	-0.19	0.00	0.08
DDB2	1643	-1.41	0.11	0.00	0.41
GALK1	2584	-1.41	-0.53	0.00	0.00
OCEL1	79629	-1.41	-0.10	0.00	0.55
C8orf13	83648	-1.41	0.42	0.00	0.00
PLIN2	123	-1.41	0.61	0.00	0.00
PHF15	23338	-1.42	0.26	0.00	0.04
LOC653879	653879	-1.42	0.33	0.00	0.00
ZBTB4	57659	-1.42	0.00	0.00	0.99
GRINA	2907	-1.43	0.47	0.00	0.00
PLA2G4C	8605	-1.43	0.07	0.00	0.54
BHLHB2	8553	-1.43	0.62	0.00	0.00
FOXQ1	94234	-1.43	0.64	0.00	0.00
IMPA2	3613	-1.44	-0.34	0.00	0.01
LOC728431	728431	-1.44	0.45	0.00	0.00
IGFBP5	3488	-1.44	0.18	0.00	0.18
AHNAK	79026	-1.44	0.16	0.00	0.28
MYH13	8735	-1.44	-1.54	0.00	0.00
PLXNB1	5364	-1.44	-0.10	0.00	0.52
MT2A	4502	-1.44	0.08	0.00	0.49
SPTLC3	55304	-1.44	0.24	0.00	0.02
IRAK2	3656	-1.45	0.51	0.00	0.00

Table S4.1 continued

CCDC8	83987	-1.45	-0.09	0.00	0.55
ASNS	440	-1.45	0.71	0.00	0.00
ATL3	25923	-1.46	1.35	0.00	0.00
PLEKHA4	57664	-1.46	1.33	0.00	0.00
PPP3CC	5533	-1.46	0.26	0.00	0.02
TRIP6	7205	-1.46	0.00	0.00	0.99
LOC387763	387763	-1.46	0.20	0.00	0.06
CYP26B1	56603	-1.46	0.41	0.00	0.00
LIMA1	51474	-1.46	0.26	0.00	0.04
AGTRAP	57085	-1.47	0.62	0.00	0.00
RUSC1	23623	-1.47	0.03	0.00	0.80
P2RX6	9127	-1.47	-0.74	0.00	0.00
RALGDS	5900	-1.48	0.07	0.00	0.65
C14orf4	64207	-1.48	0.05	0.00	0.70
PSME2	5721	-1.48	0.43	0.00	0.00
PTPRU	10076	-1.48	0.26	0.00	0.04
GBP4	115361	-1.48	2.34	0.00	0.00
RGS20	8601	-1.48	0.09	0.00	0.61
RRBP1	6238	-1.48	0.10	0.00	0.48
PARP3	10039	-1.48	0.16	0.00	0.10
MIOS	54468	-1.48	0.12	0.00	0.31
DNAJB2	3300	-1.49	-0.02	0.00	0.93
ABCC3	8714	-1.49	0.13	0.00	0.28
MYBPHL	343263	-1.49	-0.52	0.00	0.00
CABC1	56997	-1.49	-0.18	0.00	0.20
IRAK3	11213	-1.50	1.05	0.00	0.00
MMP3	4314	-1.51	0.28	0.00	0.02
NFKB1	4790	-1.51	0.06	0.00	0.64
RBM43	375287	-1.52	0.32	0.00	0.03
LOC389386	389386	-1.52	1.37	0.00	0.00
CEBPD	1052	-1.52	0.16	0.00	0.13
PDK4	5166	-1.52	0.03	0.00	0.91
DDIT4	54541	-1.52	0.35	0.00	0.00
CA9	768	-1.53	0.52	0.00	0.00
MT1A	4489	-1.53	-0.12	0.00	0.37
COL8A1	1295	-1.53	0.47	0.00	0.00
TRNP1	388610	-1.53	0.01	0.00	0.96
VEZF1	7716	-1.53	-0.69	0.00	0.00
TRAF3IP2	10758	-1.53	0.73	0.00	0.00
PRKCD	5580	-1.53	0.35	0.00	0.00
TCP11L1	55346	-1.53	0.78	0.00	0.00
SLC25A28	81894	-1.53	0.73	0.00	0.00

Table S4.1 continued

HOXC6	3223	-1.54	-0.19	0.00	0.06
FAM175A	84142	-1.54	0.27	0.00	0.05
IDH3B	3420	-1.54	0.07	0.00	0.57
ZFP36L2	678	-1.54	0.73	0.00	0.00
ZC3H12A	80149	-1.55	0.54	0.00	0.00
ISCU	23479	-1.55	0.04	0.00	0.82
APOL3	80833	-1.55	0.98	0.00	0.00
LAP3	51056	-1.55	1.40	0.00	0.00
CLDN15	24146	-1.55	-0.37	0.00	0.01
PHLDA3	23612	-1.56	0.20	0.00	0.09
NT5E	4907	-1.56	0.49	0.00	0.00
TNS3	64759	-1.56	-0.16	0.00	0.12
SLC22A18	5002	-1.56	-0.05	0.00	0.75
RCAN1	1827	-1.57	0.99	0.00	0.00
ESPNL	339768	-1.57	-0.30	0.00	0.01
STC1	6781	-1.57	0.81	0.00	0.00
NFKBIZ	64332	-1.57	-0.26	0.00	0.05
ARID3A	1820	-1.57	0.29	0.00	0.02
BTG2	7832	-1.57	-0.12	0.00	0.32
PRDM1	639	-1.57	-0.09	0.00	0.52
TMEM45A	55076	-1.57	0.41	0.00	0.00
FHL2	2274	-1.58	-0.09	0.00	0.51
SESN1	27244	-1.58	-0.11	0.00	0.38
SLC16A3	9123	-1.58	0.49	0.00	0.00
MYF5	4617	-1.59	0.29	0.00	0.06
TNFAIP2	7127	-1.59	0.06	0.00	0.72
PSMB10	5699	-1.60	0.75	0.00	0.00
CYP27A1	1593	-1.60	-0.06	0.00	0.69
FPR1	2357	-1.60	0.93	0.00	0.00
MSC	9242	-1.61	0.25	0.00	0.03
ERAP2	64167	-1.61	0.77	0.00	0.00
GPX8	493869	-1.61	-0.33	0.00	0.00
ZFP90	146198	-1.61	-0.20	0.00	0.09
GFPT2	9945	-1.61	0.05	0.00	0.72
FUCA1	2517	-1.61	0.29	0.00	0.01
ADM	133	-1.62	0.64	0.00	0.00
C18orf56	494514	-1.62	0.21	0.00	0.08
BTN3A3	10384	-1.62	0.40	0.00	0.00
PDPN	10630	-1.62	-0.29	0.00	0.04
CSF3	1440	-1.63	0.81	0.00	0.00
KRT17	3872	-1.63	0.46	0.00	0.00
C1RL	51279	-1.63	-0.16	0.00	0.28

Table S4.1 continued

ID3	3399	-1.63	-0.07	0.00	0.69
NRXN2	9379	-1.63	-0.24	0.00	0.03
PTGES	9536	-1.63	0.31	0.00	0.09
RBCK1	10616	-1.64	0.33	0.00	0.01
DDX60L	91351	-1.64	0.66	0.00	0.00
PCTK3	5129	-1.64	0.59	0.00	0.00
IFIT5	24138	-1.65	-0.28	0.00	0.01
PLXNB2	23654	-1.65	0.20	0.00	0.09
HCG4	54435	-1.65	0.68	0.00	0.00
MOV10	4343	-1.66	0.69	0.00	0.00
KLF6	1316	-1.66	0.64	0.00	0.00
RTTN	25914	-1.66	0.01	0.00	0.96
SERPING1	710	-1.67	1.21	0.00	0.00
TNFRSF1A	7132	-1.67	0.27	0.00	0.03
SIX5	147912	-1.67	-0.36	0.00	0.00
DCN	1634	-1.67	0.26	0.00	0.01
LOC392437	392437	-1.67	0.11	0.00	0.37
FAM110B	90362	-1.67	-0.77	0.00	0.00
TNFRSF6B	8771	-1.68	0.01	0.00	0.97
HAS2	3037	-1.68	0.69	0.00	0.00
XBP1	7494	-1.68	0.34	0.00	0.00
IL32	9235	-1.68	0.06	0.00	0.65
ZNF337	26152	-1.69	-0.04	0.00	0.79
NINJ1	4814	-1.69	0.21	0.00	0.06
SQSTM1	8878	-1.70	0.18	0.00	0.12
TCEA3	6920	-1.71	-0.12	0.00	0.37
MAOA	4128	-1.71	-0.38	0.00	0.01
IFITM2	10581	-1.71	0.01	0.00	0.97
MOCOS	55034	-1.72	0.42	0.00	0.00
TSC22D3	1831	-1.73	0.20	0.00	0.20
GAS1	2619	-1.73	0.01	0.00	0.97
RTKN	6242	-1.73	-0.27	0.00	0.05
MUC1	4582	-1.74	0.35	0.00	0.02
RHBDF2	79651	-1.74	0.38	0.00	0.01
PPAP2B	8613	-1.74	0.06	0.00	0.67
CNTNAP1	8506	-1.74	0.15	0.00	0.17
HES4	57801	-1.75	1.07	0.00	0.00
CRISPLD2	83716	-1.75	-0.45	0.00	0.00
FKBP5	2289	-1.76	-0.10	0.00	0.42
CABYR	26256	-1.76	-0.04	0.00	0.79
BTN3A2	11118	-1.76	0.34	0.00	0.00
VASN	114990	-1.76	0.29	0.00	0.01

Table S4.1 continued

ZFHX3	463	-1.76	0.50	0.00	0.00
ITPRIP	85450	-1.76	0.55	0.00	0.00
SHISA5	51246	-1.76	-0.27	0.00	0.01
GSDMD	79792	-1.76	0.57	0.00	0.00
MSI2	124540	-1.76	1.33	0.00	0.00
TNFSF13B	10673	-1.76	2.19	0.00	0.00
PCK2	5106	-1.78	0.42	0.00	0.00
C4orf18	51313	-1.78	-0.51	0.00	0.00
FILIP1L	11259	-1.78	0.07	0.00	0.62
NACC2	138151	-1.79	0.98	0.00	0.00
HLA-C	3107	-1.79	0.60	0.00	0.00
MLPH	79083	-1.79	0.42	0.00	0.00
HLA-H	3136	-1.80	0.33	0.00	0.02
DAB2	1601	-1.81	0.11	0.00	0.37
FST	10468	-1.81	0.91	0.00	0.00
TMEM173	340061	-1.82	0.37	0.00	0.00
NFIL3	4783	-1.82	0.32	0.00	0.03
ATF5	22809	-1.83	0.50	0.00	0.00
MGC16121	84848	-1.83	0.01	0.00	0.96
DKFZp451A211	400169	-1.83	0.31	0.00	0.01
GALNTL2	117248	-1.83	0.99	0.00	0.00
KIAA1618	57714	-1.84	-0.30	0.00	0.01
TAPBP	6892	-1.84	0.47	0.00	0.00
FADS1	3992	-1.85	0.42	0.00	0.00
MAMDC2	256691	-1.86	-0.03	0.00	0.83
IHPK3	117283	-1.86	-0.09	0.00	0.49
PFKFB4	5210	-1.86	0.46	0.00	0.00
SLC39A14	23516	-1.87	0.48	0.00	0.00
SRGN	5552	-1.87	0.20	0.00	0.07
ANGPTL2	23452	-1.89	-0.78	0.00	0.00
APCDD1	147495	-1.89	-0.43	0.00	0.00
TXNIP	10628	-1.90	-0.01	0.00	0.95
WARS	7453	-1.92	1.71	0.00	0.00
HMOX1	3162	-1.92	0.98	0.00	0.00
RETSAT	54884	-1.93	-0.17	0.00	0.18
HLA-F	3134	-1.93	0.94	0.00	0.00
IGFBP4	3487	-1.94	0.40	0.00	0.00
CFLAR	8837	-1.94	0.53	0.00	0.00
SEMA4B	10509	-1.94	0.15	0.00	0.26
CYBASC3	220002	-1.95	-0.42	0.00	0.00
FTHL12	2504	-1.95	0.39	0.00	0.01
IFITM3	10410	-1.95	-0.13	0.00	0.26

Table S4.1 continued

EVC	2121	-1.95	-0.39	0.00	0.01
DGKA	1606	-1.96	0.10	0.00	0.46
SLC2A5	6518	-1.96	0.81	0.00	0.00
IRF7	3665	-1.96	1.14	0.00	0.00
TP53I3	9540	-1.96	0.19	0.00	0.11
UNC93B1	81622	-1.96	0.94	0.00	0.00
SPATA18	132671	-1.99	-0.21	0.00	0.07
CMBL	134147	-1.99	0.07	0.00	0.55
CXCL6	6372	-1.99	0.17	0.00	0.39
APBB3	10307	-1.99	0.04	0.00	0.77
IFI16	3428	-2.01	0.68	0.00	0.00
APOBEC3G	60489	-2.01	1.41	0.00	0.00
FTH1	2495	-2.03	-0.42	0.00	0.00
ZBTB16	7704	-2.04	0.49	0.00	0.00
CES2	8824	-2.04	0.06	0.00	0.66
PYGB	5834	-2.05	-0.12	0.00	0.32
PARP10	84875	-2.06	0.54	0.00	0.00
MT1E	4493	-2.07	0.13	0.00	0.31
C14orf159	80017	-2.07	0.52	0.00	0.00
ATOH8	84913	-2.07	-0.29	0.00	0.03
FTHL11	2503	-2.09	0.60	0.00	0.01
SCHIP1	29970	-2.09	0.06	0.00	0.67
SNAI2	6591	-2.09	0.20	0.00	0.06
C20orf127	140851	-2.10	0.44	0.00	0.00
PARP9	83666	-2.10	0.44	0.00	0.00
LOC441019	441019	-2.11	0.26	0.00	0.02
ANPEP	290	-2.12	0.36	0.00	0.00
S1PR3	1903	-2.12	1.13	0.00	0.00
LGALS3BP	3959	-2.12	0.31	0.00	0.03
COL7A1	1294	-2.12	-0.19	0.00	0.10
ZNFX1	57169	-2.13	0.82	0.00	0.00
CBS	875	-2.13	0.54	0.00	0.00
KIAA0247	9766	-2.13	0.37	0.00	0.00
NUPR1	26471	-2.14	0.83	0.00	0.00
GDF15	9518	-2.17	0.92	0.00	0.00
SERPINE1	5054	-2.17	0.46	0.00	0.00
PHGDH	26227	-2.18	0.31	0.00	0.03
GAL3ST4	79690	-2.19	0.30	0.00	0.02
TNFAIP6	7130	-2.21	1.31	0.00	0.00
DDR2	4921	-2.22	0.63	0.00	0.00
LOC643384	643384	-2.22	1.02	0.00	0.00
RARRES3	5920	-2.22	1.27	0.00	0.00

Table S4.1 continued

SERPINA3	12	-2.22	-0.29	0.00	0.14
STXBP6	29091	-2.23	-0.14	0.00	0.24
DUSP1	1843	-2.24	0.41	0.00	0.00
GBP1	2633	-2.24	1.01	0.00	0.00
PSMB8	5696	-2.25	0.71	0.00	0.00
TNIP1	10318	-2.25	0.28	0.00	0.01
IRF9	10379	-2.25	0.29	0.00	0.02
PAPPA	5069	-2.26	0.40	0.00	0.00
IL7R	3575	-2.26	0.57	0.00	0.00
PHF11	51131	-2.27	0.66	0.00	0.00
RTP4	64108	-2.28	1.21	0.00	0.00
SP100	6672	-2.28	1.22	0.00	0.00
DKK1	22943	-2.29	0.47	0.00	0.00
STOM	2040	-2.29	0.62	0.00	0.00
EIF2AK2	5610	-2.30	0.54	0.00	0.00
NDRG1	10397	-2.33	0.66	0.00	0.00
ADAR	103	-2.34	0.31	0.00	0.01
C19orf66	55337	-2.35	0.62	0.00	0.00
FTHL8	2501	-2.35	0.51	0.00	0.00
DDX60	55601	-2.35	0.45	0.00	0.00
MTE	644314	-2.38	0.50	0.00	0.00
DUSP19	142679	-2.38	0.75	0.00	0.00
RSAD2	91543	-2.40	2.36	0.00	0.00
GBP2	2634	-2.40	0.66	0.00	0.00
IFI44	10561	-2.42	0.29	0.00	0.00
H1F0	3005	-2.43	0.50	0.00	0.00
CEBPB	1051	-2.44	0.28	0.00	0.01
MT1X	4501	-2.44	0.32	0.00	0.02
XPC	7508	-2.46	0.26	0.00	0.03
DDX58	23586	-2.47	0.02	0.00	0.91
CXCL5	6374	-2.47	0.82	0.00	0.00
SLC7A2	6542	-2.48	-0.11	0.00	0.38
USP18	11274	-2.50	1.00	0.00	0.00
C9orf169	375791	-2.51	-0.05	0.00	0.78
TRIM25	7706	-2.51	0.51	0.00	0.00
		-2.51	-0.13	0.00	0.40
CCL5	6352	-2.53	1.89	0.00	0.00
SAMD9L	219285	-2.53	1.26	0.00	0.00
UBA7	7318	-2.56	0.57	0.00	0.00
FTHL3	2498	-2.58	0.45	0.00	0.05
TRIM22	10346	-2.58	-0.02	0.00	0.90
PRIC285	85441	-2.61	0.86	0.00	0.00

Table S4.1 continued

AGRN	375790	-2.61	0.38	0.00	0.00
CA12	771	-2.63	0.95	0.00	0.00
C10orf10	11067	-2.64	0.14	0.00	0.35
IRF1	3659	-2.64	0.36	0.00	0.01
LOC729009	729009	-2.65	0.66	0.00	0.00
CCL2	6347	-2.68	0.44	0.00	0.00
STAT2	6773	-2.69	0.19	0.00	0.14
CHI3L2	1117	-2.70	0.26	0.00	0.12
OAS2	4939	-2.71	0.46	0.00	0.00
TNFRSF14	8764	-2.72	0.57	0.00	0.00
PTX3	5806	-2.74	0.94	0.00	0.00
HLA-B	3106	-2.75	0.34	0.00	0.01
PARP14	54625	-2.83	0.52	0.00	0.00
C1R	715	-2.83	0.49	0.00	0.00
DHX58	79132	-2.84	0.70	0.00	0.00
SAMD9	54809	-2.86	1.40	0.00	0.00
TNFAIP3	7128	-2.90	0.28	0.00	0.01
STAT1	6772	-2.92	0.82	0.00	0.00
MT1M	4499	-2.92	1.19	0.00	0.00
ISG20	3669	-2.94	2.36	0.00	0.00
SP110	3431	-2.94	1.15	0.00	0.00
TMEM140	55281	-2.95	1.00	0.00	0.00
MLKL	197259	-2.99	1.46	0.00	0.00
NFKBIA	4792	-3.00	0.28	0.00	0.03
VCAM1	7412	-3.01	0.63	0.00	0.00
UBE2L6	9246	-3.08	0.58	0.00	0.00
PSMB9	5698	-3.14	0.95	0.00	0.00
PARP12	64761	-3.17	0.78	0.00	0.00
HERC5	51191	-3.20	1.33	0.00	0.00
LY6E	4061	-3.25	0.25	0.00	0.07
TAP1	6890	-3.26	0.70	0.00	0.00
VWCE	220001	-3.32	-0.12	0.00	0.33
CXCL1	2919	-3.52	0.20	0.00	0.42
XAF1	54739	-3.62	0.62	0.00	0.00
IFIH1	64135	-3.70	1.66	0.00	0.00
HERC6	55008	-3.73	0.52	0.00	0.00
SLC15A3	51296	-3.75	0.78	0.00	0.00
C1QTNF1	114897	-3.85	0.77	0.00	0.00
IFI35	3430	-3.85	0.92	0.00	0.00
IFIT3	3437	-3.86	1.81	0.00	0.00
IL8	3576	-3.88	1.19	0.00	0.00
OAS3	4940	-4.03	0.87	0.00	0.00

Table S4.1 continued

MX2	4600	-4.07	0.96	0.00	0.00
LOC100129681	100129681	-4.07	0.52	0.00	0.00
EPSTI1	94240	-4.12	0.66	0.00	0.00
SAA1	6288	-4.14	0.26	0.00	0.03
IFI6	2537	-4.24	-0.04	0.00	0.82
BST2	684	-4.25	0.21	0.00	0.06
ECGF1	1890	-4.30	0.62	0.00	0.00
ISG15	9636	-4.31	0.13	0.00	0.31
IFIT2	3433	-4.31	2.07	0.00	0.00
IFIT1	3434	-4.38	0.72	0.00	0.00
OAS1	4938	-4.61	1.45	0.00	0.00
SOD2	6648	-4.62	1.53	0.00	0.00
IFI44L	10964	-5.12	0.47	0.00	0.00
CFB	629	-5.62	0.63	0.00	0.00
MX1	4599	-5.67	0.26	0.00	0.02
IFITM1	8519	-6.12	0.34	0.00	0.01
IFI27	3429	-6.53	0.27	0.00	0.01

symbol = gene symbol
fc = fold change
fdr = false discovery rate

Table S4.2. Gene Ontology Analysis of genes upregulated by DUX4-fl

OR	EC	Count	Size	Term
11.4369	0.56126	5	22	RNA polymerase II transcription mediator activity
9.93702	0.62393	5	25	mediator complex
6.29225	2.62052	14	105	nuclear speck
6.24826	1.29488	7	52	nuclear mRNA splicing, via spliceosome
5.07493	1.10245	5	44	axon guidance
4.88496	1.14003	5	45	mRNA binding
4.75221	1.40668	6	56	response to unfolded protein
4.51467	3.24445	13	130	spliceosomal complex
4.50874	2.4868	10	99	RNA splicing, via transesterification reactions
4.48374	1.23084	5	49	mesenchymal cell differentiation
4.44298	2.26073	9	90	nucleic acid transport
4.44298	2.26073	9	90	establishment of RNA localization
3.94896	3.96666	14	163	RNA splicing
3.31757	1.62248	5	70	regulation of metabolic process
3.28233	1.63275	5	65	positive regulation of cell development
2.93327	4.4271	12	183	mRNA processing
2.85139	1.85882	5	74	regulation of cell morphogenesis involved in differentiation
2.76598	4.67218	12	186	nuclear transport
2.73419	1.93257	5	77	urogenital system development
2.73168	1.93418	5	77	microtubule-based movement
2.51447	2.93895	7	117	leukocyte differentiation
2.12713	5.4763	11	233	multicellular organismal development
2.12364	6.47331	13	258	interspecies interaction between organisms
2.10256	3.97755	8	159	gamete generation
2.0952	7.0648	14	309	nucleoplasm
2.03526	4.62194	9	184	spermatogenesis
1.99266	5.25652	10	209	apoptosis
1.85637	40.1563	65	1609	intracellular organelle lumen
1.84932	7.93768	14	316	multicellular organism reproduction
1.7945	41.2794	65	1654	membrane-enclosed lumen
1.72114	8.48494	14	341	regulation of multicellular organismal development
1.68006	99.5252	132	4000	nucleus
1.67516	12.5327	20	512	RNA binding
1.66302	31.5739	46	1223	regulation of cellular process
1.64141	16.0475	25	643	nucleolus
1.57575	38.1464	54	1498	regulation of macromolecule metabolic process
1.54395	72.7072	95	2819	cellular nitrogen compound metabolic process

OR = odds ratio
EC = expected count
Count = actual count in data set
Size = size of GO term
Term = GO term

Table S4.3. Gene Ontology Analysis of genes downregulated by DUX4-fl

OR	EC	Count	Size	Term
17.99454	0.461189	6	22	chemokine activity
17.02899	0.40639	5	18	activation of immune response
15.93194	1.893662	21	84	response to virus
13.69643	0.478537	5	21	activation of plasma proteins involved in acute inflammatory response
12.88936	0.501325	5	22	regulation of viral reproduction
11.47143	3.073569	26	150	immune response
10.07058	1.626719	13	73	innate immune response
8.863863	0.664982	5	34	defense response
7.869162	0.74721	5	33	leukocyte migration
7.688316	0.761086	5	34	antigen processing and presentation
7.615385	1.400399	9	68	negative regulation of biological process
7.570055	0.93004	6	41	humoral immune response
6.296703	1.089983	6	48	response to glucocorticoid stimulus
5.753133	0.979862	5	43	response to hydrogen peroxide
5.03714	1.101595	5	49	response to steroid hormone stimulus
5.011867	1.551271	7	74	G-protein-coupled receptor binding
4.965368	1.116587	5	49	cell chemotaxis
4.709181	3.12766	13	140	anti-apoptosis
4.534895	1.4584	6	64	signal initiation by diffusible mediator
4.498795	6.540011	25	287	interspecies interaction between organisms
4.411115	2.005544	8	90	regulation of immune response
4.336898	1.254502	5	60	transferase activity, transferring glycosyl groups
4.336606	1.777424	7	78	protein maturation
4.250743	1.283484	5	57	response to lipopolysaccharide
4.243549	3.160098	12	139	positive regulation of immune system process
4.042572	1.34314	5	60	chemotaxis
3.681051	2.054385	7	98	pattern binding
3.618056	2.400128	8	111	inflammatory response
3.579944	2.107496	7	101	cytokine receptor binding
3.542467	8.041374	25	380	extracellular space
3.519614	1.836402	6	81	response to bacterium
3.50269	2.158862	7	95	response to nutrient
3.452381	2.179423	7	105	carbohydrate binding
3.377922	5.255012	16	239	response to external stimulus
3.339047	2.255962	7	99	aging
3.323799	2.596461	8	116	positive regulation of response to stimulus
3.260458	3.668786	11	161	locomotory behavior

Table S4.3 continued

3.204482	1.663487	5	73	negative regulation of binding
3.167963	4.840465	14	216	negative regulation of cell proliferation
3.157695	1.686275	5	74	cellular response to extracellular stimulus
3.072478	6.859036	19	301	negative regulation of cell death
2.9375	4.040065	11	179	cell-cell signaling
2.911979	4.074034	11	182	cell adhesion
2.892365	1.823791	5	87	growth factor activity
2.78715	2.666137	7	117	leukocyte differentiation
2.749994	4.676694	12	221	soluble fraction
2.677219	19.78601	45	935	extracellular region
2.650262	1.983392	5	90	immune effector process
2.556128	3.301196	8	156	microsome
2.399631	2.620562	6	115	regulation of transcription regulator activity
2.390475	3.076312	7	135	muscle tissue development
2.374029	2.641353	6	126	transcription corepressor activity
2.363266	4.94011	11	218	locomotion
2.237124	5.195548	11	228	regulation of response to stress
2.182383	8.340223	17	366	positive regulation of cell death
2.147643	10.29764	20	617	response to stimulus
2.127693	8.050932	16	368	cell differentiation
2.031607	7.274202	14	347	peptidase activity
2.00897	5.742448	11	252	immune system development
1.986315	5.263911	10	231	blood vessel development
1.622598	21.94903	33	961	signal transduction

OR = odds ratio
EC = expected count
Count = actual count in data set
Size = size of GO term
Term = GO term

Table S4.4. Gene Ontology Analysis of genes upregulated by DUX4-fl 8-fold or more

OR	EC	Count	Size	Term
58.85380117	0.0435084	2	11	phosphate binding
10.1148429	0.2166402	2	73	M phase of meiotic cell cycle
6.215384615	0.3472178	2	117	leukocyte differentiation
6.116685083	0.5460519	3	184	spermatogenesis
4.643037975	0.7122417	3	240	gamete generation
4.609476512	5.5730248	17	1409	zinc ion binding
4.1721673	0.7894012	3	266	regulation of immune system process
3.570087799	10.248195	22	2591	ion binding

OR = odds ratio
EC = expected count
Count = actual count in data set
Size = size of GO term
Term = GO term

Table S4.5. Repeat families bound by DUX4

Repeat class	DUX4-fl binding prevalence	overall genome prevalence	DUX4-fl enrichment
LTR/ERVL-MaLR	0.35716	0.036	9.92
LTR/ERV	0.00032	6.00E-05	5.33
LTR/ERVK	0.00803	0.0027	2.97
LTR/ERVL	0.04558	0.01823	2.50
rRNA	1.00E-04	6.00E-05	1.67
SINE/tRNA	0.00011	7.00E-05	1.57
Unknown	0.00063	0.00043	1.47
DNA/TcMar-Mariner	0.00105	0.00092	1.14
DNA/TcMar-Tigger	0.0124	0.01121	1.11
LTR/Gypsy	0.00081	0.00076	1.07
LINE/CR1	0.00356	0.00356	1.00
DNA/hAT?	0.00016	0.00017	0.94
LINE/L2	0.02978	0.03443	0.86
SINE/MIR	0.02317	0.0281	0.82
LTR/ERV1	0.01878	0.02604	0.72
Satellite	0.00068	0.00103	0.66
LINE/L1	0.0938	0.16059	0.58
DNA/hAT-Blackjack	0.00064	0.00113	0.57
Simple_repeat	0.00442	0.00836	0.53
DNA/hAT-Charlie	0.00761	0.01486	0.51
DNA/hAT-Tip100	0.00098	0.0022	0.45
Satellite/centr	0.00047	0.00243	0.19
SINE/Alu	0.00777	0.10171	0.08

DUX4-fl binding prevalence: fraction of all DUX4-fl peaks
overall genome prevalence: fraction of whole genome
DUX4-fl enrichment: (DUX4-fl binding prevalence)/(overall genome prevalence)

Table S4.6. Non-repeat element DUX4-fl binding sites associated with expressed genes

space	max.cov	Full.fc	Symbol	dist2tss
chr22	124	8.4	RFPL1	-3042
chr7	114	8.3	hCG_1651160	-892
chr7	215	8.3	hCG_1651160	-192
chr6	232	8.3	RFPL4B	-321
chr19	364	8.3	ZSCAN4	1430
chr2	85	8.1	TRIM43	-5584
chr2	47	8.1	TRIM43	1385
chr11	178	7.9	TRIM48	-151
chr1	99	7.8	PRAMEF12	-1230
chr13	56	7.5	CCNA1	-2425
chr13	58	7.5	CCNA1	1874
chr22	118	7.3	RFPL2	-2853
chr22	129	7.3	RFPL2	3057
chr14	344	7	PNP	-104
chr11	85	6.9	TRIM49L1	200
chr11	109	6.9	TRIM49L2	-202
chr8	55	6.4	DEFB103A	-2289
chr19	110	5.5	ZNF296	182
chr11	373	4.2	SFRS2B	42
chr5	91	4.1	PPP2R2B	2498
chr20	55	4.1	ZNF217	-2168
chr20	122	4.1	ZNF217	3546
chr12	224	3.9	ZNF705A	-5106
chr22	118	3.8	PANX2	-2040
chr19	98	3.8	ZSCAN5B	-5014
chr19	67	3.8	ZSCAN5B	-4101
chr19	89	3.8	ZSCAN5B	4892
chr16	118	3.7	SIAH1	-2337
chr12	129	3.6	FAM90A1	-1597
chr12	54	3.6	PRR4	83
chr3	117	3.4	DBR1	3981
chr11	105	3.3	SPTY2D1	-797
chr11	74	3.3	SPTY2D1	5934
chr14	117	3.2	FBXO33	-2226
chr19	83	3.2	GTF2F1	134
chr17	177	3.2	JUP	-701
chr22	234	3.2	TFIP11	-1353
chr21	96	3.1	CLDN14	-2523
chr20	53	2.9	CSE1L	-5914

Table S4.6 continued

chr20	57	2.9	CSE1L	-1935
chr2	85	2.9	PELI1	4936
chr7	63	2.7	BZW2	1939
chr7	86	2.7	BZW2	1955
chr10	265	2.7	CCNJ	-637
chr1	116	2.6	DENND2C	-3169
chr14	257	2.6	PABPN1	-529
chr7	128	2.6	SRRT	1546
chr19	104	2.6	USP29	2090
chr19	66	2.6	USP29	3676
chr14	159	2.5	C14orf102	-558
chr11	46	2.4	CTR9	1913
chr21	229	2.4	SYNJ1	-766
chr6	159	2.3	NFYA	49
chr1	132	2.2	C1orf63	-1326
chr17	283	2.2	HOXB2	-2471
chr3	51	2.2	PVRL3	2678
chr6	250	2.1	C6orf191	61
chr10	70	2.1	CBARA1	2340
chr10	250	2.1	CBARA1	4786
chr21	244	2.1	SON	-2951
chr10	56	2	AVPI1	126
chr10	200	2	FRG2B	598
chr16	46	2	RBBP6	-4040
chr16	323	2	RBBP6	-1922
chr16	76	2	RBBP6	-1619
chr1	139	1.9	EXOSC10	150
chr2	150	1.9	GPBAR1	-224
chr9	85	1.9	NANS	1843
chr20	62	1.9	SNAI1	769
chr18	65	1.9	TAF4B	-1741
chr21	274	1.8	C21orf91	-803
chr2	163	1.8	CLK1	4634
chr1	93	1.8	KDM5B	301
chr12	73	1.8	KIF21A	-4249
chr12	37	1.8	KIF21A	-1030
chr17	43	1.8	MED13	-2755
chr10	172	1.8	SEC61A2	-1861
chr11	133	1.8	SPRYD5	-2855
chr14	50	1.7	C14orf138	-4696

Table S4.6 continued

chr14	53	1.7	C14orf138	326
chr12	55	1.7	DDX47	-4416
chr5	269	1.7	MAST4	-1253
chr10	57	1.7	PRPF18	3230
chr6	105	1.7	PTP4A1	-2227
chr9	198	1.6	CTNNAL1	641
chr1	205	1.6	EGLN1	-1
chr9	69	1.6	MAPKAP1	-4397
chr9	153	1.6	MAPKAP1	3619
chr2	52	1.6	RTN4	-2562
chr2	115	1.6	RTN4	1853
chr10	43	1.6	SAMD8	3338
chr9	88	1.6	SH3GL2	-4600

space: chromosome location of binding site
max.cov: peak height
full.fc: expression fold change for DUX4-fl targets
dist2tss: distance to TSS

Table S4.7. DUX4-fl expression in FSHD and control muscle

Primary Human Myoblasts

Sample #	Formal Identifer	DUX4-fl expression	Disease Status
1	MB135	not detected	control
2	MB196	not detected	control
3	MB201	not detected	control
4	MB209	not detected	control
5	MB230	not detected	control
6	MB54-1*	not detected	control*
7	MB073	detected	FSHD1
8	MB183	detected	FSHD1
9	MB197	detected	FSHD1
10	MB216	detected	FSHD1
11	MB200	detected	FSHD2
12	MB54-2*	detected	FSHD1*

Muscle Biopsies

Sample #	Formal Identifer	DUX4-fl expression	Disease Status
1	C-20	not detected	control
2	C-22	not detected	control
3	C-33	not detected	control
4	C-38	not detected	control
5	C-40	not detected	control
6	C-2333/C-2397	not detected	control
7	F-2315	not detected	FSHD1
8	F-2316	detected	FSHD1
9	F-2319	not detected	FSHD1
10	F-2326	not detected	FSHD1
11	F-2331	detected	FSHD1
12	F-2367	detected	FSHD1
13	F-2369	detected	FSHD1
14	F-2377	detected	FSHD1

Table S4.8. Genes induced by lenti-GFP and lenti-DUX4-s but poorly induced by lenti-DUX4-fl

Gene symbol	GFP v. NoLenti (Fc)	Full v. GFP (Fc)	Short v. GFP (Fc)	Full v. NoLenti (Fc)	Short vs. NoLenti (Fc)
ABCA1	2.29	-1.19	-0.32	1.10	1.97
ABI3BP	1.29	-1.24	-0.40	0.04	0.89
ACSM5	1.17	-1.21	0.05	-0.04	1.22
ADAR	1.36	-2.32	0.32	-0.96	1.68
ADAR	1.49	-2.18	0.36	-0.69	1.84
ADCK3	1.64	-1.42	-0.17	0.22	1.47
AGAP1	1.39	-1.27	0.52	0.12	1.91
AGRN	2.63	-2.53	0.38	0.10	3.01
AK3	1.07	-1.25	-0.42	-0.18	0.65
ALDH3A2	1.13	-1.24	-0.19	-0.12	0.93
ALOX15B	1.08	-1.21	0.72	-0.13	1.80
ANGPT1	1.03	-1.08	0.22	-0.06	1.25
ANKRA2	1.04	-1.17	0.33	-0.14	1.36
ANKRA2	1.14	-1.34	0.25	-0.20	1.39
ANPEP	1.40	-2.09	0.36	-0.69	1.76
APCDD1	1.82	-1.70	-0.41	0.11	1.41
APOBEC3G	1.36	-1.74	1.36	-0.39	2.72
ATL3	1.17	-1.43	1.34	-0.26	2.51
BATF2	1.05	-1.07	1.27	-0.02	2.32
BCL3	1.08	-1.20	-0.04	-0.12	1.03
BCL6	1.24	-1.25	0.36	-0.01	1.60
BIRC3	1.78	-2.14	1.06	-0.36	2.84
BST2	4.69	-4.11	0.21	0.58	4.91
BTG2	1.24	-1.52	-0.12	-0.28	1.12
BTN3A2	1.40	-1.58	0.32	-0.18	1.72
BTN3A3	1.34	-1.39	0.37	-0.05	1.72
C10orf10	2.19	-2.51	0.14	-0.32	2.33
C14orf159	1.26	-1.96	0.51	-0.70	1.77
C18orf56	1.45	-1.51	0.21	-0.06	1.66
C19orf66	1.71	-2.18	0.60	-0.46	2.32
C1QTNF1	3.51	-3.64	0.76	-0.12	4.28
C1R	3.43	-2.76	0.49	0.67	3.92
C1R	4.10	-2.99	0.48	1.11	4.57
C1RL	1.21	-1.45	-0.15	-0.24	1.06
C1S	3.51	-2.74	0.54	0.76	4.05
C4orf34	1.46	-1.04	0.60	0.42	2.06
C6orf138	1.28	-1.18	0.95	0.11	2.23

Table S4.8 continued

C9orf169	1.55	-2.34	-0.05	-0.79	1.50
CA12	1.51	-2.36	0.93	-0.85	2.44
CA9	1.38	-1.31	0.49	0.07	1.87
CABYR	1.13	-1.59	-0.04	-0.46	1.09
CCL2	2.39	-2.38	0.43	0.01	2.82
CCL5	2.47	-2.25	1.86	0.22	4.33
CCL5	2.79	-3.05	1.43	-0.26	4.22
CCND2	1.35	-1.05	-0.53	0.30	0.82
CCND2	1.21	-1.01	-0.54	0.20	0.67
CD68	2.37	-1.09	1.18	1.28	3.54
CDK18	1.48	-1.41	0.56	0.07	2.03
CDKN1A	1.58	-2.01	0.15	-0.43	1.73
CEBPB	1.32	-2.45	0.28	-1.13	1.59
CEBPD	2.03	-1.54	0.16	0.49	2.19
CES2	1.21	-2.02	0.06	-0.80	1.27
CES2	1.46	-1.32	0.30	0.14	1.76
CFB	5.39	-5.32	0.66	0.07	6.04
CFD	1.46	-1.09	0.24	0.37	1.70
CFLAR	1.17	-1.87	0.53	-0.70	1.71
CHI3L2	2.88	-2.45	0.25	0.43	3.14
CIDEC	1.30	-1.13	-0.03	0.17	1.27
CLDN15	1.76	-2.32	-0.50	-0.55	1.26
CLMP	1.10	-1.14	0.02	-0.04	1.12
CMBL	1.78	-1.96	0.07	-0.18	1.85
COL7A1	1.67	-2.09	-0.19	-0.41	1.49
COPS8	1.08	-1.24	0.14	-0.15	1.23
CORO6	1.15	-1.09	-0.29	0.06	0.86
CSF3	1.31	-1.38	0.75	-0.07	2.06
CSF3	1.17	-1.29	0.51	-0.11	1.68
CST3	1.35	-1.11	0.17	0.24	1.53
CXCL1	3.16	-3.21	0.20	-0.04	3.36
CXCL2	1.09	-1.12	0.00	-0.04	1.09
CXCL5	1.80	-2.16	0.80	-0.37	2.60
CXCL5	2.55	-2.80	0.75	-0.24	3.30
CXCL6	1.61	-1.71	0.16	-0.10	1.77
CXCL6	2.89	-2.76	0.36	0.12	3.25
CYBASC3	1.57	-1.92	-0.42	-0.35	1.15
CYFIP2	1.39	-2.08	0.27	-0.69	1.66
CYGB	1.24	-1.22	0.14	0.02	1.39
CYP26B1	1.56	-1.35	0.40	0.20	1.96

Table S4.8 continued

CYP27A1	1.60	-1.56	-0.06	0.04	1.54
DCN	1.78	-1.56	0.26	0.22	2.04
DCN	1.26	-1.36	0.52	-0.11	1.77
DCN	2.54	-2.12	0.21	0.42	2.74
DDB2	1.07	-1.23	0.10	-0.16	1.16
DDR2	2.23	-2.18	0.63	0.05	2.86
DDX58	2.12	-2.30	0.02	-0.18	2.14
DDX60	1.86	-2.11	0.44	-0.26	2.30
DDX60L	1.17	-1.45	0.63	-0.29	1.80
DGKA	1.15	-1.73	0.10	-0.59	1.24
DGKA	1.54	-2.06	0.22	-0.52	1.76
DHX58	2.35	-2.52	0.68	-0.17	3.04
DKK1	2.06	-2.30	0.47	-0.24	2.53
DRAM1	1.17	-1.06	-0.36	0.11	0.81
DUSP10	1.05	-1.72	-0.08	-0.67	0.97
DUSP19	1.52	-2.36	0.76	-0.84	2.28
EGFR	1.24	-1.29	1.19	-0.06	2.42
EIF2AK2	2.72	-2.27	0.55	0.45	3.27
EPSTI1	4.36	-4.03	0.68	0.34	5.05
ERAP2	1.62	-1.56	0.77	0.05	2.38
FAM160B1	1.90	-1.55	1.54	0.34	3.44
FAM198B	1.37	-1.75	-0.51	-0.38	0.86
FAM198B	1.56	-1.67	-0.29	-0.10	1.28
FBXO32	2.57	-1.16	0.76	1.41	3.33
FBXO32	2.99	-1.80	0.77	1.20	3.76
FILIP1L	1.21	-1.72	0.06	-0.51	1.28
FKBP5	1.06	-1.77	-0.11	-0.71	0.95
FOXQ1	1.49	-1.21	0.59	0.27	2.07
FRMD3	1.26	-1.01	-0.27	0.26	1.00
FST	1.88	-1.80	0.92	0.08	2.79
FTH1	1.17	-2.00	-0.41	-0.84	0.76
FTH1P3	1.58	-2.55	0.46	-0.97	2.04
FUCA1	1.55	-1.53	0.28	0.02	1.84
GALNTL2	1.63	-1.60	0.95	0.03	2.58
GAS1	1.25	-1.73	0.01	-0.48	1.26
GAS1	1.31	-1.55	0.09	-0.24	1.40
GBP1	1.66	-2.20	1.00	-0.54	2.66
GBP1	1.54	-1.91	1.03	-0.37	2.57
GBP2	2.19	-2.39	0.68	-0.19	2.87
GBP4	1.34	-1.25	2.24	0.08	3.58

Table S4.8 continued

GDF15	1.77	-2.05	0.92	-0.28	2.69
GFPT2	1.73	-1.59	0.05	0.14	1.78
GRINA	1.77	-1.40	0.47	0.37	2.24
GRINA	1.16	-1.20	0.42	-0.04	1.58
H1F0	1.21	-2.36	0.50	-1.15	1.71
HCG4	1.25	-1.63	0.68	-0.38	1.92
HECW2	1.22	-1.05	0.06	0.17	1.27
HERC5	5.37	-3.16	1.35	2.21	6.71
HERC6	3.89	-3.57	0.52	0.32	4.41
HIPK2	1.27	-1.16	0.36	0.12	1.63
HLA-A	1.22	-1.07	0.57	0.15	1.80
HLA-B	3.45	-2.75	0.33	0.70	3.78
HLA-C	2.04	-1.73	0.60	0.32	2.65
HLA-E	2.49	-1.12	0.81	1.37	3.30
HLA-F	1.87	-1.74	0.91	0.14	2.79
HLA-F	2.70	-2.11	0.61	0.59	3.32
HLA-G	1.08	-1.21	0.11	-0.12	1.19
HLA-H	1.30	-1.67	0.32	-0.37	1.62
HLA-H	2.55	-2.19	0.35	0.36	2.91
HOXC13	1.00	-1.09	0.44	-0.09	1.45
IFI16	1.84	-1.98	0.68	-0.14	2.52
IFI27	7.30	-6.38	0.30	0.92	7.60
IFI35	3.41	-3.73	0.92	-0.32	4.33
IFI44	3.29	-2.39	0.30	0.90	3.59
IFI44L	5.75	-5.01	0.46	0.74	6.22
IFI6	5.09	-4.26	-0.04	0.82	5.05
IFIH1 (MDA5)	3.69	-3.45	1.64	0.25	5.34
IFIT1	4.79	-4.36	0.72	0.43	5.50
IFIT2	4.37	-4.16	2.08	0.20	6.44
IFIT3	3.68	-3.64	1.81	0.04	5.49
IFIT3	5.23	-4.77	1.10	0.46	6.33
IFIT3	3.37	-3.15	0.53	0.21	3.90
IFIT5	1.07	-1.43	-0.26	-0.36	0.81
IFITM1	6.73	-6.00	0.35	0.73	7.08
IFITM2	1.29	-1.75	0.01	-0.46	1.29
IFITM3	1.98	-2.00	-0.14	-0.01	1.84
IGDCC4	1.22	-1.59	-0.62	-0.38	0.59
IGFBP4	2.22	-1.93	0.42	0.30	2.64
IGFBP5	1.22	-1.28	0.17	-0.07	1.38

Table S4.8 continued

IGFBP5	1.54	-1.31	0.28	0.23	1.82
IL18BP	2.01	-1.16	1.37	0.85	3.37
IL1R1	1.32	-1.23	0.29	0.09	1.61
IL32	2.03	-2.21	0.15	-0.18	2.18
IL7R	1.51	-2.00	0.55	-0.49	2.05
IL8	3.52	-3.65	1.17	-0.13	4.69
IL8	5.33	-5.45	0.93	-0.12	6.26
IRAK3	1.35	-1.34	1.02	0.01	2.37
IRF7	2.86	-1.84	1.13	1.02	3.99
IRF7	2.76	-2.16	0.96	0.59	3.71
IRF9	2.43	-2.23	0.30	0.20	2.72
ISG15	5.06	-4.32	0.14	0.74	5.20
ISG20	2.30	-2.68	2.33	-0.38	4.63
ITPRIP	1.10	-1.71	0.55	-0.61	1.66
KIAA0247	2.30	-2.01	0.36	0.29	2.66
KLF9	1.02	-1.25	-0.05	-0.22	0.97
KRT17	1.38	-1.41	0.43	-0.03	1.81
KYNU	1.12	-1.09	1.34	0.03	2.46
LAP3	1.75	-1.54	1.41	0.21	3.16
LGALS3BP	2.68	-2.11	0.32	0.57	3.00
LNPEP	1.76	-1.28	2.05	0.48	3.82
LTBR	1.17	-1.15	0.40	0.02	1.57
LUM	1.15	-1.11	0.39	0.04	1.54
LY6E	3.69	-3.24	0.26	0.44	3.94
MAMDC2	1.73	-1.84	-0.03	-0.11	1.71
MAOA	1.42	-1.50	-0.35	-0.08	1.07
MLKL	1.57	-2.72	1.44	-1.15	3.01
MME	1.09	-1.03	-0.18	0.05	0.90
MMP3	1.24	-1.29	0.26	-0.05	1.50
MMP7	1.28	-1.06	0.19	0.22	1.48
MOCOS	1.13	-1.66	0.42	-0.53	1.55
MR1	1.19	-1.18	0.66	0.01	1.85
MSI2	1.97	-1.70	1.32	0.28	3.29
MT1F	1.22	-1.15	-0.53	0.08	0.70
MT1G	1.29	-1.24	0.26	0.05	1.55
MT1M	2.70	-2.87	1.20	-0.17	3.90
MT1X	1.98	-2.45	0.31	-0.47	2.29
MTSS1	1.33	-1.07	-0.39	0.26	0.94
MTSS1	2.17	-1.35	-0.44	0.81	1.72
MUC1	1.43	-1.58	0.34	-0.14	1.77

Table S4.8 continued

MUC1	1.20	-2.10	0.32	-0.90	1.53
MUSK	1.32	-1.18	0.59	0.14	1.91
MX1	7.21	-5.62	0.29	1.58	7.49
MX2	3.85	-3.78	0.95	0.07	4.80
MYBPHL	1.41	-1.38	-0.50	0.02	0.91
NACC2	1.18	-1.62	0.96	-0.43	2.14
NDRG1	1.63	-2.29	0.65	-0.66	2.29
NDUFA4L2	1.15	-1.12	0.18	0.03	1.33
NFE2L2	1.28	-1.09	0.32	0.20	1.61
NFIL3	1.84	-1.77	0.32	0.07	2.16
NFKBIA	2.19	-3.01	0.27	-0.83	2.46
NFKBIZ	1.85	-1.54	-0.27	0.30	1.58
NRCAM	1.11	-1.11	-0.10	0.00	1.01
NTPCR	1.00	-1.09	0.35	-0.09	1.36
OAS1	4.38	-4.25	1.44	0.13	5.82
OAS1	3.60	-3.70	1.30	-0.10	4.90
OAS1	3.76	-3.81	1.29	-0.05	5.05
OAS2	2.38	-2.38	0.44	0.00	2.82
OAS2	1.56	-1.60	0.99	-0.04	2.55
OAS2	5.94	-5.52	0.39	0.42	6.33
OAS3	3.73	-3.82	0.85	-0.09	4.59
OASL	1.25	-1.12	1.63	0.13	2.88
OASL	3.54	-3.37	2.06	0.16	5.60
PAPPA	1.77	-2.04	0.39	-0.26	2.16
PARP10	1.33	-2.54	0.40	-1.21	1.73
PARP12	2.85	-3.02	0.77	-0.17	3.62
PARP14	2.58	-2.67	0.52	-0.09	3.10
PARP9	2.13	-1.98	0.43	0.15	2.56
PARP9	3.28	-2.70	0.44	0.57	3.72
PCBP3	1.04	-1.02	0.20	0.02	1.24
PDK4	1.62	-1.33	0.02	0.29	1.64
PDPN	1.45	-1.54	-0.28	-0.09	1.17
PHF11	1.57	-2.23	0.65	-0.66	2.22
PHF11	1.49	-2.46	0.51	-0.98	2.00
PHLDA3	1.18	-1.52	0.20	-0.34	1.38
PLA2G4C	1.01	-1.30	0.07	-0.29	1.09
PLEKHA4	1.64	-1.40	1.32	0.24	2.95
PLXNB1	1.02	-1.37	-0.10	-0.34	0.93
PPAP2A	1.21	-1.08	0.55	0.13	1.75
PRIC285	3.01	-2.56	0.85	0.45	3.87

Table S4.8 continued

PSMB8	1.74	-2.13	0.71	-0.39	2.45
PSMB8	1.87	-2.20	0.65	-0.33	2.51
PSMB8	2.08	-2.27	0.55	-0.18	2.64
PSMB9	2.43	-2.97	0.94	-0.53	3.37
PSME1	1.49	-1.16	0.27	0.33	1.76
PSME2	1.33	-1.47	0.45	-0.13	1.78
PSTPIP2	1.09	-1.18	0.44	-0.09	1.53
PTGER2	1.00	-1.18	-0.20	-0.18	0.80
PTGES	1.15	-1.41	0.29	-0.26	1.44
PTGFR	1.45	-1.00	0.14	0.45	1.59
PTGFR	1.60	-1.17	0.21	0.43	1.81
PTX3	2.26	-2.44	0.92	-0.18	3.19
PYGB	1.11	-2.02	-0.12	-0.91	0.99
RARRES3	2.25	-2.05	1.25	0.20	3.51
RBCK1	1.15	-1.37	0.36	-0.22	1.51
RBM43	1.29	-1.36	0.31	-0.07	1.59
RCAN1	1.36	-2.04	0.59	-0.69	1.95
RELB	1.33	-1.20	0.27	0.13	1.59
RNF213	1.03	-1.74	-0.29	-0.72	0.74
RSAD2	2.05	-2.11	2.33	-0.06	4.38
RTN1	1.42	-1.19	-0.05	0.23	1.37
RTP4	1.88	-1.97	1.17	-0.09	3.05
S1PR3	2.56	-2.03	1.12	0.53	3.68
SAA1	3.85	-3.82	0.26	0.03	4.11
SAA1	1.90	-2.09	0.28	-0.19	2.18
SAA2	1.04	-1.02	0.25	0.02	1.29
SAMD9	2.93	-2.79	1.40	0.13	4.32
SAMD9L	2.37	-2.40	1.25	-0.04	3.62
SCHIP1	1.18	-1.99	0.06	-0.81	1.24
SEMA4B	1.09	-1.75	0.15	-0.66	1.23
SERPINA3	4.12	-2.23	-0.29	1.89	3.83
SERPINE2	1.44	-1.37	0.20	0.07	1.65
SERPING1	1.69	-1.46	1.17	0.23	2.87
SESN1	1.94	-1.54	-0.11	0.40	1.83
SHISA5	1.98	-1.76	-0.28	0.22	1.69
SLC15A3	3.13	-3.64	0.78	-0.51	3.91
SLC22A18	1.21	-1.44	-0.05	-0.23	1.16
SLC2A5	1.72	-1.82	0.80	-0.10	2.52
SLC39A14	1.10	-1.75	0.47	-0.65	1.57
SLC39A8	1.24	-1.10	0.86	0.15	2.11

Table S4.8 continued

SLC44A1	1.71	-1.10	0.78	0.61	2.49
SLC7A11	1.17	-1.14	1.26	0.03	2.43
SLC7A2	2.26	-2.39	-0.11	-0.13	2.15
SNAI2	1.34	-2.05	0.21	-0.71	1.56
SOD2	5.21	-4.54	1.54	0.67	6.75
SOD2	5.00	-3.78	0.35	1.22	5.34
SOD2	1.60	-1.59	-0.06	0.01	1.54
SP100	2.10	-2.02	1.19	0.08	3.29
SP100	1.51	-1.64	1.18	-0.13	2.68
SP110	2.52	-2.85	1.14	-0.33	3.66
SP110	1.72	-1.88	1.06	-0.16	2.78
SP110	2.69	-2.89	1.10	-0.20	3.79
SP110	1.31	-1.49	0.74	-0.18	2.05
SPATA18	1.94	-1.86	-0.21	0.08	1.73
SPTLC3	1.19	-1.30	0.23	-0.11	1.42
SPTLC3	1.18	-1.41	-0.38	-0.23	0.80
SRGN	1.08	-1.88	0.20	-0.80	1.28
SRGN	1.24	-1.68	0.21	-0.44	1.45
SSBP2	1.01	-1.26	0.35	-0.26	1.36
SSH2	1.20	-1.25	-0.31	-0.05	0.89
STAT1	3.95	-2.87	0.84	1.08	4.79
STAT1	4.00	-2.85	0.47	1.15	4.47
STAT1	2.91	-2.59	-0.09	0.32	2.81
STAT2	1.92	-2.69	0.19	-0.77	2.11
STC1	1.03	-1.46	0.79	-0.43	1.82
STOM	2.17	-2.27	0.64	-0.10	2.80
STXBP6	1.69	-2.21	-0.14	-0.52	1.54
STXBP6	1.36	-1.55	-1.01	-0.19	0.34
SUSD1	1.17	-1.22	0.50	-0.05	1.67
SUSD2	2.00	-1.30	-1.84	0.69	0.15
TAP1	2.55	-3.25	0.70	-0.70	3.26
TAP2	1.14	-1.10	1.30	0.04	2.44
TAP2	1.50	-1.57	-0.06	-0.07	1.43
TAPBP	1.78	-1.81	0.47	-0.04	2.25
TCEA3	1.19	-1.69	-0.12	-0.50	1.07
TGFBR3	1.64	-1.21	0.44	0.43	2.08
TLR3	1.00	-1.07	1.56	-0.06	2.56
TMEM140	2.61	-2.64	0.99	-0.03	3.59
TMEM179B	1.06	-1.06	0.49	0.00	1.55
TNFAIP2	1.01	-1.43	0.06	-0.42	1.06

Table S4.8 continued

TNFAIP3	2.22	-2.60	0.27	-0.38	2.50
TNFAIP6	3.80	-2.18	1.30	1.63	5.10
TNFRSF14	1.21	-2.57	0.57	-1.36	1.78
TNFRSF6B	1.31	-1.60	0.01	-0.29	1.32
TNFRSF6B	1.05	-1.22	0.04	-0.18	1.08
TNFRSF6B	1.61	-1.78	-0.02	-0.17	1.59
TNFSF13B	1.65	-1.53	2.13	0.12	3.78
TNFSF13B	1.50	-1.43	2.15	0.07	3.65
TP53I3	1.22	-1.94	0.19	-0.72	1.41
TP53INP1	1.37	-1.11	-0.56	0.26	0.81
TRIM21	1.94	-1.32	0.97	0.63	2.91
TRIM22	2.24	-2.50	-0.02	-0.26	2.22
TRIM25	2.18	-2.45	0.51	-0.27	2.69
TRIM5	1.36	-1.32	1.33	0.04	2.69
TRIM55	1.01	-1.05	0.20	-0.04	1.22
TSC22D3	1.68	-1.64	0.20	0.03	1.87
TSC22D3	2.37	-1.91	0.28	0.45	2.64
TSC22D3	2.25	-1.52	0.18	0.73	2.43
TTC39B	1.18	-1.15	1.20	0.03	2.37
TYMP	3.59	-4.25	0.62	-0.66	4.21
TYMP	1.81	-1.83	0.99	-0.02	2.80
TYMP	2.13	-2.19	1.23	-0.06	3.36
TYMP	1.90	-1.94	1.00	-0.04	2.89
UBA7	1.72	-2.41	0.57	-0.69	2.29
UBE2L6	2.26	-3.09	0.57	-0.83	2.84
UBE2L6	1.96	-2.11	0.94	-0.15	2.90
UGP2	1.15	-1.13	0.11	0.02	1.26
UNC93B1	1.23	-1.87	0.94	-0.64	2.17
USP18	2.32	-2.23	0.97	0.09	3.29
VCAM1	2.68	-2.79	0.63	-0.11	3.31
VWCE	2.84	-3.05	-0.12	-0.22	2.72
XAF1	2.66	-3.49	0.61	-0.83	3.27
XAF1	1.36	-1.76	0.79	-0.40	2.15
XPC	1.95	-2.43	0.26	-0.48	2.21
YPEL3	1.25	-1.24	-0.27	0.01	0.98
ZBTB16	1.97	-1.86	0.48	0.11	2.45
ZBTB16	1.99	-2.20	0.41	-0.20	2.40
ZC3H12A	1.66	-1.37	0.52	0.29	2.18
ZFHX3	1.11	-1.70	0.50	-0.59	1.61
ZNFX1	1.75	-2.10	0.81	-0.35	2.56

GFP v. NoLenti (Fc) = log2 fold change lenti-GFP versus uninfected
Full_v. GFP (Fc) = log2 fold change lenti-DUX4-fl versus lenti-GFP
Short_v. GFP (Fc) = log2 fold change lenti-DUX4-s versus lenti-GFP
Full v._NoLenti (Fc) = log2 fold change lenti-DUX4-fl versus uninfected
Short_vs. NoLenti (Fc) = log2 fold change lenti-DUX4-s versus uninfected

Table S4.9. DEFB103 suppresses the induction of skeletal muscle differentiation genes

Symbol	log_2 FC defMT/MT	log_2 FC MT/MB
ACTA1	-2.65	5.42
MYH8	-2.62	4.58
MYH3	-2.20	6.93
CASQ2	-2.00	3.49
CKM	-1.90	3.28
MYL4	-1.86	4.59
SMPX	-1.76	2.92
MYH7	-1.73	2.11
CACNG1	-1.58	2.51
TNNT3	-1.55	2.85
MYLPF	-1.53	6.54
TNNT3	-1.52	2.89
ENO3	-1.51	3.38
MYBPH	-1.49	5.13
TNNC2	-1.49	2.29
LOC389827	-1.48	2.30
ENO3	-1.47	2.51
TNNC1	-1.45	5.92
TPM2	-1.41	3.53
HRC	-1.40	3.31
LOC389827	-1.38	2.20
HES6	-1.24	3.80
VASH2	-1.22	1.78
MYOM1	-1.20	1.70
MYL1	-1.18	5.71
AIF1L	-1.16	1.95
CKB	-1.16	2.98
CTGF	-1.12	0.82
MYL1	-1.11	5.58
HBEGF	-1.10	2.07
PRAGMIN	-1.10	1.38
FOLR1	-1.10	1.45
ZFP106	-1.09	1.55
MYL4	-1.08	1.98
SMYD1	-1.07	2.42
ARPP-21	-1.06	2.86
CYP2J2	-1.05	1.39
ATP2A2	-1.04	2.48
HFE2	-1.04	2.89

Table S4.9 continued

RASSF4	-1.03	2.49
IL32	-1.03	1.85
FOLR1	-1.03	1.38
LMCD1	-1.03	1.24
MYL6B	-1.00	1.59
TNNT1	-1.00	2.27
NDRG1	1.01	-0.09
ANGPTL4	1.21	-0.76
DKK1	1.22	-3.56
MME	1.27	-1.36
AKR1C2	1.41	2.32
PLIN2	1.61	-0.76
MT1X	1.65	-1.50
HMOX1	1.66	-0.15
PLIN2	1.79	-0.93

log2 FC DefMT/MT is the log 2 ratio of expression in DEF103 treated muscle cells compared to control muscle cells
log2 FC MT/MB is the log2 ratio of expression in differentiated muscle cells to myoblasts

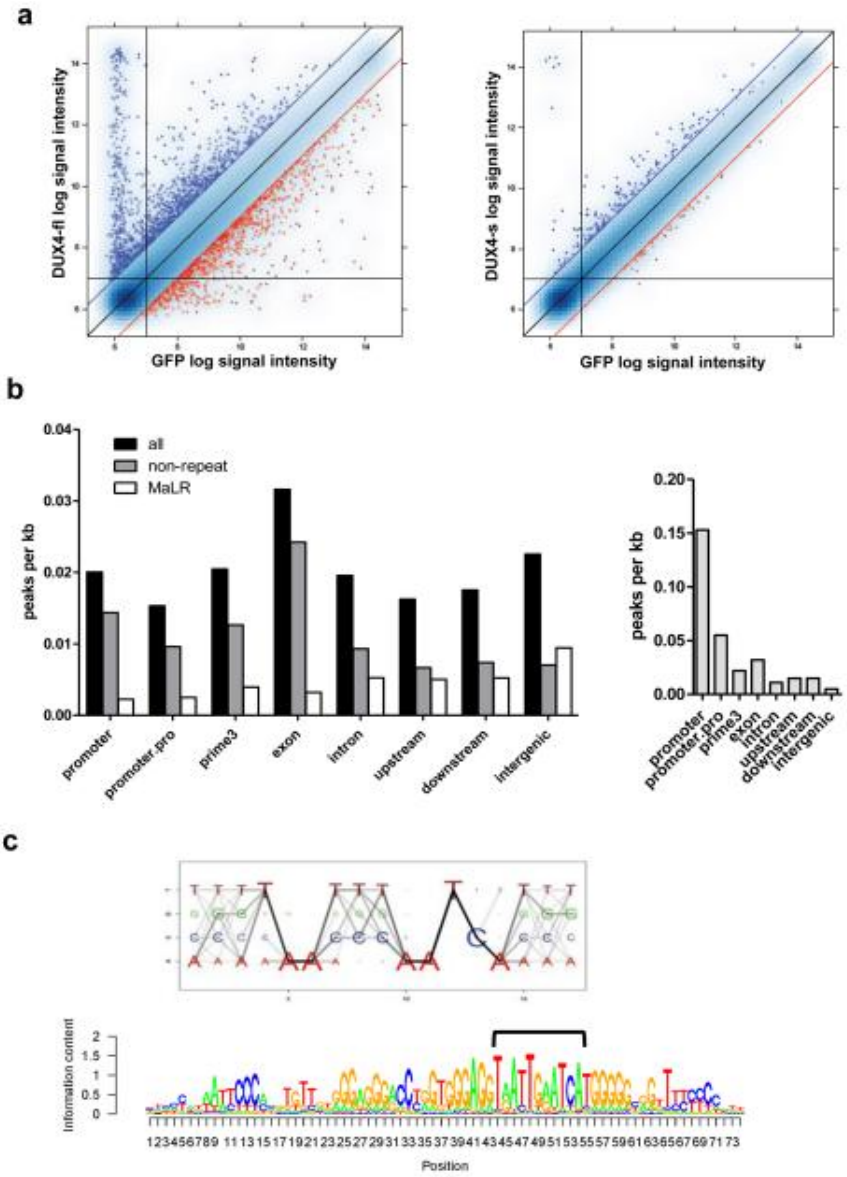


Figure 4.1. DUX4-fl activates the expression of germline genes and binds a double-homeobox motif. (a) Pairwise comparison of normalized array data from DUX4-fl full vs GFP (left), and DUX4-s vs GFP (right). Blue dots: upregulated genes; red dots: downregulated genes; blue and red diagonal lines represent 2-fold change. Vertical and horizontal lines represent signal thresholds for calling genes present or absent. (b) Gene context of DUX4-fl binding sites (left) or MyoD binding sites (right) represented by peak density, adjusting for prevalence of gene context category in the genome. Promoter: +/-500 bp from the transcription start site (TSS); promoter.pro: +/-2 kb from the TSS; prime3: +/-500 nt from the end of the transcript; upstream: -2 kb to -10 kb upstream of the TSS; downstream: +2 kb to +10 kb from the end of the transcript; intergenic: >10 kb from any annotated gene. (c) Top: DUX4-fl motif Logo. The size of each nucleotide at a given position is proportional to the frequency of the nucleotide at that position, and the darkness of the line connecting two adjacent nucleotides represents corresponding dinucleotide frequency. Bottom: DUX4 binding motif matches MaLR repeat consensus sequence. We identified the best DUX4 binding sites (bracket) within the MaLR repeats annotated in the RepeatMasker track provided by the UCSC genome browser (hg18) and extended the motif in the flanking regions to reflect general MaLR repeat consensus.

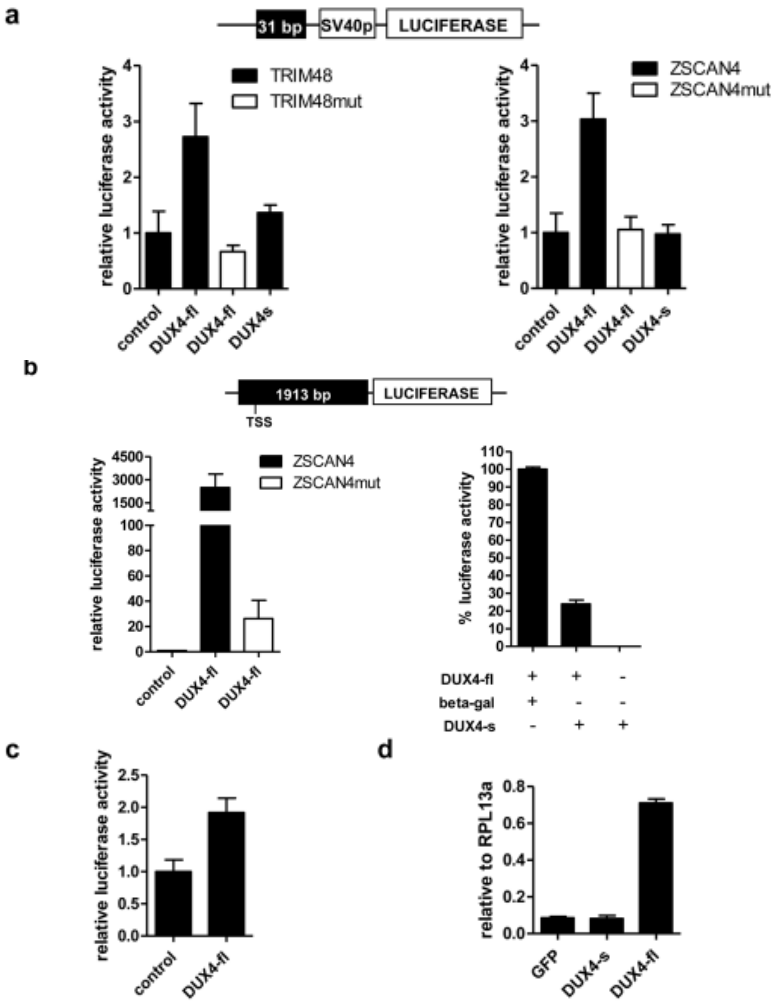


Figure 4.2. DUX4-fl activates transcription in vivo and DUX4-s can interfere with its activity. (a) Genomic fragments near theTRIM48 and ZSCAN4 genes containing DUX4 binding sites were cloned into pGL3-promoter reporter vector (schematic, top) and transfected into human rhabdomyoscaroma cell line RD. Cells were co-transfected with DUX4-fl or DUX4-s. pCS2-β galactosidase (beta gal) was used to balance DNA amount in control condition. TRIM48mut and ZSCAN4mut, mutated binding sites. Luciferase activity is set relative to control. (b) Genomic fragment upstream of the ZSCAN4 translation start site containing four DUX4 binding sites was cloned into pGL3-basic luciferase vector (schematic, top). DUX4-fl highly activates the luciferase expression, whereas mutation of the binding sites (ZSCAN4mut) drastically reduces this induction (left). Co-transfection of equal amounts of DUX4-fl and DUX4-s diminishes luciferase activity (right). Luciferase activity from DUX4-fl co-transfected with equal amount of beta gal is set at 100%. (c) Genomic fragment from the LTR of THE1D MaLR element containing the DUX4 binding site were cloned into pGL3-promoter vector and tested for response to DUX4-fl as in (a). (d) Transcripts from endogenous retroelement MaLRs are upregulated by lentiviral transduction of DUX4-fl into primary human myoblasts. No upregulation is seen with lentiviral transduction of GFP or DUX4-s. Real-time RT-PCR quantitation is reported relative to internal standard RPL13a. All data represent mean +/- SD from at least triplicates.

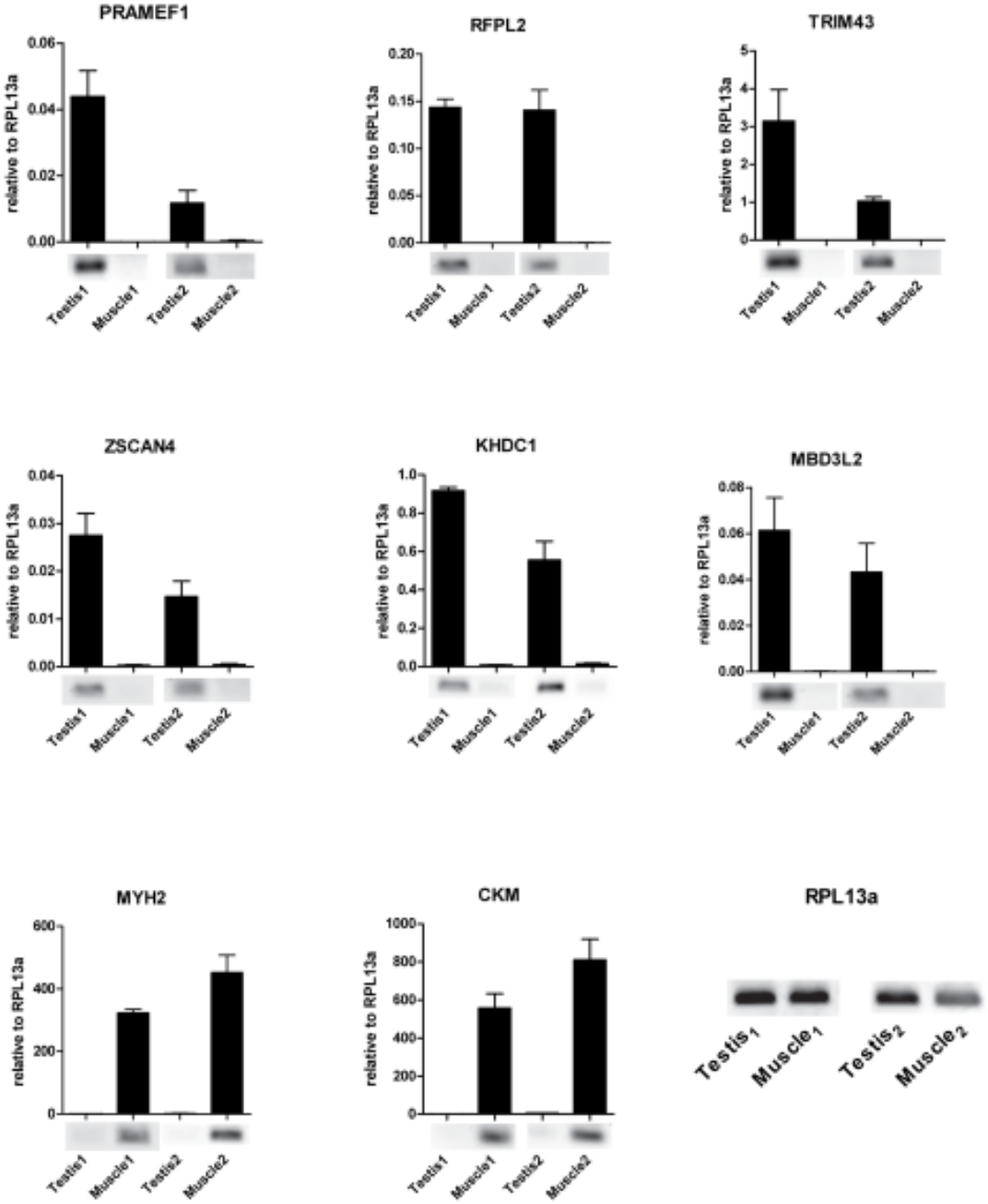


Figure 4.3. DUX4 targets are normally expressed in human testis but not in healthy skeletal muscle. DUX4-fl targets expression in human testis versus matched skeletal muscle tissue from two healthy donors. Real-time RT-qPCR analysis of gene expression is presented relative to internal standard RPL13a and error bars represent standard deviation of PCR triplicates. MYH2 and CKM are markers of skeletal muscle. Qualitative RT-PCR gel panels are shown below qPCR graphs.

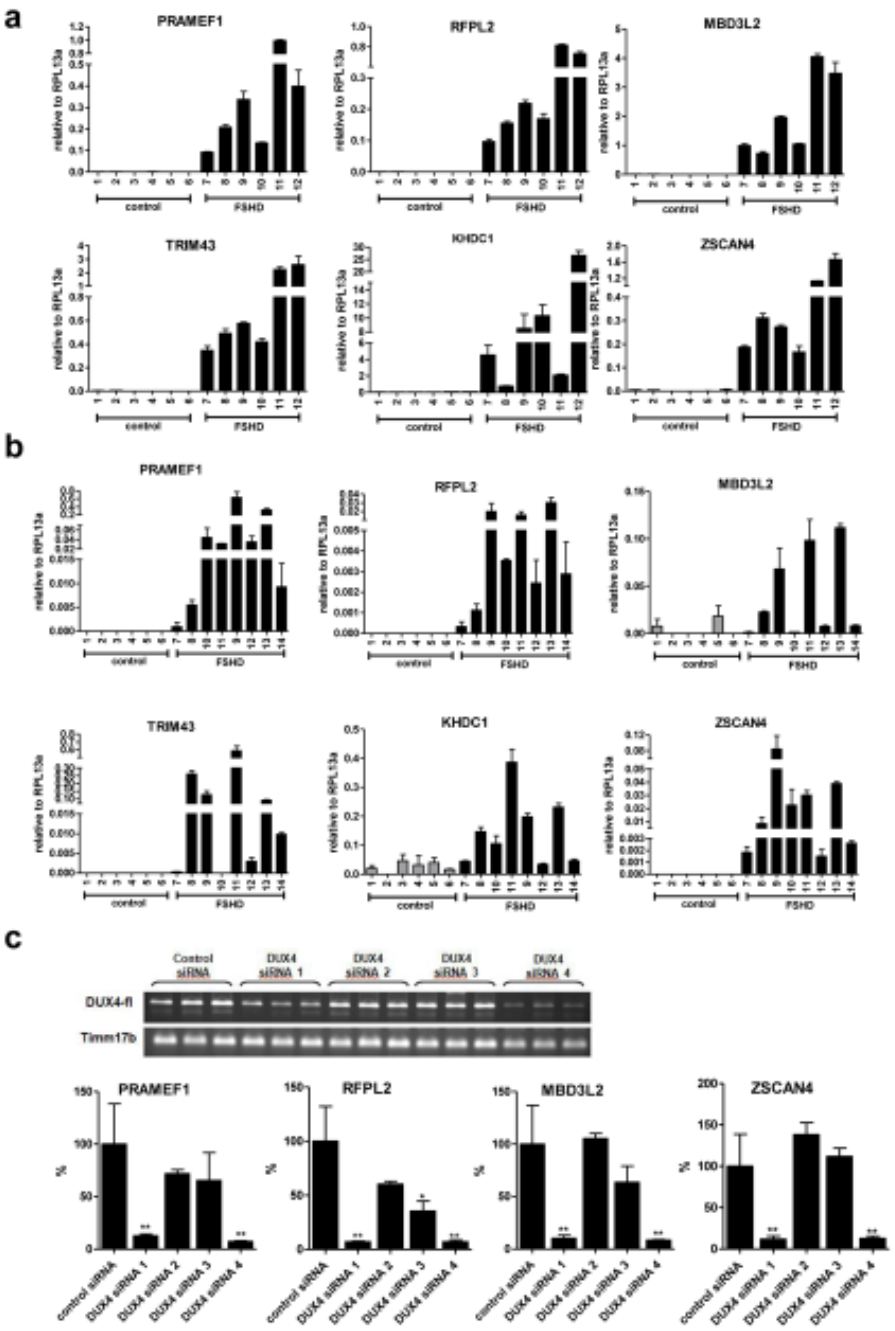


Figure 4.4. DUX4 regulated genes normally expressed in the testis are aberrantly expressed in FSHD muscle. Real-time RT-PCR quantitation of six DUX4-fl target genes, PRAMEF1, RFPL2, TRIM43, ZSCAN4, KHDC1 and MBD3L2. Values are expressed as relative to internal standard RPL13a and represent mean +/- SD from triplicates. See Supplementary Table 7 for all sample names and endogenous DUX4-fl expression status. (a) Cultured control and FSHD muscle cells. (b) Control and FSHD muscle biopsies. (c) Top: RT-PCR gel showing siRNA knockdown of endogenous DUX4-fl in cultured FSHD muscle cells, done in triplicate with Timm17b as an internal standard. Negative control siRNA is against unrelated luciferase gene. Bottom: Levels of DUX4-fl target genes relative to the control treated samples were also reduced, as measured by qPCR. Error bars represent standard deviation of triplicates; *P<0.05, **P<0.01 between DUX4 siRNA and control siRNA treated cells.

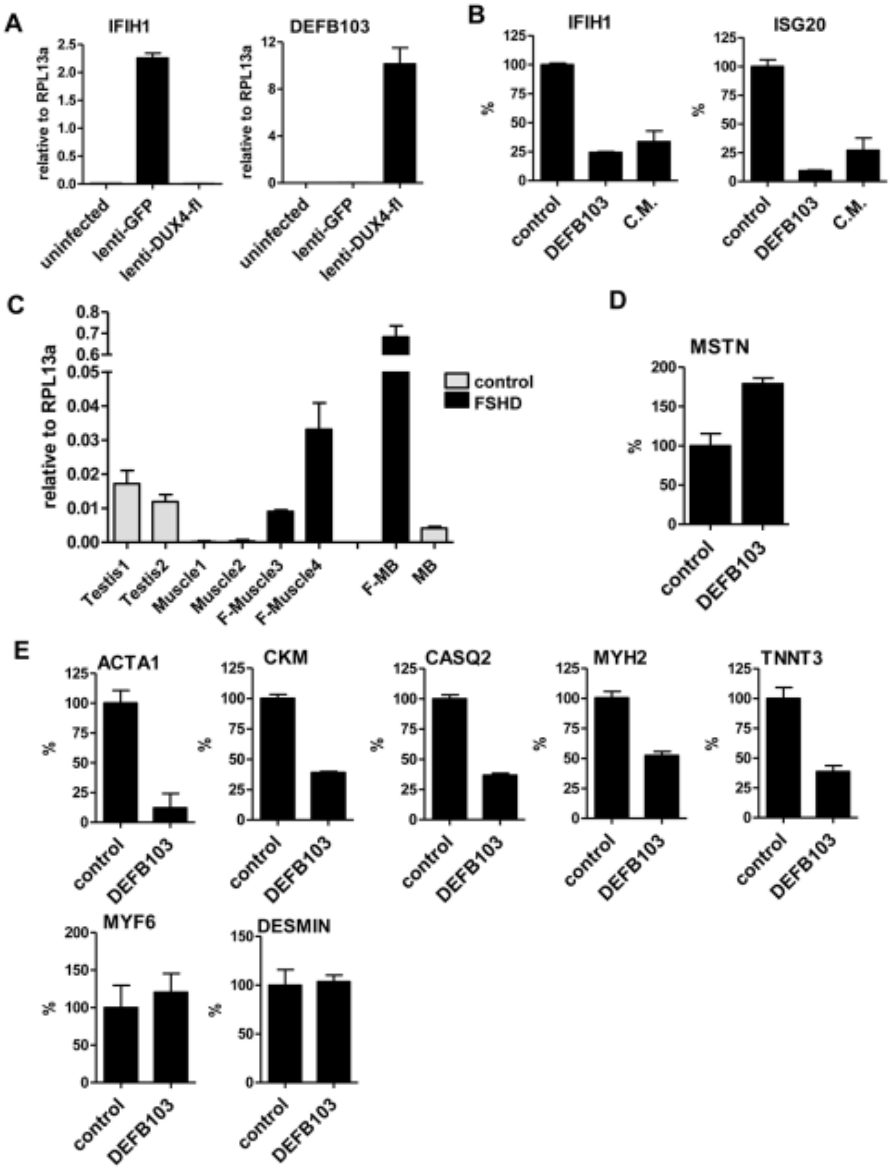


Figure 4.5. DEFB103 inhibits innate immune response to viral infection and inhibits muscle differentiation. Real time RT-PCR quantitation of innate immune responsive genes and genes involved in muscle differentiation. Values represent mean +/- SD from triplicates and are either expressed as relative to internal standard RPL13a or as percentage relative to control condition after being normalized to RPL13a. (a) Expression levels of innate immune responder IFIH1 and secreted factor DEFB103 after infection with lenti-GFP or lenti-DUX4-fl. (b) Expression levels of innate immune responders IFIH1 and ISG20 after infection with lenti-GFP in either media supplemented with human β-defensin 3 peptide or conditioned media from lenti-DUX4-fl. (c) Endogenous expression of DEFB103 in control testis and skeletal muscle tissues, FSHD muscle biopsies and cultured FSHD and control muscle cells. (d) Upregulation of myostatin (MSTN) in β-defensin 3-treated myoblasts cultured in growth media. (e) Expression levels of various muscle marker genes in response to human β-defensin when added to myoblasts cultured in differentiation media. Myf6 and Desmin were included as genes that were unchanged on the arrays.

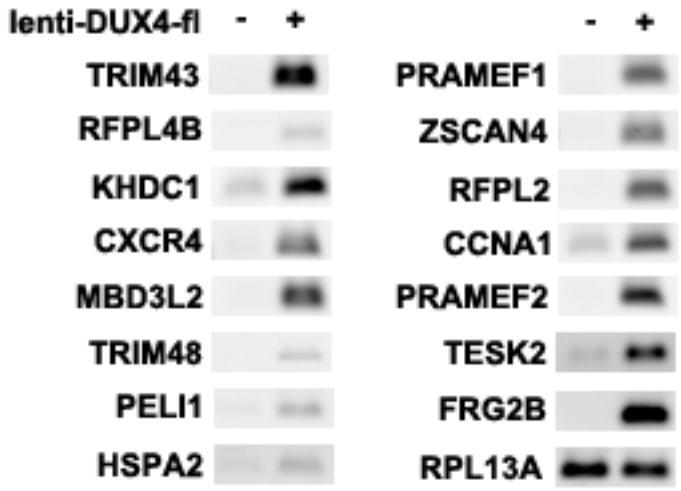


Figure S4.1. RT-PCR validation of DUX4-fl target genes from expression microarray. RNA was collected from cultured control skeletal muscle either transduced with a DUX4-fl expressing lentivirus (+) or not transduced (-). RPL13A was used as an internal standard.

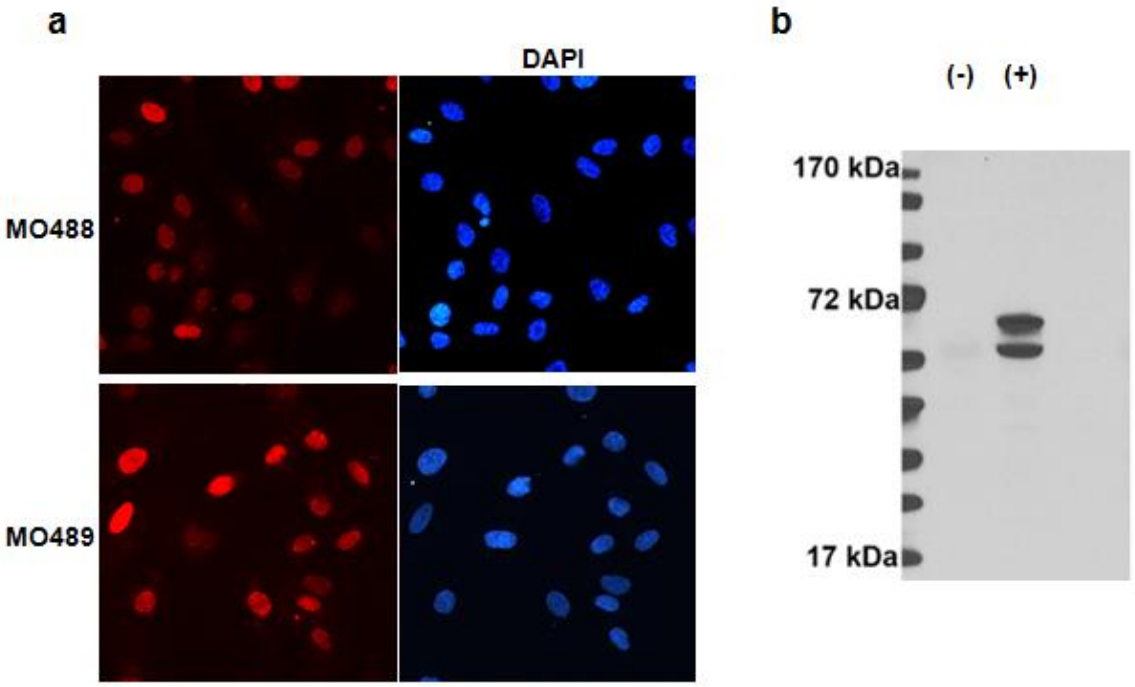


Figure S4.2. Antibody characterization. (a) MO488 and MO489 are custom rabbit polyclonal antibodies are against the DUX4 C-terminus. Human myoblasts were transduced with lenti-DUX4-fl and fixed for immunofluorescence 24hr later. Nuclei were counterstained with DAPI. Non-transduced human myoblasts did not have any nuclear staining with either antisera (data not shown) (b) Immunoprecipitation with both rabbit antibodies combined and probed on western blot with mouse anti-DUX4 antibody P4H2. (-) myoblast lysates from cells not transduced with DUX4-fl; (+) myoblast lysates expressing DUX4-fl. Doublet band is due to expression construct containing additional upstream translation start site.

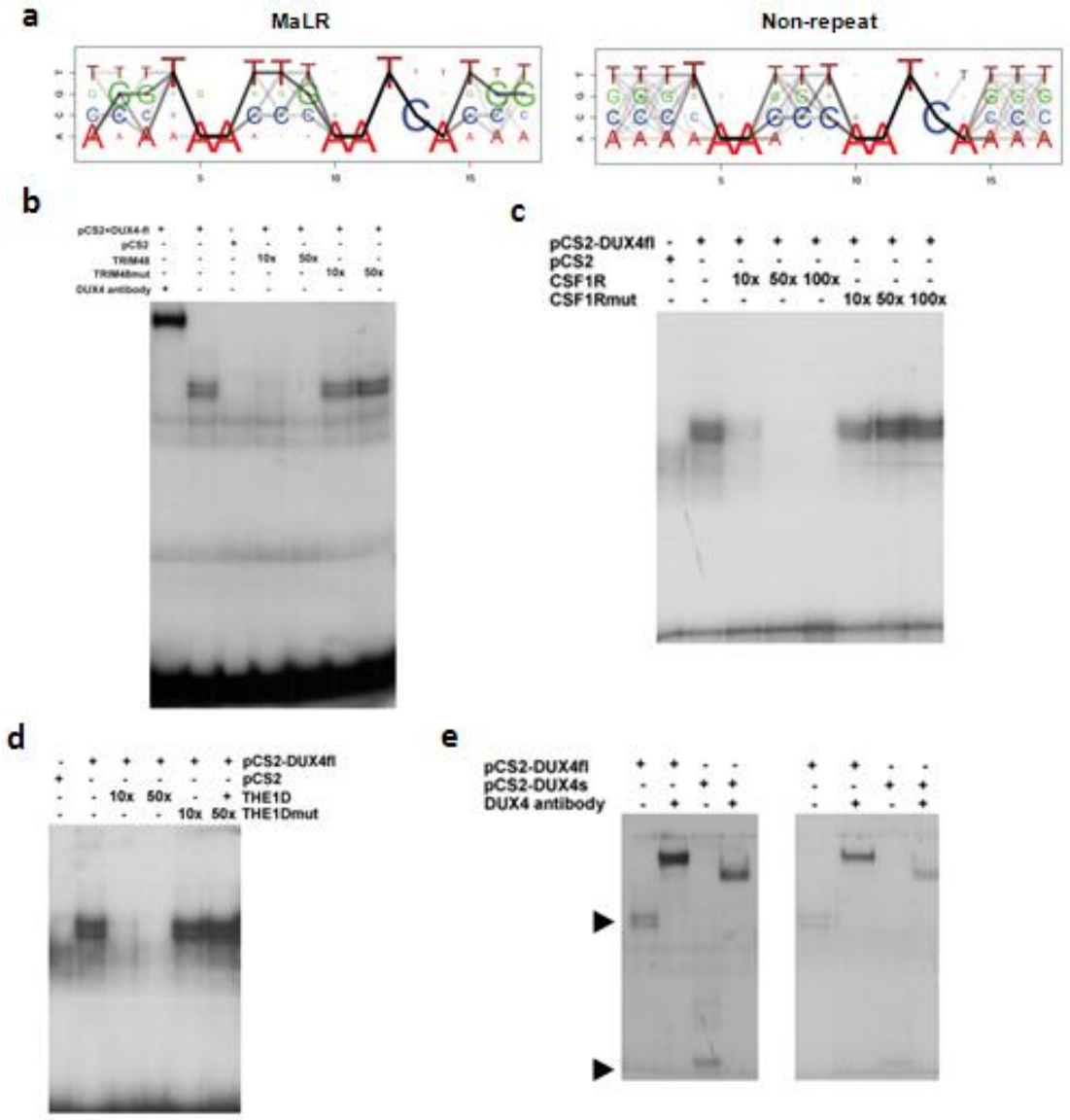


Figure S4.3. DUX4 binding in repeat and non-repeat regions and EMSA validation of DUX4 binding to ChIP-seq determined motifs. (a) DUX4-fl motifs for MaLR-associated binding sites or non-repeat associated binding sites. (b) in vitro translated DUX4-fl binds to radiolabeled oligos containing the TAATTTAATCA core sequence found near the TRIM48 gene. Competition with cold TRIM48 oligos reduces binding whereas competition with cold TRIM48mut oligos, containing the mutated core sequence TACTTTTATGA, does not. Supershift with DUX4 antibody E14-3 confirmed the specificity of binding. (c,d) In vitro translated DUX4-fl binds to radiolabeled oligos containing the TAATTGAATCA core sequence found within the LTR of a THE1B retroelement near the CSF1R gene (c) or to oligos containing the TAATCCAATCA core sequence found within the LTR of the THE1D retroelement (d). Competition with the cold CSF1R and THE1D probes to their respective radioactive oligos inhibited binding, whereas competition with cold mutant CSF1Rmut and THE1Dmut oligos, containing sites TACTTCTATG and TACTCCTATGA, respectively, do not. (e) DUX4-s also binds the same core motifs; left, CSF1R; right, TRIM48. Supershift with anti-DUX4 N-terminus antibody E14-3 confirmed specificity of binding.

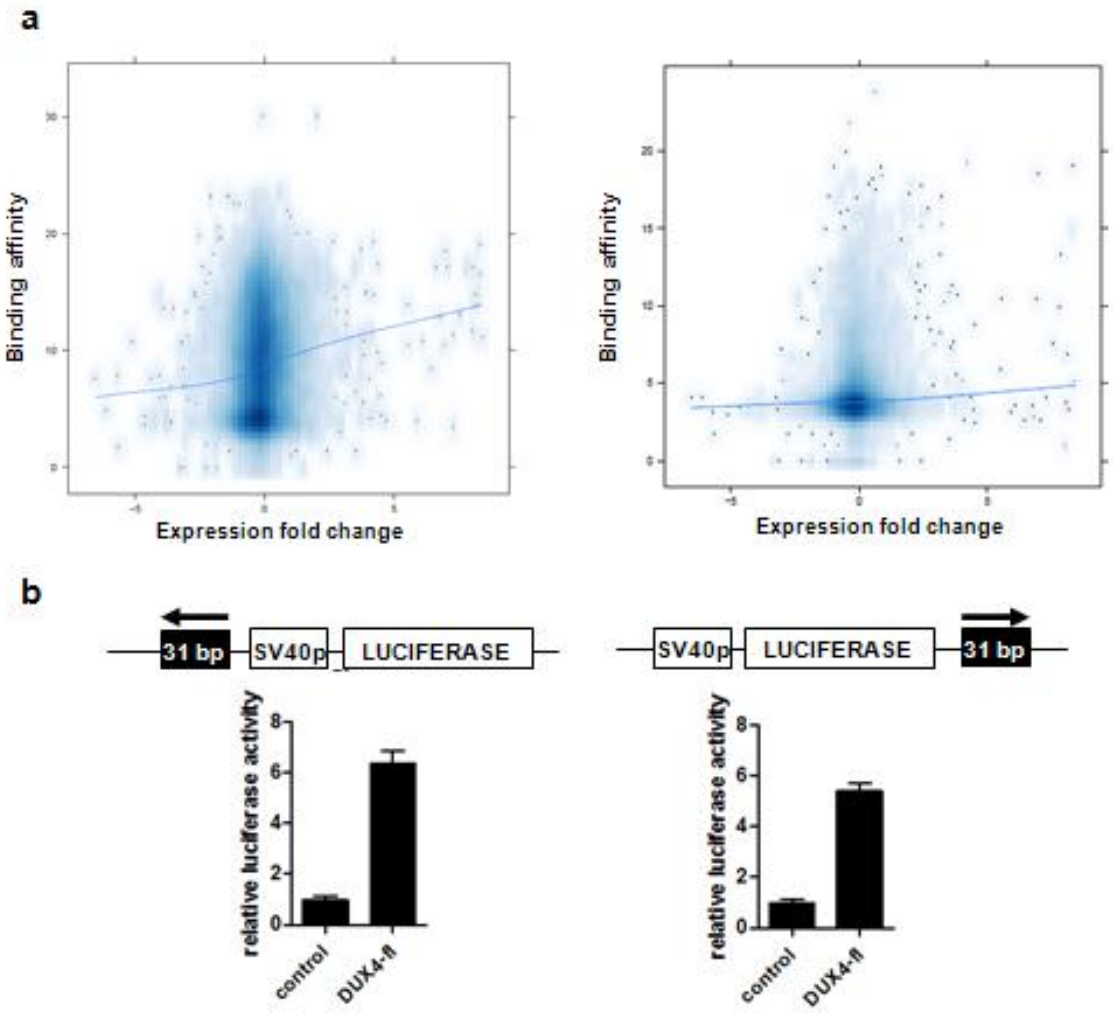


Figure S4.4. Global DUX4-fl binding is moderately associated with the expression of its targets, but DUX4-fl can act as an enhancer at certain loci. (a) Fold change on the x-axis and binding affinity as defined by square root of peak height on the y-axis with trend line drawn showing positive trend. Left: binding sites in gene-regulatory regions defined by those falling within CTCF domains. Right: binding sites in gene-regulatory regions defined by those falling within +/-2kb of TSS. (b) DUX4-fl can act as an enhancer. Left: DUX4 binding site in reverse orientation upstream of the SV40 promoter. Right: DUX4 binding site in original orientation but moved downstream of the reporter gene. Luciferase activity set relative to control plasmid condition and error bars represent standard deviation of triplicates.

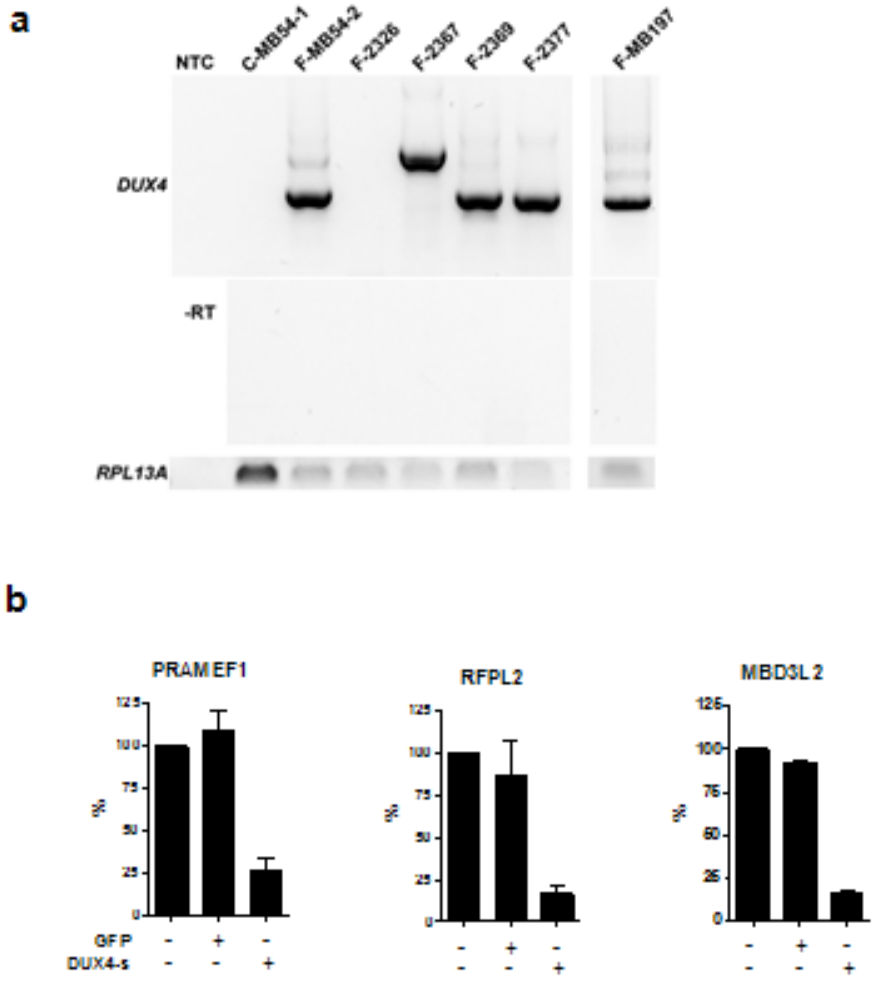


Figure S4.5. DUX4-fl expression status in muscle samples and inhibition with dominant negative DUX4-s. (a) Nested DUX4-fl PCR on cDNA from cultured muscle cells or biopsies. RPL13A PCR was used for an internal standard. For coded sample names and complete status information, see Supplementary Table 7. Remaining samples were previously characterized for DUX4 expression1. (b) DUX4-s blocks expression of DUX4-fl target genes in FSHD muscle cells. Real-time RT-PCR quantitation of three target genes, PRAMEF1, RFPL2 and MBD3L2 in FSHD cultured muscle cells transduced with lenti-GFP or lenti-DUX4-s or untransduced. Abundance of targets was calculated relative to internal standard RPL13a and then set as percentages relative to the untransduced condition. Values represent mean +/- SEM from three independent experiments.

Chapter 5: Discussion

The goal of my thesis was to understand the role of the double homeobox transcription factor DUX4 in the disease facioscapulohumeral muscular dystrophy and its possible role in normal human cell biology. I have developed multiple monoclonal and polyclonal antibodies that recognize and distinguish the different isoforms of DUX4 and can be used as tools to assess expression and activity of DUX4 protein. Using these antibodies in combination with small-pool RT-PCR, I found that the very low level of endogenous DUX4 expression in a population of FSHD muscle cells represented a relatively high expression of DUX4 mRNA and protein in a subset of cells at any given point in time.

I demonstrated that while the full length DUX4 induces toxicity in muscle cells, the shorter isoform without the C-terminal end, DUX4-s, exhibited no obvious detrimental effects on muscle cells. Through ChIP-seq studies we showed that DUX4 protein binds to a DNA target sequence that contains two closely-spaced canonical homeodomain binding sequences TAAT in tandem. These binding sites are present in both unique regions of the genome as well as within LTRs of a family of endogenous retrotransposons called MaLR. Using EMSA, we demonstrated that DUX4-s can bind to the same sequences, but, interestingly, DUX4-s was unable to activate transcription in luciferase reporter assays and, in fact, could act as a dominant negative to inhibit full length DUX4's activity.

Using expression microarrays, we showed that DUX4 can upregulate over a thousand genes involved in multiple processes, such as germ cell development, RNA splicing and immune modulation. I found that these targets are indeed upregulated in cultured FSHD muscle cells as well as biopsies, providing further support for the causal role of DUX4 in FSHD. The repertoire of DUX4 targets raises interesting questions about not only the pathophysiological but also physiological function of DUX4.

DUX4 and germ cell development

DUX4 is a retrogene, and retrogenes that are passed on to progeny must have been created in the germ cells, indicating that the ancestral gene was expressed in the germ cells and perhaps had a functional role there. Most retrotransposed sequences

accumulate mutations and become defunct pseudogenes, though some survive to retain the function of the parental gene or go on to acquire new functions. Prior to our work, DUX4 was mostly regarded as a relic pseudogene with no known purpose other than to perhaps cause a debilitating human disease when insufficiently repressed. The only hints that DUX4, or any double homeobox gene for that matter, may have a role in normal biology was that the coding sequence for DUX4 was conserved (Clapp et al., 2007) and that mouse DUXs may be involved in germ cell and T-cell development (Kawazu et al., 2007; Wu et al., 2010).

We found that DUX4 is normally present in the human testis and absent from somatic tissues. This led us to hypothesize that DUX4 performs a physiological function in the testis and when inappropriately expressed in somatic tissues like skeletal muscle, may lead to a pathology such as FSHD. Many of the genes most highly induced by DUX4 when we ectopically expressed it in human muscle cells were normally expressed in the testis and involved in gametogenesis and early development. For example, CCNA1 is a cyclin that primarily functions in the control of the germline meiotic cell cycle and highly expressed in pachytene spermatocytes (Muller-Tidow et al., 2003). HSPA2 is a testis-specific member of the heat-shock protein 70 family and a transition chaperone that associates with spermatid DNA-packaging proteins thought to be important for the completion of meiosis during spermatogenesis (Govin et al., 2006; Quenet et al., 2009). Absence or downregulation of these DUX4 targets have been correlated with subtypes of male infertility (Cedenho et al., 2006; Haraguchi et al., 2009; Schrader et al., 2002a; Schrader et al., 2002b). Inappropriate activation of these meiotic and spermiogenic genes in post-mitotic skeletal muscle as seen in FSHD may lead to cell stress and apoptosis.

DUX4 also induces a number of cancer-testis antigens (CTAs). Because the testis is an immune-privileged site, many testis-specific gene products are not recognized as "self" and thus are targeted by the immune system when misexpressed elsewhere, such as on tumor cells (Simpson et al., 2005). For example, the members of the PRAME gene family are expressed in human melanomas and recognized by cytolytic T lymphocytes and not normally expressed in most somatic tissues (Santamaria et al., 2008). CSAG3

(also known as TRAG-3 or CSAG2) is another cancer-testis antigen found in a variety of tumors including chondrosarcomas, melanomas and breast cancers (Janjic et al., 2006). An inflammatory component to FSHD has been recognized for many years, and, more specifically, it was shown recently that FSHD skeletal muscle contained focal infiltrates of CD4+ and CD8+ T-cells (Frisullo et al., 2011). It is reasonable to hypothesize, then, that these T-cells may be targeting muscle cells presenting antigenic peptides from one or more of these DUX4-induced CTAs. Furthermore, it is not known what factor(s) activate(s) testis genes in tumor cells. Since DUX4 has been found to be endogenously expressed in tumor lines (Kowaljow et al., 2007), it is interesting to speculate that perhaps DUX4 is responsible for the induction of CTAs in certain cancers.

Thus, DUX4 may effect pathogenesis in FSHD through its induction of testis genes by either directly causing cellular stress from reactivation of the gametogenic program in post-mitotic cells or indirectly causing cell destruction by targeting the muscle cells for immune attack. In addition to providing a novel pathogenic model for FSHD, these testis-specific genes also make good candidates for biomarkers of FSHD since they are absent, or nearly absent, in healthy skeletal muscle but robustly upregulated in FSHD muscle. Future work will be needed to elucidate the exact function of DUX4 in normal human reproductive biology, though this will be difficult to assess by knockout studies since there are hundreds of copies DUX4 in the human genome. It would be interesting to determine whether there is a correlation between DUX4 levels and fertility. For FSHD, future work will be needed to determine the exact antigens that these T-cells are recognizing.

DUX4 and early embryonic development

As described in Chapter 3, DUX4 is expressed in induced pluripotent stem (iPS) cells made from fibroblasts of both control and FSHD individuals. However, when these iPS cells were differentiated to embryoid bodies, the control cells repressed the expression of DUX4 whereas DUX4 expression persisted in the FSHD cells. Therefore, we propose that FSHD represents a developmental disease in which insufficient

repression of the D4Z4 macrosatellite in muscle cells results in leaky expression of DUX4 and, consequently, the inappropriate induction of early stem cell genes.

When ectopically expressed in control human muscle cells, DUX4 upregulated many genes involved in early embryonic development. For example, ZSCAN4 is a specific marker of the two-cell embryo and embryonic stem cells; it is thought to be involved in telomere maintenance, upregulation of meiosis-specific homologous recombination genes and essential for preimplantation development (Falco et al., 2007; Zalzman et al., 2010). The KHDC1 gene family encode RNA-binding proteins that are expressed in oocytes and embryonic stem cells and have undergone rapid evolution since their emergence in eutherian mammalian genomes (Pierre et al., 2007). The RFPL gene family is regulated by the corticogenic transcription factor PAX6 and expressed at the onset of neurogenesis in differentiating human embryonic stem cells; RFPL genes are restricted to the subdivision of higher primates called Catarrhini and thought to contribute to the organization and size of the primate neocortex, regulate cell cycle progression and possibly even human-specific traits (Bonnefont et al., 2011; Bonnefont et al., 2008). Similar to the activation of germ cell genes, the inappropriate spatiotemporal upregulation of early stem cell genes may contribute to FSHD pathogenesis by inhibiting muscle differentiation and maturation, causing cellular distress and even leading to apoptosis.

DUX4 and RNA splicing

DUX4 upregulated numerous RNA processing factors and components of the spliceosome, such as THOC4, SFRS2B and 17A, CWC15 and TFIP11. As discussed in Chapter 4, this is particularly interesting considering that splicing abnormalities have been reported in FSHD, though previously attributed to another candidate disease gene FRG1 (Gabellini et al., 2006). In order to determine what proteins interacted with the DUX4 protein, I performed co-immunoprecipitation of DUX4 under native conditions preserving chromatin integrity followed by mass spectrometry and identified some of the same splicing factors, like THOC4 (data not shown). Future experiments using next generation RNA-sequencing will be needed to globally assess the possible splicing

alterations in the transcriptome of FSHD muscle cells as a result of DUX4. Since DUX4 might be closely associating with these RNA-processing factors, it is reasonable to wonder whether DUX4 itself might be able to bind and interact with RNA in addition to DNA. Though traditionally known as DNA-binding transcription factors, there have been reports of homeodomain proteins regulating expression through direct binding of messenger RNA (Chan and Struhl, 1997; Dubnau and Struhl, 1996). Techniques such as cross-linking immunoprecipitation and high-throughput sequencing (CLIP-seq or HITS-CLIP) of RNA can be used to map DUX4 protein-RNA binding in vivo.

DUX4 and endogenous retrotransposons

Endogenous retrotransposons make up a significant portion of the human genome, but we are just beginning to understand what regulates the expression and activity of these repetitive elements. The binding sites for a few transcription factors, including p53, ESR1, and CTCF, have been found to reside in repetitive elements (Bourque et al., 2008). p53 is a well-studied tumor suppressor and has been recently shown to have the ability to bind and activate long interspersed nuclear elements-1 (L1s) (Harris et al., 2009) as well as bind to the LTRs of class I endogenous retroviral (ERV) elements, which in turn act as enhancers for the transcription of nearby genes (Wang et al., 2007). This is very similar to what we found for DUX4, which can bind and activate the transcription of certain classes of endogenous LTR retrotransposons. DUX4 binding sites were most highly enriched in primate-specific subclasses of MaLR called the transposable human element (THE-1) family. Not much is known about the THE-1 elements, but studies have shown that other mammalian MaLRs can be reactivated during zygote formation and early embryonic development and change the landscape of the transcriptome by, for example, forming chimeras with genic transcripts or acting as alternative promoters or enhancers of neighboring genes (Macfarlan et al., 2011; Peaston et al., 2004). Future experiments such as RNA deep sequencing will be needed to determine whether these events are occurring in FSHD muscle or human testis as a result of DUX4 activity.

Another interesting question this raises is in regards to evolutionary relationship between transcription factors and repetitive elements. One possibility is that DUX4 acquired the ability to bind sequences in the repetitive elements in order to expand its regulatory network to new genes, as has been proposed for p53 (Bourque et al., 2008; Wang et al., 2007). Another possibility is that the repetitive element co-opted the binding site for a strong transcriptional activator such as DUX4 in order to increase its own activity and amplify in the genome. Both of these scenarios may be true, in which case exploring the histories of the DUX4 gene in conjunction with these retrotransposons may offer insight into the co-evolutionary balance between host and "parasite."

DUX4 and immune modulators

I found that DUX4 induces a human β-defensin DEFB103, which, when added at the time of infection, can inhibit the innate immune response to lentiviral infection. The canonical function of defensins are to neutralize microbial pathogens such as invading bacteria and viruses, so it was surprising that DEFB103 could also block the antiviral response in muscle cells. It will be interesting to determine whether this block in antiviral response by DEFB103 is generalizable to other stimuli such as DNA viruses, interferon signaling, etc. Very recently, it was shown that DEFB103 can inhibit the transcription of pro-inflammatory genes in TLR4-stimulated macrophages (Semple et al., 2011), though further investigation is needed to delineate the exact molecular mechanisms by which DEFB103 acts to suppress innate immune responses. Although purely speculative, this raises the question of whether FSHD patients might be more susceptible to viral infection of their skeletal muscle cells.

DEFB103 can also directly inhibit myogenic differentiation when added to cultured human myoblasts. Though the mechanism is unclear, one possibility is that it acts through the receptor CXCR4. DEFB103 has been shown to be an antagonist of CXCR4 (Feng et al., 2006), and CXCR4 is expressed on myoblasts and is important for muscle cell fusion and differentiation (Melchionna et al., 2010).

DUX4 and other pathways

As discussed in Chapter 4, we have demonstrated that DUX4 is a potent regulator of numerous pathways and processes, including not only the aforementioned germ cell development, RNA splicing, endogenous retrotransposons and immune mediators but also developmentally regulated components of the Pol II transcription complex and ubiquitin pathway. Given the complex and variable phenotype of FSHD seen even among affected relatives in the same family, it is likely that this is not a simple disease caused by a single mechanism, but rather a complex disease caused by a combination of dysregulated pathways. Future work will be required to determine which of these pathways is most relevant to DUX4's role in disease pathogenesis and normal cell biology.

Targeting DUX4 for FSHD therapeutics

There is currently no cure for FSHD, though efforts are under way to inhibit the activity of DUX4 and other candidate disease genes. We have shown that the shorter DUX4-s can act as a dominant negative against full length DUX4. If a safe mode of gene delivery that targets skeletal muscle is developed for clinical use, then DUX4-s might be one way to mitigate the toxic effects of DUX4. Another possibility is the use of small RNAs to target the DUX4 gene product for destruction or downregulation such as with the siRNAs we have used to knockdown endogenous DUX4 in cultured FSHD muscle cells. Safe delivery of such molecules is, again, a concern, but new advances in nanotechnology offer a promising avenues to explore (Davis, 2009). Yet another possibility is to prevent the leaky expression of DUX4 in the first place by modulating the chromatin state of the D4Z4 locus.

In conclusion, FSHD represents a disease where incomplete developmental silencing of a retrogene DUX4 results in inappropriate expression of germ cell and early stem cell genes in post-mitotic skeletal muscle. DUX4 also upregulates the expression of DEFB103, which not only suppresses the innate immune response to viral infection but also inhibits myogenic differentiation. Additionally, DUX4 reactivates the expression of endogenous retrotransposons, raising interesting questions about their co-evolutionary

relationship. Future investigation will be needed to determine which of the DUX4-regulated pathways are most relevant to FSHD pathogenesis and which are conserved functions of double homeodomain proteins in normal cell biology.

LIST OF REFERENCES

Allo, M., Buggiano, V., Fededa, J.P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S.A.*, et al.* (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. Nat Struct Mol Biol *16*, 717-724.

Ansseau, E., Laoudj-Chenivesse, D., Marcowycz, A., Tassin, A., Vanderplanck, C., Sauvage, S., Barro, M., Mahieu, I., Leroy, A., Leclercq, I.*, et al.* (2009). DUX4c is up-regulated in FSHD. It induces the MYF5 protein and human myoblast proliferation. PLoS One *4*, e7482.

Bakker, E., Wijmenga, C., Vossen, R.H., Padberg, G.W., Hewitt, J., van der Wielen, M., Rasmussen, K., and Frants, R.R. (1995). The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter. Muscle Nerve *2*, S39-44.

Banerjee-Basu, S., Sink, D.W., and Baxevanis, A.D. (2001). The Homeodomain Resource: sequences, structures, DNA binding sites and genomic information. Nucleic Acids Res *29*, 291-293.

Belancio, V.P., Roy-Engel, A.M., and Deininger, P.L. (2010). All y'all need to know 'bout retroelements in cancer. Semin Cancer Biol *20*, 200-210.

Bodega, B., Ramirez, G.D., Grasser, F., Cheli, S., Brunelli, S., Mora, M., Meneveri, R., Marozzi, A., Mueller, S., Battaglioli, E.*, et al.* (2009). 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.

Boncinelli, E. (1997). Homeobox genes and disease. Curr Opin Genet Dev *7*, 331-337.

Bonnefont, J., Laforge, T., Plastre, O., Beck, B., Sorce, S., Dehay, C., and Krause, K.H. (2011). Primate-specific RFPL1 gene controls cell-cycle progression through cyclin B1/Cdc2 degradation. Cell Death Differ *18*, 293-303.

Bonnefont, J., Nikolaev, S.I., Perrier, A.L., Guo, S., Cartier, L., Sorce, S., Laforge, T., Aubry, L., Khaitovich, P., Peschanski, M.*, et al.* (2008). Evolutionary forces shape the human RFPL1,2,3 genes toward a role in neocortex development. Am J Hum Genet *83*, 208-218.

Booth, H.A., and Holland, P.W. (2007). Annotation, nomenclature and evolution of four novel homeobox genes expressed in the human germ line. Gene *387*, 7-14.

Bosch, N., Caceres, M., Cardone, M.F., Carreras, A., Ballana, E., Rocchi, M., Armengol, L., and Estivill, X. (2007). Characterization and evolution of the novel gene family FAM90A in primates originated by multiple duplication and rearrangement events. Hum Mol Genet *16*, 2572-2582.

Bosnakovski, D., Daughters, R.S., Xu, Z., Slack, J.M., and Kyba, M. (2009). Biphasic myopathic phenotype of mouse DUX, an ORF within conserved FSHD-related repeats. PLoS One *4*, e7003.

Bosnakovski, D., Lamb, S., Simsek, T., Xu, Z., Belayew, A., Perlingeiro, R., and Kyba, M. (2008a). DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. Exp Neurol *214*, 87-96.

Bosnakovski, D., Xu, Z., Gang, E.J., Galindo, C.L., Liu, M., Simsek, T., Garner, H.R., Agha-Mohammadi, S., Tassin, A., Coppee, F.*, et al.* (2008b). An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. Embo J *27*, 2766-2779.

Bourque, G., Leong, B., Vega, V.B., Chen, X., Lee, Y.L., Srinivasan, K.G., Chew, J.L., Ruan, Y., Wei, C.L., Ng, H.H., *et al.* (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res *18*, 1752-1762.

Candille, S.I., Kaelin, C.B., Cattanach, B.M., Yu, B., Thompson, D.A., Nix, M.A., Kerns, J.A., Schmutz, S.M., Millhauser, G.L., and Barsh, G.S. (2007). A -defensin mutation causes black coat color in domestic dogs. Science *318*, 1418-1423.

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G.J., Parker, M.H., MacQuarrie, K.L., Davison, J., Morgan, M.T., Ruzzo, W.L., *et al.* (2010). Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. Dev Cell *18*, 662-674.

Cedenho, A.P., Lima, S.B., Cenedeze, M.A., Spaine, D.M., Ortiz, V., and Oehninger, S. (2006). Oligozoospermia and heat-shock protein expression in ejaculated spermatozoa. Hum Reprod *21*, 1791-1794.

Chan, S.K., and Struhl, G. (1997). Sequence-specific RNA binding by bicoid. Nature *388*, 634.

Chang, T.C., Yang, Y., Yasue, H., Bharti, A.K., Retzel, E.F., and Liu, W.S. (2011). The expansion of the PRAME gene family in Eutheria. PLoS One *6*, e16867.

Clapp, J., Mitchell, L.M., Bolland, D.J., Fantes, J., Corcoran, A.E., Scotting, P.J., Armour, J.A., and Hewitt, J.E. (2007). Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. Am J Hum Genet *81*, 264-279.

Davis, M.E. (2009). The first targeted delivery of siRNA in humans via a self-assembling, cyclodextrin polymer-based nanoparticle: from concept to clinic. Mol Pharm *6*, 659-668.

de Greef, J.C., Lemmers, R.J., Camano, P., Day, J.W., Sacconi, S., Dunand, M., van Engelen, B.G., Kiuru-Enari, S., Padberg, G.W., Rosa, A.L., *et al.* (2010). Clinical features of facioscapulohumeral muscular dystrophy 2. Neurology *75*, 1548-1554.

de Greef, J.C., Lemmers, R.J., van Engelen, B.G., Sacconi, S., Venance, S.L., Frants, R.R., Tawil, R., and van der Maarel, S.M. (2009). Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD. Hum Mutat *30*, 1449-1459.

Deidda, G., Cacurri, S., Grisanti, P., Vigneti, E., Piazzo, N., and Felicetti, L. (1995). Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. Eur J Hum Genet *3*, 155-167.

Dixit, M., Ansseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., Sauvage, S., Matteotti, C., van Acker, A.M., Leo, O., *et al.* (2007). DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proc Natl Acad Sci U S A *104*, 18157-18162.

Dmitriev, P., Lipinski, M., and Vassetzky, Y.S. (2009). Pearls in the junk: dissecting the molecular pathogenesis of facioscapulohumeral muscular dystrophy. Neuromuscul Disord *19*, 17-20.

Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics *24*, 1547-1548.

Dubnau, J., and Struhl, G. (1996). RNA recognition and translational regulation by a homeodomain protein. Nature *379*, 694-699.

Falco, G., Lee, S.L., Stanghellini, I., Bassey, U.C., Hamatani, T., and Ko, M.S. (2007). Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. Dev Biol *307*, 539-550.

Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*, 257-258.

Feng, Z., Dubyak, G.R., Lederman, M.M., and Weinberg, A. (2006). Cutting edge: human beta defensin 3--a novel antagonist of the HIV-1 coreceptor CXCR4. J Immunol *177*, 782-786.

Frisullo, G., Frusciante, R., Nociti, V., Tasca, G., Renna, R., Iorio, R., Patanella, A.K., Iannaccone, E., Marti, A., Rossi, M.*, et al.* (2011). CD8(+) T cells in facioscapulohumeral muscular dystrophy patients with inflammatory features at muscle MRI. J Clin Immunol *31*, 155-166.

Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M.A., Bottinelli, R.*, et al.* (2006). Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. Nature *439*, 973-977.

Gabellini, D., Green, M.R., and Tupler, R. (2002). Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell *110*, 339-348.

Gabriels, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D.*, et al.* (1999). Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene *236*, 25-32.

Gehring, W.J. (1992). The homeobox in perspective. Trends Biochem Sci *17*, 277-280.

Geng, L.N., Tyler, A.E., and Tapscott, S.J. (2011). Immunodetection of human double homeobox 4. Hybridoma (Larchmt) *30*, 125-130.

Govin, J., Caron, C., Escoffier, E., Ferro, M., Kuhn, L., Rousseaux, S., Eddy, E.M., Garin, J., and Khochbin, S. (2006). Post-meiotic shifts in HSPA2/HSP70.2 chaperone activity during mouse spermatogenesis. J Biol Chem *281*, 37888-37892.

Griffin, C.A., Apponi, L.H., Long, K.K., and Pavlath, G.K. (2010). Chemokine expression and control of muscle cell migration during myogenesis. J Cell Sci *123*, 3052-3060.

Haraguchi, T., Ishikawa, T., Yamaguchi, K., and Fujisawa, M. (2009). Cyclin and protamine as prognostic molecular marker for testicular sperm extraction in patients with azoospermia. Fertil Steril *91*, 1424-1426.

Harris, C.R., Dewan, A., Zupnick, A., Normart, R., Gabriel, A., Prives, C., Levine, A.J., and Hoh, J. (2009). p53 responsive elements in human retrotransposons. Oncogene *28*, 3857-3865.

Hewitt, J.E., Lyle, R., Clark, L.N., Valleley, E.M., Wright, T.J., Wijmenga, C., van Deutekom, J.C., Francis, F., Sharpe, P.T., Hofker, M.*, et al.* (1994). Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum Mol Genet *3*, 1287-1295.

Holland, P.W., Booth, H.A., and Bruford, E.A. (2007). Classification and nomenclature of all human homeobox genes. BMC Biol *5*, 47.

Janjic, B., Andrade, P., Wang, X.F., Fourcade, J., Almunia, C., Kudela, P., Brufsky, A., Jacobs, S., Friedland, D., Stoller, R.*, et al.* (2006). Spontaneous CD4+ T cell responses against TRAG-3 in patients with melanoma and breast cancers. J Immunol *177*, 2717-2727.

Ji, Z., and Tian, B. (2009). Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. PLoS One *4*, e8419.

Jiang, C.L., Jin, S.G., Lee, D.H., Lan, Z.J., Xu, X., O'Connor, T.R., Szabo, P.E., Mann, J.R., Cooney, A.J., and Pfeifer, G.P. (2002). MBD3L1 and MBD3L2, two new proteins homologous to the methyl-CpG-binding proteins

MBD2 and MBD3: characterization of MBD3L1 as a testis-specific transcriptional repressor. Genomics *80*, 621-629.

Jiang, G., Yang, F., van Overveld, P.G., Vedanarayanan, V., van der Maarel, S., and Ehrlich, M. (2003). 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.

Jin, G., Kawsar, H.I., Hirsch, S.A., Zeng, C., Jia, X., Feng, Z., Ghosh, S.K., Zheng, Q.Y., Zhou, A., McIntyre, T.M.*, et al.* (2010). An antimicrobial peptide regulates tumor-associated macrophage trafficking via the chemokine receptor CCR2, a model for tumorigenesis. PLoS One *5*, e10993.

Jin, S.G., Tsark, W., Szabo, P.E., and Pfeifer, G.P. (2008). Haploid male germ cell- and oocyte-specific Mbd3l1 and Mbd3l2 genes are dispensable for early development, fertility, and zygotic DNA demethylation in the mouse. Dev Dyn *237*, 3435-3443.

Kaessmann, H., Vinckenbosch, N., and Long, M. (2009). RNA-based gene duplication: mechanistic and evolutionary insights. Nat Rev Genet *10*, 19-31.

Kawamura-Saito, M., Yamazaki, Y., Kaneko, K., Kawaguchi, N., Kanda, H., Mukai, H., Gotoh, T., Motoi, T., Fukayama, M., Aburatani, H.*, et al.* (2006). Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Hum Mol Genet *15*, 2125-2137.

Kawazu, M., Yamamoto, G., Yoshimi, M., Yamamoto, K., Asai, T., Ichikawa, M., Seo, S., Nakagawa, M., Chiba, S., Kurokawa, M.*, et al.* (2007). Expression profiling of immature thymocytes revealed a novel homeobox gene that regulates double-negative thymocyte development. J Immunol *179*, 5335-5345.

Klooster, R., Straasheijm, K., Shah, B., Sowden, J., Frants, R., Thornton, C., Tawil, R., and van der Maarel, S. (2009). Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. Eur J Hum Genet *17*, 1615-1624.

Knoepfler, P.S., Bergstrom, D.A., Uetsuki, T., Dac-Korytko, I., Sun, Y.H., Wright, W.E., Tapscott, S.J., and Kamps, M.P. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with pbx-Meis1/Prep1. Nucleic Acids Res *27*, 3752-3761.

Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O.*, et al.* (2007). The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord *17*, 611-623.

Lai, Y., and Gallo, R.L. (2009). AMPed up immunity: how antimicrobial peptides have multiple roles in immune defense. Trends Immunol *30*, 131-141.

Laoudj-Chenivesse, D., Carnac, G., Bisbal, C., Hugon, G., Bouillot, S., Desnuelle, C., Vassetzky, Y., and Fernandez, A. (2005). Increased levels of adenine nucleotide translocator 1 protein and response to oxidative stress are early events in facioscapulohumeral muscular dystrophy muscle. J Mol Med *83*, 216-224.

Leidenroth, A., and Hewitt, J.E. (2010). A family history of DUX4: phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. BMC Evol Biol *10*, 364.

Lemmers, R.J., van der Vliet, P.J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J.G., Snider, L., Straasheijm, K.R., van Ommen, G.J., Padberg, G.W.*, et al.* (2010a). A unifying genetic model for facioscapulohumeral muscular dystrophy. Science *329*, 1650-1653.

Lemmers, R.J., van der Vliet, P.J., van der Gaag, K.J., Zuniga, S., Frants, R.R., de Knijff, P., and van der Maarel, S.M. (2010b). Worldwide population analysis of the 4q and 10q subtelomeres identifies only four discrete interchromosomal sequence transfers in human evolution. Am J Hum Genet *86*, 364-377.

Lemmers, R.J., Wohlgemuth, M., van der Gaag, K.J., van der Vliet, P.J., van Teijlingen, C.M., de Knijff, P., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2007). Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. Am J Hum Genet *81*, 884-894.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. Science *327*, 996-1000.

Lunt, P.W., Jardine, P.E., Koch, M.C., Maynard, J., Osborn, M., Williams, M., Harper, P.S., and Upadhyaya, M. (1995). 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.

Lyle, R., Wright, T.J., Clark, L.N., and Hewitt, J.E. (1995). The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. Genomics *28*, 389-397.

Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B.*, et al.* (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev *25*, 594-607.

Melchionna, R., Di Carlo, A., De Mori, R., Cappuzzello, C., Barberi, L., Musaro, A., Cencioni, C., Fujii, N., Tamamura, H., Crescenzi, M.*, et al.* (2010). Induction of myogenic differentiation by SDF-1 via CXCR4 and CXCR7 receptors. Muscle Nerve *41*, 828-835.

Midorikawa, K., Ouhara, K., Komatsuzawa, H., Kawai, T., Yamada, S., Fujiwara, T., Yamazaki, K., Sayama, K., Taubman, M.A., Kurihara, H.*, et al.* (2003). Staphylococcus aureus susceptibility to innate antimicrobial peptides, beta-defensins and CAP18, expressed by human keratinocytes. Infect Immun *71*, 3730-3739.

Molnar, M., Dioszeghy, P., and Mechler, F. (1991). Inflammatory changes in facioscapulohumeral muscular dystrophy. Eur Arch Psychiatry Clin Neurosci *241*, 105-108.

Muller-Tidow, C., Readhead, C., Cohen, A.H., Asotra, K., Idos, G., Diederichs, S., Cauvet, T., Yang, R., Berdel, W.E., Serve, H.*, et al.* (2003). Successive increases in human cyclin A1 promoter activity during spermatogenesis in transgenic mice. Int J Mol Med *11*, 311-315.

Munsat, T.L., Piper, D., Cancilla, P., and Mednick, J. (1972). Inflammatory myopathy with facioscapulohumeral distribution. Neurology *22*, 335-347.

Nelson, J.D., Denisenko, O., and Bomsztyk, K. (2006). Protocol for the fast chromatin immunoprecipitation (ChIP) method. Nat Protoc *1*, 179-185.

Orrell, R.W. (2011). Facioscapulohumeral dystrophy and scapuloperoneal syndromes. Handb Clin Neurol *101*, 167-180.

Ostlund, C., Garcia-Carrasquillo, R.M., Belayew, A., and Worman, H.J. (2005). Intracellular trafficking and dynamics of double homeodomain proteins. Biochemistry *44*, 2378-2384.

Padberg, G.W., and van Engelen, B.G. (2009). Facioscapulohumeral muscular dystrophy. Curr Opin Neurol *22*, 539-542.

Palii, C.G., Perez-Iratxeta, C., Yao, Z., Cao, Y., Dai, F., Davison, J., Atkins, H., Allan, D., Dilworth, F.J., Gentleman, R.*, et al.* (2011). Differential genomic targeting of the transcription factor TAL1 in alternate haematopoietic lineages. Embo J *30*, 494-509.

Pandya, S., King, W.M., and Tawil, R. (2008). Facioscapulohumeral dystrophy. Phys Ther *88*, 105-113.

Parker, H.G., VonHoldt, B.M., Quignon, P., Margulies, E.H., Shao, S., Mosher, D.S., Spady, T.C., Elkahloun, A., Cargill, M., Jones, P.G.*, et al.* (2009). An expressed fgf4 retrogene is associated with breed-defining chondrodysplasia in domestic dogs. Science *325*, 995-998.

Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell *7*, 597-606.

Pierre, A., Gautier, M., Callebaut, I., Bontoux, M., Jeanpierre, E., Pontarotti, P., and Monget, P. (2007). Atypical structure and phylogenomic evolution of the new eutherian oocyte- and embryo-expressed KHDC1/DPPA5/ECAT1/OOEP gene family. Genomics *90*, 583-594.

Quenet, D., Mark, M., Govin, J., van Dorsselear, A., Schreiber, V., Khochbin, S., and Dantzer, F. (2009). Parp2 is required for the differentiation of post-meiotic germ cells: identification of a spermatid-specific complex containing Parp1, Parp2, TP2 and HSPA2. Exp Cell Res *315*, 2824-2834.

Reed, P.W., Corse, A.M., Porter, N.C., Flanigan, K.M., and Bloch, R.J. (2007). Abnormal expression of mu-crystallin in facioscapulohumeral muscular dystrophy. Exp Neurol *205*, 583-586.

Santamaria, C., Chillon, M.C., Garcia-Sanz, R., Balanzategui, A., Sarasquete, M.E., Alcoceba, M., Ramos, F., Bernal, T., Queizan, J.A., Penarrubia, M.J.*, et al.* (2008). The relevance of preferentially expressed antigen of melanoma (PRAME) as a marker of disease activity and prognosis in acute promyelocytic leukemia. Haematologica *93*, 1797-1805.

Sass, V., Schneider, T., Wilmes, M., Korner, C., Tossi, A., Novikova, N., Shamova, O., and Sahl, H.G. (2010). Human beta-defensin 3 inhibits cell wall biosynthesis in Staphylococci. Infect Immun *78*, 2793-2800.

Schrader, M., Muller-Tidow, C., Ravnik, S., Muller, M., Schulze, W., Diederichs, S., Serve, H., and Miller, K. (2002a). Cyclin A1 and gametogenesis in fertile and infertile patients: a potential new molecular diagnostic marker. Hum Reprod *17*, 2338-2343.

Schrader, M., Ravnik, S., Muller-Tidow, C., Muller, M., Straub, B., Diedrichs, S., Serve, H., and Miller, K. (2002b). Quantification of cyclin A1 and glyceraldehyde-3-phosphate dehydrogenase expression in testicular biopsies of infertile patients by fluorescence real-time RT-PCR. Int J Androl *25*, 202-209.

Schulz, W.A. (2006). L1 retrotransposons in human cancers. J Biomed Biotechnol *2006*, 83672.

Scott, M.P., Tamkun, J.W., and Hartzell, G.W., 3rd (1989). The structure and function of the homeodomain. Biochim Biophys Acta *989*, 25-48.

Semple, F., Macpherson, H., Webb, S., Cox, S.L., Mallin, L.J., Tyrrell, C., Grimes, G.R., Semple, C.A., Nix, M.A., Millhauser, G.L.*, et al.* (2011). Human beta-defensin 3 affects the activity of pro-inflammatory pathways associated with MyD88 and TRIF. Eur J Immunol.

Semple, F., Webb, S., Li, H.N., Patel, H.B., Perretti, M., Jackson, I.J., Gray, M., Davidson, D.J., and Dorin, J.R. (2010). Human beta-defensin 3 has immunosuppressive activity in vitro and in vivo. Eur J Immunol *40*, 1073-1078.

Simpson, A.J., Caballero, O.L., Jungbluth, A., Chen, Y.T., and Old, L.J. (2005). Cancer/testis antigens, gametogenesis and cancer. Nat Rev Cancer *5*, 615-625.

Small, R.G. (1968). Coats' disease and muscular dystrophy. Trans Am Acad Ophthalmol Otolaryngol *72*, 225-231.

Smit, A.F. (1993). Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic Acids Res *21*, 1863-1872.

Snider, L., Asawachaicharn, A., Tyler, A.E., Geng, L.N., Petek, L.M., Maves, L., Miller, D.G., Lemmers, R.J., Winokur, S.T., Tawil, R.*, et al.* (2009). 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.

Snider, L., Geng, L.N., Lemmers, R.J., Kyba, M., Ware, C.B., Nelson, A.M., Tawil, R., Filippova, G.N., van der Maarel, S.M., Tapscott, S.J.*, et al.* (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet *6*, e1001181.

Stanghellini, I., Falco, G., Lee, S.L., Monti, M., and Ko, M.S. (2009). Trim43a, Trim43b, and Trim43c: Novel mouse genes expressed specifically in mouse preimplantation embryos. Gene Expr Patterns *9*, 595-602.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M.*, et al.* (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature *453*, 534-538.

Tawil, R. (2008). Facioscapulohumeral muscular dystrophy. Neurotherapeutics *5*, 601-606.

Tawil, R., and Van Der Maarel, S.M. (2006). Facioscapulohumeral muscular dystrophy. Muscle Nerve *34*, 1-15.

Taylor, D.A., Carroll, J.E., Smith, M.E., Johnson, M.O., Johnston, G.P., and Brooke, M.H. (1982). Facioscapulohumeral dystrophy associated with hearing loss and Coats syndrome. Ann Neurol *12*, 395-398.

Tupler, R., Berardinelli, A., Barbierato, L., Frants, R., Hewitt, J.E., Lanzi, G., Maraschio, P., and Tiepolo, L. (1996). Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy. J Med Genet *33*, 366-370.

van der Maarel, S.M., Tawil, R., and Tapscott, S.J. (2011). Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends Mol Med *17*, 252-258.

van Overveld, P.G., Lemmers, R.J., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. Nat Genet *35*, 315-317.

Voit, T., Lamprecht, A., Lenard, H.G., and Goebel, H.H. (1986). Hearing loss in facioscapulohumeral dystrophy. Eur J Pediatr *145*, 280-285.

Wallace, L.M., Garwick, S.E., Mei, W., Belayew, A., Coppee, F., Ladner, K.J., Guttridge, D., Yang, J., and Harper, S.Q. (2010). DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. Ann Neurol.

Wallace, L.M., Garwick, S.E., Mei, W., Belayew, A., Coppee, F., Ladner, K.J., Guttridge, D., Yang, J., and Harper, S.Q. (2011). DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. Ann Neurol *69*, 540-552.

Wang, T., Zeng, J., Lowe, C.B., Sellers, R.G., Salama, S.R., Yang, M., Burgess, S.M., Brachmann, R.K., and Haussler, D. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc Natl Acad Sci U S A *104*, 18613-18618.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T.*, et al.* (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature *453*, 539-543.

Wettenhall, J.M., and Smyth, G.K. (2004). limmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics *20*, 3705-3706.

Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R.*, et al.* (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet *2*, 26-30.

Wu, S.L., Tsai, M.S., Wong, S.H., Hsieh-Li, H.M., Tsai, T.S., Chang, W.T., Huang, S.L., Chiu, C.C., and Wang, S.H. (2010). Characterization of genomic structures and expression profiles of three tandem repeats of a mouse double homeobox gene: Duxbl. Dev Dyn *239*, 927-940.

Wuebbles, R.D., Long, S.W., Hanel, M.L., and Jones, P.L. (2010). Testing the effects of FSHD candidate gene expression in vertebrate muscle development. Int J Clin Exp Pathol *3*, 386-400.

Yang, D., Chertov, O., Bykovskaia, S.N., Chen, Q., Buffo, M.J., Shogan, J., Anderson, M., Schroder, J.M., Wang, J.M., Howard, O.M.*, et al.* (1999). Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. Science *286*, 525-528.

Zalzman, M., Falco, G., Sharova, L.V., Nishiyama, A., Thomas, M., Lee, S.L., Stagg, C.A., Hoang, H.G., Yang, H.T., Indig, F.E.*, et al.* (2010). Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature *464*, 858-863.

Zeng, W., de Greef, J.C., Chen, Y.Y., Chien, R., Kong, X., Gregson, H.C., Winokur, S.T., Pyle, A., Robertson, K.D., Schmiesing, J.A.*, et al.* (2009). Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). PLoS Genet *5*, e1000559.

Zhong, Y.F., and Holland, P.W. (2011). The dynamics of vertebrate homeobox gene evolution: gain and loss of genes in mouse and human lineages. BMC Evol Biol *11*, 169.

VITA

Linda N. Geng was born in Suzhou, China and has called many places her home, including Gothenburg, Sweden; Boston; Pittsburgh; Houston; and, finally, Seattle. In 2006, she graduated from Rice University with a Bachelor of Science in Biochemistry and Cell Biology as well as a Bachelor of Arts in Psychology. In 2011, she earned a Doctor of Philosophy at the University of Washington in Molecular and Cellular Biology. She performed her doctorate work at Fred Hutchinson Cancer Research Center in the laboratory of Dr. Stephen Tapscott. In 2012, she returned to medical school to complete her training in the Medical Scientist Training Program at the University of Washington.