Telomere Position Effect (TPE) Regulates DUX4 in Human Facioscapulohumeral Muscular Dystrophy (FSHD)

Published in final edited form as:
Nat Struct Mol Biol. 2013 June ; 20(6): 671–678. doi:10.1038/nsmb.2571.

Telomere Position Effect (TPE) Regulates *DUX4* in Human Facioscapulohumeral Muscular Dystrophy (FSHD)

Guido Stadler[1,2], **Fedik Rahimov**[2,3], **Oliver D. King**[2,4,5], **Jennifer C. J. Chen**[2,4,5], **Jerome D. Robin**[1,2], **Kathryn R. Wagner**[2,6,7], **Jerry W. Shay**[1,8], **Charles P. Emerson Jr.**[2,4,5], and **Woodring E. Wright**[1,2]

[1]Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390

[2]Senator Paul D. Wellstone Muscular Dystrophy Cooperative Research Center

[3]Program in Genomics, Division of Genetics, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115

[4]Boston Biomedical Research Institute, Watertown, MA 02472

[6]The Hugo W. Moser Research Institute at Kennedy Krieger Institute, Baltimore, MD 21205

[7]Department of Neurology and Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD 21205

[8]CEGMR, King Abulaziz University, Jeddah, Saudi Arabia

Abstract

Telomeres may regulate human disease by at least two independent mechanisms. 1) Replicative senescence occurs once short telomeres generate DNA damage signals that produce a barrier to tumor progression. 2) Telomere Position Effect (TPE) can change gene expression at intermediate telomere lengths in cultured human cells. We here report a human disease, facioscapulohumeral muscular dystrophy (FSHD) where telomere length may well contribute to its pathogenesis. FSHD is age-related and genetically only 25-60 kb from the end of chromosome 4q. We used a floxable telomerase to generate isogenic clones with different telomere lengths from patients and their unaffected siblings. DUX4, the primary candidate for FSHD pathogenesis, is upregulated >10-fold in FSHD myoblasts-myotubes with short versus long telomeres, and its expression is inversely proportional to telomere length. FSHD may represent a human disease in which TPE contributes to its age-related phenotype.

Keywords

aging; muscle; D4Z4 repeats; chromosome 4q; telomere shortening

Facioscapulohumeral muscular dystrophy (FSHD1A, MIM 158900) is one of the most prevalent myopathies, affecting approximately 1:20,000 individuals [1-4]. It is genetically

Corresponding author: Woodring E. Wright, M.D., Ph.D. Department of Cell Biology UT Southwestern Medical Center 5323 Harry Hines Blvd. Dallas, TX 75390-9039 Tel: 214-633-1996 Woodring.wright@utsouthwestern.edu.
[5]Current address: Wellstone Program, Department of Cell & Developmental Biology and Neurology, University of Massachusetts Medical School, Worcester, MA 01655

AUTHOR CONTRIBUTIONS G.S. helped design and executed most of the experiments and contributed to the writing of the manuscript, W.E.W. designed the experiments and wrote the manuscript, KRW and DGL obtained phenotypes and muscle biopsies. J.C.J.C established the primary muscle cultures, F.R. performed the microfluidics expression analysis, O.D.K. analyzed the probability of expression per cell and other statistical measures, and all authors edited the manuscript.

linked to a reduced number of tandemly repeated 3.3 kb D4Z4 elements present near the telomere of chromosome 4q. Normal individuals contain up to 100 repeats while only 1-10 repeats are present in patients with FSHD [5,6]. The disease has a highly variable clinical expression of progressive atrophy and weakness of the facial, scapular and upper arm muscles (hence the name) and has the unusual characteristic for myopathies of delayed appearance, where symptoms often do not appear until the second/third decade or later [7]. The disease requires additional genetic markers present on both sides of the repeats (Fig.1a).

The current leading candidate mediating pathogenesis in FSHD is the DUX4 homeobox protein in the final D4Z4 repeat [8-10]. The long, but not the short, form of DUX4 is toxic when overexpressed [11-13], though toxicity of endogenous DUX4 has not yet been demonstrated. It has become increasingly clear that there are additional factors contributing to the pathogenesis of FSHD. Although FSHD has traditionally been considered an autosomal dominant disease, the prevalence of its genetic signature (a reduced number of D4Z4 repeats on chromosome 4 in combination with the 4A161PAS haplotype) is present in ~1% of the general population [14], roughly two orders of magnitude higher than the incidence of the disease. Moreover, DUX4-fl mRNA and protein is expressed in some unaffected subjects without D4Z4 deletions [15]. This suggests other unknown factors are important.

Telomere shortening has primarily been investigated as a tumor-suppression mechanism. It limits the number of available divisions before some telomeres become sufficiently short to induce a non-repairable DNA damage signal. This causes growth arrest and prevents pre-malignant cells from continuing to accumulate the changes needed to form progressive tumors [16]. A second function of telomere shortening that has received relatively little attention is its potential for regulation of gene expression by trapping adjacent genes in heterochromatin known as Telomere Position Effect (TPE)[17].

There is no explanation for the delayed onset of FSHD. There has been speculation about Position Effect Variegation (PEV), the term derived from the Drosophila melanogaster literature to describe the repression of genes by adjacent heterochromatin [18]. TPE describes the special case of PEV that occurs for genes located near telomeres. TPE, originally described in D. melanogaster and S. cerevisiae[19,20], exists in mammalian cells expressing artificially produced reporter constructs [21-24]. A single endogenous human gene located 1Mb from the end of chromosome 1p, *ISG15*, has been found to be regulated by telomere length [25]. Because many genes distal to *ISG15* are not regulated by telomere length, it does not conform to classical TPE in which repression extends in a continuously decreasing fashion from the telomeres. *ISG15* is involved in innate immunity, and although an attractive candidate for contributing to an increased aging inflammatory response ("inflammaging"), it has so many inputs driving its expression that it will be difficult to establish that telomere shorting is an important contributor.

The presence of the FSHD locus adjacent to the end of chromosome 4q raises the possibility that >300 kb of D4Z4 repeats function as an insulator [26,27]-repressor [26] that both blocks expression of *DUX4* (repressor function) and the spreading of telomeric heterochromatin (insulator function) in normal individuals. The loss of these activities in contracted alleles would then permit telomere shortening to regulate the expression of internal genes. This could explain the age-associated manifestation of symptoms, since significant shortening might be required before genes such as *DUX4* were upregulated (Fig. 1b). To answer this question, we first obtained muscle samples from both affected FSHD patients and their unaffected family members. We then immortalized myoblasts using a floxable telomerase (*TERT*) cDNA [28], isolated individual clones, excised *TERT* at different times, and examined the effects of telomere shortening on gene expression in isogenic cell lines in the absence of confounding factors. We demonstrate that DUX4 full-length (DUX4-fl)

expression is upregulated by telomere shortening in cells from FSHD subjects; in some cases 100-fold between cells with long and short telomeres. The effect is progressive with decreasing telomere length and occurs long before terminal telomere shortening would induce replicative senescence. The targets of *DUX4* expression are also upregulated as a consequence of the TPE induced increased expression of *DUX4*. The effect of TPE is most prominent with *DUX4*, the gene nearest the telomere. The effect of TPE is present but less prominent compared to *DUX4* with *FRG2*, approximately 70 kb more internal, and not observed with *FRG1*, an additional 90 kb from the telomere (Fig. 1a). TPE in FSHD thus follows the classical model of continuous heterochromatin spreading that declines with distance from the telomere. FSHD is thus the first human disease in which classic TPE may play a role in its age-related pathogenesis.

ONLINE METHODS

Cell culture

Human myoblasts [28] were seeded in dishes coated with 0.1% pigskin gelatin (Sigma Aldrich) in 4:1 DMEM:Medium 199 supplemented with 15% FBS, 0.02 M HEPES, 1.4 mg/l vitamin B12, 0.03 mg/l $ZnSO_4$, 0.055 mg/l dexamethasone, 2.5 μg/l hepatocyte growth factor, and 10 μg/l basic-FGF. Cultures were passaged at ~50% confluency in 2-5% oxygen. Population doublings (PDs) were calculated as PD = ln[(final number of cells)/(initial number of cells)]/ln(2). Cell strains used in this study are described in Supplementary Table 1. GM17731 and 38/03 were from Coriell Cell Repository (Camden, NJ) and Muscle Tissue Culture Collection (University of Munich), respectively.

For differentiation, cells seeded in growth medium were switched to differentiation medium (2% horse serum in 4:1 Dulbecco modified Eagle medium: Medium 199) when 70-90% confluent.

Reversible immortalization

Primary human myoblasts prematurely growth arrest in culture for unknown reasons unrelated to telomere length (e.g., lack of an appropriate growth factor). We have shown expression of *CDK4* can bypass this premature growth arrest in myoblasts without affecting normal cell cycle kinetics or the ability to differentiate normally [28,47]. Primary cultures were first transduced with *CDK4* to maintain CD56+ myogenic cells, then with an excisable telomerase (Lox-*TERT*-hygromycin). Lox-*TERT*-hygromycin was excised by Cre-recombinase, by transient transfection with pOG231 [48] or stable integration of retroviral Cre. Excision of *TERT*-hygromycin was confirmed in clones by testing for hygromycin sensitivity and telomerase activity.

Terminal restriction fragment (TRF) assay

TRF assays were done as described [49].

Quantitative real-time RT-PCR

Cells were lysed (RNeasy plus kit (Qiagen)) after washing with PBS, scraped (BD Biosciences) and sheared by centrifugation through Qiashredder columns (Qiagen). Total RNA purified according to the manufacturer's instructions was quantified on a Nanodrop 1000 spectrophotometer (Thermo Scientific). 2 μg total RNA was reverse transcribed in 20 μl (Transcriptor first strand cDNA synthesis kit (Roche), including random hexamers and oligo$(dT)_{18}$ primers). The cDNA was diluted 1:4 in water for quantitative RT-PCR (qRT-PCR) in triplicates (Lightcycler 480 (Roche)).

Supplementary Table 3 lists primers and PCR conditions. Melting curves were analyzed (SYBR green or EvaGreen assays) to exclude nonspecific amplification products. We confirmed amplicon size at least once on agarose gels. Crossing- point (Cp) values were the second derivative maximum. Cp values were normalized by subtracting the geometric mean of three housekeeping genes (*GAPDH*, *PPIA* and *HPRT1*). All Cp values were corrected by their PCR efficiency, determined by 1:2 or 1:4 cDNA dilution series. Cp values are on a reversed axis where a decreased Cp indicates increased mRNA.

High throughput quantitative real-time RT-PCR

High-throughput qRT-PCR used the BioMark 96.96 Dynamic Array (Fluidigm) with TaqMan Gene Expression Assays (Applied Biosystems). 1.25 μl of cDNA was pre-amplified (pooled and 1:100 diluted TaqMan assay mix, final concentration 0.2X) for 14 cycles, then diluted 1:5 with 1X TE buffer. 2.5 μl of each diluted sample was mixed with TaqMan Gene Expression Mastermix, distributed into BioMark 96.96 Dynamic Array nanoliter reaction chambers and simultaneously injected with 2.5 μl of 10X TaqMan assay mix. Each reaction was analyzed in triplicate.

Qualitative PCR for *DUX4* splicing

Splicing of *DUX4* transcripts [13,15] used Phusion Hot Start II High-Fidelity DNA Polymerase (Fisher Scientific). Primers were 14 for and 183 rev nested with 15A for and 184 rev (Supplementary Table 3). PCR conditions: 98°C 2 min, 25 cycles 98°C 15 s, 62°C 20 s, 72°C 1 min, final extension 72°C 10 min. 8% of this PCR reaction was used for nested PCR: 98°C 2 min, 20-30 cycles 98°C 15 s, 62°C 15 s, 72°C 50 s, final extension 72°C 10 min. PCR products from 1-2% agarose gels were gel purified and sequenced.

Reporter assay

Myoblasts co-transfected with *DUX4*-promoter- or *FRG2*-promoter-Firefly luciferase constructs [9,33] and an SV40-driven Renilla luciferase (Promega), using PolyJet (SignaGen) were either harvested 48 h after transfection or after an additional 5 days in differentiation medium. Luciferase was measured using the Dual-Luciferase Reporter Assay System (Promega) on an Optocomp I luminometer (MGM Instruments).

Statistical analysis

Rates of *DUX4* detection in qRT-PCR data were fit with a binomial mixed-effect model using R package lme4 [15,50]. Fixed effects were D4Z4 locus length (contracted or normal), telomere length (long or short), cell state (cycling or differentiated), and muscle type (biceps or deltoid), including all interactions between these. Random effects for qRT-PCR date and batch (each consisting of three technical replicates), days post differentiation, and subject of origin accounted for these additional sources of variation. Each fixed effect significance was computed using a likelihood ratio test of the full model versus with the fixed effect deleted.

Determination of the fraction of cells expressing DUX4-fl from samples containing limiting cell numbers

A multi-level probabilistic model calculated whether increased expression of *DUX4* was from increased expression per cell or a greater fraction of cells expressing *DUX4*.

Differentiated cells were trypsinized, and a nuclear aliquot taken (after centrifugation, the pellet was incubated in hypotonic solution (0.1M Na-citrate pH 7.5, 2% Triton-X-100, 0.5% SDS, 5 min), vortexed and hemocytometer counted with trypan blue). Remaining cells were centrifuged, washed (cold PBS), resuspended (cold PBS), and diluted to the desired concentration of nuclei. Aliquots (50 ul PBS) were added to 50 ul 2x lysis buffer (200 mM

Tris-HCl pH7.5, 1M LiCl, 20mM EDTA, 2% LiDS, 10 mM DTT) and frozen (−80°C). mRNA was isolated from individual aliquots using Dynabeads mRNA Direct kit (Life Technologies), eluted by heat and reverse transcribed (oligo(dT)$_{18}$ primers, 25 ul, Transcriptor first strand cDNA synthesis kit (Roche)). 2 ul served as template for DUX4-fl 3′ PCR [13] in triplicates. Primers 182 for and 183 rev were nested with 1A for and 184 rev. Cycling conditions were as above, with decreased extension times (40 s first PCR, 30 s second PCR). Amplicons were run on 2% agarose gels, gel purified and sequenced.

We used a multi-level probabilistic model that accounts for (1) variability in the number of copies of *DUX4* in any aliquot, due to variability per cell; (2) variability in the number of copies of *DUX4* per replicate given the number per aliquot, due to stochastic multinomial sampling; and (3) variability in whether *DUX4* is detected in a replicate given the number of copies of *DUX4* in the replicate, due to stochastic effects in PCR.

For (1), we use a zero-inflated Poisson (ZIP) model [38] for the number of copies of *DUX4* per cell. This models a situation in which at the time of mRNA extraction, with probability β the number of *DUX4* transcripts in a cell follows a Poisson distribution with mean λ, and zero transcripts has a probability of $1 - \beta$. Note that a cell from the Poisson component of this mixture may also have count zero (with probability $e^{-\lambda}$), so the overall distribution X for each cell is $\Pr(X=k|\beta,\lambda) = \beta \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda}$ for $k > 0$ and $\Pr(X = 0|\beta,\lambda) = (1 - \beta) + \beta \cdot e^{-\lambda}$

If X follows this ZIP distribution for each cell, then the distribution for the total number Y of *DUX4* transcripts in an aliquot of n cells is given by

$$\Pr(Y=k|\beta,\lambda,n) = \sum_{i=0}^{n} \left\{ \binom{n}{i} \beta^i (1-\beta)^{n-i} \cdot \frac{(i\lambda)\,k}{k!} \cdot e^{-i\lambda} \right\}$$

where $\binom{n}{i} = \frac{n!}{i!\,(n-i)\,!}$ is the binomial coefficient.

For (2), if k is the number of *DUX4* transcripts in a cell aliquot, which has volume 25 μl, then the number of *DUX4* transcripts r_i in each of the three 2 μl replicates sampled from this aliquot follows a multinomial distribution, with

$$\Pr(r_1=k_1, r_2=k_2, r_3=k_3|Y=k) = \binom{k}{k_1}\binom{k-k_1}{k_2}\binom{k-k_1-k_2}{k_3}\left(\frac{2}{25}\right)^{k_1+k_2+k_3}\left(\frac{19}{25}\right)^{k-k_1-k_2-k_3}$$

For (3), we model the probability of detecting *DUX4* in a replicate that initially has k_i copies of *DUX4* by $\Pr(d_i = 1|r_i = k_i,\gamma) = 1 - (1 - \gamma)^{k_i}$ where the probability γ represents the probability of detection when there is initially 1 copy of the transcript, and d_i in {0,1} is an indicator variable for detection. To make the parameter estimates more robust to outliers we include a 2% probability that detection will be random for any given replicate, and a 2% probability that all three replicates from any given aliquot will take the same random value, so

$$\Pr(d_1, d_2, d_3 | k_1, k_2, k_3, \gamma) = 0.01 \cdot \delta_{d_1 d_2} \delta_{d_2 d_3} + 0.99 \cdot \left(\prod_{i=1}^{3} (0.01 + 0.99) \cdot \left\{ \delta_{1 d_i} + (-1)^{d_i} \right\} (1-\gamma)^{k_i} \right)$$

where δ_{ij} is the Kronecker delta.

With the parameters $\pi = (\beta, \lambda, \gamma)$ specified we can compute the joint probability of detecting *DUX4* in any number of replicates for an aliquot by summing the products of the probabilities from (1-3) over all k in (1) and all $k_1, k_2, k_3 \leq k$ in (2). (For computational efficiency we lumped together all $k > 15k^{1/2} + 50/\log(1/(1-\gamma))$ in (1) and approximated the multinomial in (2) by the product of three binomial distributions for $k > 1000$; since the probabilities in (3) converge for large k and are bounded away from zero, this had a negligible effect on the log-probabilities.)

Then applying Bayes' rule we can compute the likelihood of the parameters [51] given the combined observed data in Supplementary Table 2. Fig. 6a and b show the log-likelihood surfaces for β and λ when γ is fixed at 0.5. Fig. 6c and d show the marginal posterior probability distributions for β and λ, using a uniform [0, 1] prior for β and an exponential (scale=100) prior for λ. The estimates for β were largely insensitive to the choice of γ in the range from 0.1 to 0.9 examined (Fig. 6e), whereas the estimates of λ varied inversely with γ (Fig. 6f).

RESULTS

Generating clonal cell lines with long and short telomeres

Primary cultures are heterogeneous, both in terms of cell types and telomere lengths within a particular cell type. In order to experimentally isolate the effects of telomere length from other confounding factors, we first isolated clones and then produced isogenic subclones with different telomere lengths using a floxable *TERT*. Cells were obtained from two family cohorts (01 and 15) with affected and unaffected family members [28,29]. One additional FSHD strain (GM17731) and one myopathy control (Merosin-deficient congenital muscular dystrophy, MDC1A, 38/03) were included. Multiple isogenic subclones were derived from 12 lines (e.g., deltoid and biceps from one individual equals two lines) described in Supplementary Table 1: seven lines from FSHD subjects, four from matched unaffected sibling controls, and one from the MDC1A control myopathy. Importantly, we controlled for family effects [29] by comparing FSHD lines to lines from unaffected siblings (Supplementary Table 1: e.g. lines 15Vbic and 15Vdel are derived from biceps and deltoid of the unaffected sister of FSHD subjects 15A and 15B, both of which have contracted D4Z4 alleles). Primary muscle cultures transduced with *CDK4* [to circumvent growth arrest due to inadequate culture conditions [30] (Fig. 1c), enabling the isolation of myogenic clones] were reversibly immortalized with a floxable *TERT* [28]. Excision of *TERT* by Cre-recombinase one week after *TERT* introduction yielded subclones with slightly elongated telomeres and extended lifespan (Fig. 1d: compare 15Abic subclone 1 (orange line) with the parental *CDK4* clone (black line); see Supplementary Fig. 1 for other clones). In vitro propagation produced shortened telomeres (e.g., subclone 1 of 15Abic in Fig. 1e and other clones in Supplementary Fig. 1). An isogenic subclone with long telomeres was generated by excising *TERT* ~60 population doublings (PDs) after its introduction (e.g., subclone 2 in Fig. 1d and e). All clones and subclones were tested for efficient fusion into multinucleated myotubes when switched to differentiation medium. Clones with short telomeres were used before the end of their logarithmic growth phase, well before the onset of replicative senescence.

DUX4 and *FRG2* are induced in FSHD with short telomeres

Several genes have been reported as dysregulated in FSHD, including the current primary candidate *DUX4* [8,13], *FRG2*, *FRG1*, and *SLC25A4* [31,32]. We investigated the effects of telomere length on transcription of these genes by measuring their mRNA levels using qRT-PCR. Fig. 2 shows representative results for the three genes closest to the 4q telomere (*DUX4*, *FRG2* and *FRG1*, see Fig. 1a for genomic locations) in isogenic subclones with long (light colors) or short telomeres (dark colors), derived from two brothers with a shortened FSHD locus (15A and 15B, blue bars) and their unaffected sister without a D4Z4 contraction (15V, red bars). Data from independent repetitions are combined in Supplementary Fig. 2a and b. Subclones from unaffected 15V did not express detectable *DUX4* under any condition, as expected from their nonpermissive genotype (4qB/4qB and missing the poly(A) signal that stabilizes *DUX4* transcripts) [15]. We found that most cycling FSHD subclones with long telomeres had *DUX4* levels 5 crossing-point (Cp) values (~30x) above the qRT-PCR detection limit that increased 4-1000x upon telomere shortening. Differentiation increased *DUX4* levels as reported [15]. Short telomeres augmented this increase. 15Abic, 15Bbic and 15Bdel showed ~8 Cp difference (~250x increased expression) between short and long telomeres. The 15Adel subclone exhibited a smaller increase (3-4 Cp: ~10x-20x). *DUX4* detection rates were significantly higher in samples with shortened versus normal D4Z4 loci ($p = 1.2\times10\text{-}4$), in samples with short versus long telomeres ($p = 9.9\times10^{-11}$), and in differentiated versus cycling cells ($p = 1.1\times10^{-3}$, likelihood ratio test for binomial mixed-effect models). *FRG2* was similarly regulated by both differentiation and telomere length in FSHD cells. Although expressed by cells without D4Z4 contractions (15Vbic and 15Vdel), these controls showed no induction by differentiation or telomere shortening. *FRG1* levels were constant under all conditions tested, as were other more centromeric genes at 4q (*LRP2BP*, *PDLIM3* and *SLC25A4*, data not shown). We confirmed that all subclones, whether FSHD or controls, long or short telomeres, had similar expression profiles of several skeletal muscle differentiation markers (Supplementary Fig. 3a).

We confirmed the generality of these results in six additional subclones from three FSHD lines and five additional subclones from three control lines (Supplementary Fig. 2c-f). We found increased expression of both *DUX4* and *FRG2* in both subclones of FSHD line 01Adel, 1/3 subclones from FSHD line 01Abic, and the subclone from FSHD line GM17731. We observed *DUX4* in two subclones from unaffected subject 01Ubic, one from 01Udel, and saw it slightly upregulated in cells with short telomeres. Importantly, we did not detect *DUX4* in cells from an unrelated myopathy, and found *FRG2* levels were similar to cells from healthy donors, and marginally regulated by telomere length (Supplementary Fig. 2c).

To summarize, we demonstrated that the two genes closest to the telomere at chromosome 4q, but not more centromeric genes, are regulated by telomere length, specifically in cells with contracted D4Z4 repeats.

DUX4 expression is inversely proportional to telomere length

Fig. 3a shows the general outline of our approach to generate samples from an FSHD clone (01Adel) with a wide range of telomere lengths. The corresponding growth curves are depicted in Fig. 3b. We treated the Lox-*TERT* expressing clone (red line) with Cre-recombinase at various times to produce subclones of different telomere lengths and time in culture. We obtained subclone 1 with ~20 kb telomeres (point A in Fig. 3b) by excising Lox-*TERT* after 37 PDs (Fig. 3a). We then cultured these cells and harvested samples at multiple time points (B-G in Fig. 3b). We excised *TERT* after 180 PDs to yield subclone 2 which had ~24 kb telomeres at the time of analysis at PD 216 (Fig. 3a and b). We analyzed

telomere length, and *DUX4* and *FRG2* mRNA levels for all samples. We found a linear correlation between telomere length and *DUX4* Cp values for cycling and differentiating myoblasts, with correlation coefficients of 0.93-0.96 (Fig. 3c). We showed that the continuous increase in *DUX4* expression in subclone 1 is not simply due to accumulating PDs since subclone 2 with ~24 kb telomeres exhibited the lowest level of *DUX4* at the same PD when subclone 1 had the highest (PD 216, samples A and H in Fig. 3b and c). We also found that *FRG2* expression inversely correlated with telomere length, although the correlation was weaker (correlation coefficient 0.7 for differentiating cells, Supplementary Fig. 4).

DUX4 and *FRG2* regulation by telomere length is direct

Transiently transfected reporter constructs containing the *DUX4* or *FRG2* promoter fused to luciferase [9,33] should respond if *DUX4* and *FRG2* upregulation in cells with short telomeres were regulated in trans rather than directly by the adjacent telomere. However, we found that both promoter constructs were equally active in cells with long or short telomeres (Fig. 4). We demonstrated that these reporter constructs could effectively respond to trans-acting regulatory factors, as they were induced 10-fold with differentiation as at their native telomeric position (Fig. 4).

DUX4 induction by telomere shortening induces *DUX4* targets

DUX4-fl has two N-terminal homeodomains [34] that mediate sequence specific DNA binding and a C-terminal transactivation domain [35]. A cryptic splice site can remove the C-terminal domain producing DUX4-s [13] with dominant negative activity following overexpression [36]. DUX4-fl, but not DUX4-s, is thought to cause or at least contribute to FSHD [13]. The *DUX4* data in Figs. 2 and 3 did not distinguish between *DUX4* splice variants. The level of *DUX4* is far too low to observe by Northern analysis, requiring ~50 cycles of nested RT-PCR to detect. We consistently detected DUX4-fl in our clonal lines using RT-PCR (Fig. 5a). We barely detected DUX4-s (01Adel is shown as representative in Fig. 5a). In agreement with qRT-PCR, we observed stronger signals in samples from cells with short telomeres compared to cells with long telomeres. We verified all PCR products as chromosome 4q *DUX4* transcripts by sequencing.

DUX4 activated its target genes in these cells. We observed increased transcription from two targets (*ZSCAN4* and *KHDC1*) [36] (the samples in Fig. 2 are shown in Fig. 5b, and eight additional targets are shown in Supplementary Fig. 5a). Overall, we showed that *DUX4* target gene expression correlated with *DUX4* expression. They were expressed at low levels in cycling cells, with a trend of being higher in both FSHD versus unaffected, and FSHD with short versus long telomeres. We found that *DUX4* target genes were slightly upregulated with differentiation in unaffected clones compared to FSHD clones with long telomeres (change in Cp: ~5, ~30x), and dramatically more in FSHD clones with short telomeres (change in Cp: ~11, ~2,000x). Plotting Cp values of each target gene as a function of *DUX4* Cp gave linear correlations with R^2 values of 0.79 – 0.89 (Supplementary Fig. 5b). In summary, we demonstrated that telomere shortening led to increased expression and transcriptional activity of DUX4-fl and its potential contribution to FSHD pathology.

Telomere shortening increases the fraction of DUX4-fl cells

The data of Fig.3 did not indicate whether telomere shortening increased the *DUX4* levels per expressing cell or the fraction of expressing cells. The number of DUX4-fl-expressing nuclei in differentiating FSHD myoblasts is extremely low (1 in 1,000 to 2,000) by antibody staining [13,15]. The fraction of nuclei expressing DUX4-fl can be determined by what amounts to a fluctuation analysis using limiting numbers of nuclei[37]. We generated cDNA from multiple aliquots of 30 – 30,000 nuclei with long or short telomeres from an 01Adel

subclone. Three PCR reactions from each aliquot were scored as either positive or negative. Agarose gel images are shown in Supplementary Fig. 6. Zero, one, two or three positives denotes the number of positive RT-PCR reactions per triplicate for a given aliquot in Supplementary Table 2. We modeled the number of DUX4 transcripts per cell with a zero-inflated Poisson distribution [38], embedded in a multi-level model that also accounts for stochastic subsampling from aliquots, and stochastic RT-PCR detection. We found that the estimated fraction of cells expressing *DUX4* was ~10-fold higher for short than for long telomeres (~1 in 200 versus ~1 in 2,000), whereas the estimated average transcript number for those cells that do express *DUX4* remained roughly the same (Fig. 6). We conclude that TPE thus behaved like an on/off switch where increasing fractions of cells expressed DUX4-fl as telomeres shortened.

DISCUSSION

FSHD is a complex age-related disease in which the molecular pathogenesis is poorly understood. We demonstrated that DUX4-fl, the primary known candidate for producing muscle toxicity, is regulated by telomere length in cultured myoblasts. FSHD is thus the first human disease in which classic TPE could contribute to the age-related phenotype.

Significant telomere shortening occurs in myoblasts between neonates and adulthood [39]. Our data showing that *DUX4* expression progressively increased as telomeres shortened does not require FSHD telomeres to be shorter than those from normal individuals. However, the pathological foci of degeneration and regeneration that are observed in FSHD suggest that satellite cell proliferation with an accompanying telomere shortening might be occurring. An attractive possibility is that myoblast proliferation produces telomere shortening, which increases DUX4-fl expression and increased toxicity in a positive feedback loop that contributes to disease progression.

A central issue for determining the in vivo relevance of these in vitro results will be to determine the contribution of telomere length to the variable penetrance of FSHD that is frequently observed. There are a variety of theoretical and technical issues that need to be resolved before this data can be obtained. Average telomere length is highly variable in humans, and the average length is a complex sum of different lengths on different chromosome ends and different lengths on homologous paternally and maternally inherited chromosomes [40-42]. Much of the variability in the magnitude of TPE we have observed in different lines could be explained by differences in telomere length of the specific chromosome 4q containing the contracted allele. Differences in the inherited telomere length of the contracted 4q allele could also contribute to the large variation in penetrance observed in FSHD.

It remains to be determined how closely the relative telomere length of the contracted 4q allele compares to average telomere length in myoblasts versus a more accessible cell population such as circulating lymphocytes. Our data indicates the importance of pursuing this information. Once the average telomere length is determined, in situ hybridization techniques using probes to D4Z4 repeats, telomeric repeats and subtelomeric regions of 4q and 10q should be able to quantify the relative signal intensity of the contracted 4q chromosome compared to the total telomeric signal from all chromosome ends, and thus whether the contracted allele is relatively long or short. If this explains a significant part of the variance then it would have important clinical implications for counseling patients and their family members.

Yeast TPE only extends a few kb into the subtelomeric regions [19,43], and it was unknown how far TPE could extend in mammals. This study demonstrates that human TPE can

influence gene expression at least 80 kb from the start of the telomeric repeats. This is a minimal value, and will probably vary greatly between different chromosome ends depending on what repressors-insulators-propagators are present.

Previous studies have shown a reduction in repressive chromatin marks and CpG methylation in the subtelomeric region of FSHD compared to normal subjects [44]. This is consistent with the evolutionarily conserved role of cytidine methylation as a mechanism for recognizing and silencing repetitive sequences [45]. The presence of >10 repeats could induce these subtelomeric modifications and prevent telomeric shortening from influencing gene expression in this region. The reduction in the number of D4Z4 repeats to <11 in FSHD could remove many of these modifications and make the region permissive for TPE. Differences in epigenetic modifications could also explain differences in the magnitude of TPE we observed in different subclones from the same patient.

The genetic signature for FSHD shown in Fig.1a has recently been shown to exist in 1% of the general population, two orders of magnitude more than the prevalence of the disease [14]. This establishes that additional unknown factors contribute to this myopathy. Whether there are few or many remains to be determined. Our observations show that telomere length may be at least one of the contributing factors in this complex disease.

The increase in *DUX4* expression when myoblasts differentiate that we and others have observed remains unexplained. *DUX4* might be a developmental factor regulating the size of fetal muscles in the primate face and upper arms [13]. These very small primary fetal myofibers form the template for the adult musculature. Causing some of them to die might be a normal part of human development that modulates the size of particular adult muscles. The expression of *DUX4* in some unaffected biopsies [15] and during differentiation of normal cell cultures by us and others [15,46] would reflect a low level expression of this normal developmental program (since expression in one nucleus of a fiber containing only 3-4 nuclei could easily cause death of the entire fiber, rather than the complex focal degeneration seen in adult highly multinucleate muscle). The reduction in primary fibers would then lead to a reduction in the number of adult myofibers. This minor level of expression in normal cells would then be transformed into a major increase in FSHD as a result of the contractions and thus reduction of the repressor-insulator function of the D4Z4 repeats.

In conclusion, we demonstrated using isogenic clones with different telomere lengths that the expression of the toxic homeodomain protein DUX4-fl is inversely regulated by telomere length. We compared affected and unaffected siblings from two families (as well as additional samples), eliminating the family effects that can confound studies of FSHD [29]. FSHD appears to be the first example of a human disease in which classic TPE is able to affect the molecular pathogenesis, as judged by its effects in cultured cells. Future studies will need to extend these observations in vivo. Variations in telomere length may contribute to the variable penetrance of this complex disease, and may prove valuable in counseling of FSHD patients and their families.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

All authors were supported by the Sen. Paul Wellston Muscular Dystrophy Cooperative Research Center (US National Institutes of Health grant no. 5U54HD060848). Additional support was provided by the Austrian Science Fund and the American Federation for Aging Research (G.S.), AG01228 from the US National Institute of Aging

(W.E.W.), the Muscular Dystrophy Association (F.R.), and P50 CA70907 from the US National Institute of Health (J.W.S.). We thank Genila Bibat (Kenedy Krieger Institute, Baltimore, MD) and Daniel Perez (FSH Society, Lexington, MA) for subject recruitment and outreach, Naili Liu for technical assistance and Takako Jones for discussions. Reporter constructs were kindly provided by Silvere van der Maarel (Leiden University Medical Center, Leiden, the Netherlands) and Alexandra Belayew (University of Mons, Mons, Belgium), 38/03 MDC1A myoblasts by P. Schneiderat (Muscle Tissue Culture Collection at the University of Munich, German network on muscular dystrophies, partner of EuroBioBank and TREAT-NMD).

REFERENCES

1. Flanigan KM, et al. Genetic characterization of a large, historically significant Utah kindred with facioscapulohumeral dystrophy. Neuromuscul Disord. 2001; 11:525–9. [PubMed: 11525880]
2. Mostacciuolo ML, et al. Facioscapulohumeral muscular dystrophy: epidemiological and molecular study in a north-east Italian population sample. Clin Genet. 2009; 75:550–5. [PubMed: 19320656]
3. Padberg, GW. PhD thesis. Leiden University; Leiden, Holland: 1982. Facioscapulohumeral disease.
4. Padberg GW, et al. Facioscapulohumeral muscular dystrophy in the Dutch population. Muscle Nerve. 1995; 2:S81–4.
5. van Deutekom JC, et al. FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. Hum Mol Genet. 1993; 2:2037–42. [PubMed: 8111371]
6. Wijmenga C, et al. Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. Nat Genet. 1992; 2:26–30. [PubMed: 1363881]
7. Pandya S, King WM, Tawil R. Facioscapulohumeral dystrophy. Phys Ther. 2008; 88:105–13. [PubMed: 17986494]
8. Dixit M, et al. DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. Proc Natl Acad Sci U S A. 2007; 104:18157–62. [PubMed: 17984056]
9. Gabriels J, et al. Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. Gene. 1999; 236:25–32. [PubMed: 10433963]
10. Lemmers RJ, et al. A unifying genetic model for facioscapulohumeral muscular dystrophy. Science. 2010; 329:1650–3. [PubMed: 20724583]
11. Bosnakovski D, et al. An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. EMBO J. 2008; 27:2766–79. [PubMed: 18833193]
12. Kowaljow V, et al. The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. Neuromuscul Disord. 2007; 17:611–23. [PubMed: 17588759]
13. Snider L, et al. Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. PLoS Genet. 2010; 6:e1001181. [PubMed: 21060811]
14. Scionti I, et al. Large-scale population analysis challenges the current criteria for the molecular diagnosis of fascioscapulohumeral muscular dystrophy. Am J Hum Genet. 2012; 90:628–35. [PubMed: 22482803]
15. Jones TI, et al. Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: Evidence for disease modifiers and a quantitative model of pathogenesis. Hum Mol Genet. 2012
16. Shay JW, Wright WE. Role of telomeres and telomerase in cancer. Semin Cancer Biol. 2011; 21:349–53. [PubMed: 22015685]
17. Mondoux, M.; Zakian, VA. Telomere Position Effect: silencing near the end. In: De Lange, T.; Lundblad, V.; Blackburn, EH., editors. Telomeres. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 2006. p. 261-316.
18. Girton JR, Johansen KM. Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila. Adv Genet. 2008; 61:1–43. [PubMed: 18282501]
19. Gottschling DE, Aparicio OM, Billington BL, Zakian VA. Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription. Cell. 1990; 63:751–62. [PubMed: 2225075]
20. Hazelrigg T, Levis R, Rubin GM. Transformation of white locus DNA in drosophila: dosage compensation, zeste interaction, and position effects. Cell. 1984; 36:469–81. [PubMed: 6319027]
21. Baur JA, Zou Y, Shay JW, Wright WE. Telomere position effect in human cells. Science. 2001; 292:2075–7. [PubMed: 11408657]

22. Gao Q, et al. Telomeric transgenes are silenced in adult mouse tissues and embryo fibroblasts but are expressed in embryonic stem cells. Stem Cells. 2007; 25:3085–92. [PubMed: 17823235]
23. Koering CE, et al. Human telomeric position effect is determined by chromosomal context and telomeric chromatin integrity. EMBO Rep. 2002; 3:1055–61. [PubMed: 12393752]
24. Pedram M, et al. Telomere position effect and silencing of transgenes near telomeres in the mouse. Mol Cell Biol. 2006; 26:1865–78. [PubMed: 16479005]
25. Lou Z, et al. Telomere length regulates ISG15 expression in human cells. Aging (Albany NY). 2009; 1:608–21. [PubMed: 20157543]
26. Ottaviani A, et al. The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. PLoS Genet. 2009; 5:e1000394. [PubMed: 19247430]
27. Ottaviani A, Schluth-Bolard C, Gilson E, Magdinier F. D4Z4 as a prototype of CTCF and lamins-dependent insulator in human cells. Nucleus. 2010; 1:30–6. [PubMed: 21327102]
28. Stadler G, et al. Establishment of clonal myogenic cell lines from severely affected dystrophic muscles - CDK4 maintains the myogenic population. Skelet Muscle. 2011; 1:12. [PubMed: 21798090]
29. Homma S, et al. A unique library of myogenic cells from facioscapulohumeral muscular dystrophy subjects and unaffected relatives: family, disease and cell function. Eur J Hum Genet. 2012; 20:404–10. [PubMed: 22108603]
30. Ramirez RD, et al. Bypass of telomere-dependent replicative senescence (M1) upon overexpression of Cdk4 in normal human epithelial cells. Oncogene. 2003; 22:433–44. [PubMed: 12545164]
31. Gabellini D, Green MR, Tupler R. Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell. 2002; 110:339–48. [PubMed: 12176321]
32. Klooster R, et al. Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. Eur J Hum Genet. 2009; 17:1615–24. [PubMed: 19809486]
33. Rijkers T, et al. FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. J Med Genet. 2004; 41:826–36. [PubMed: 15520407]
34. Hewitt JE, et al. Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum Mol Genet. 1994; 3:1287–95. [PubMed: 7987304]
35. Kawamura-Saito M, et al. Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. Hum Mol Genet. 2006; 15:2125–37. [PubMed: 16717057]
36. Geng LN, et al. DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Dev Cell. 2012; 22:38–51. [PubMed: 22209328]
37. Luria SE, Delbruck M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. Genetics. 1943; 28:491–511. [PubMed: 17247100]
38. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992; 34:1–14.
39. Decary S, et al. Replicative potential and telomere length in human skeletal muscle: implications for satellite cell-mediated gene therapy. Hum Gene Ther. 1997; 8:1429–38. [PubMed: 9287143]
40. Graakjaer J, et al. The relative lengths of individual telomeres are defined in the zygote and strictly maintained during life. Aging Cell. 2004; 3:97–102. [PubMed: 15153177]
41. Jeanclos E, et al. Telomere length inversely correlates with pulse pressure and is highly familial. Hypertension. 2000; 36:195–200. [PubMed: 10948077]
42. Londono-Vallejo JA, DerSarkissian H, Cazes L, Thomas G. Differences in telomere length between homologous chromosomes in humans. Nucleic Acids Res. 2001; 29:3164–71. [PubMed: 11470873]
43. Renauld H, et al. Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and by SIR3 dosage. Genes Dev. 1993; 7:1133–45. [PubMed: 8319906]

44. van der Maarel SM, Tawil R, Tapscott SJ. Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence. Trends Mol Med. 2011; 17:252–8. [PubMed: 21288772]
45. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet. 2007; 8:272–85. [PubMed: 17363976]
46. Tassin A, et al. DUX4 expression in FSHD muscle cells: how could such a rare protein cause a myopathy? J Cell Mol Med. 2013; 17:76–89. [PubMed: 23206257]
47. Zhu CH, et al. Cellular senescence in human myoblasts is overcome by human telomerase reverse transcriptase and cyclin-dependent kinase 4: consequences in aging muscle and therapeutic strategies for muscular dystrophies. Aging Cell. 2007; 6:515–23. [PubMed: 17559502]
48. O'Gorman S, Dagenais NA, Qian M, Marchuk Y. Protamine-Cre recombinase transgenes efficiently recombine target sequences in the male germ line of mice, but not in embryonic stem cells. Proc Natl Acad Sci U S A. 1997; 94:14602–7. [PubMed: 9405659]
49. Herbert BS, Shay JW, Wright WE. Analysis of telomeres and telomerase. Curr Protoc Cell Biol. 2003 **Chapter 18**, Unit 18 6.
50. Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language. 2008; 59:390–412.
51. MacKay, DJC. Information Theory, Inference and Learning Algorithms. Cambridge University Press; 2003.

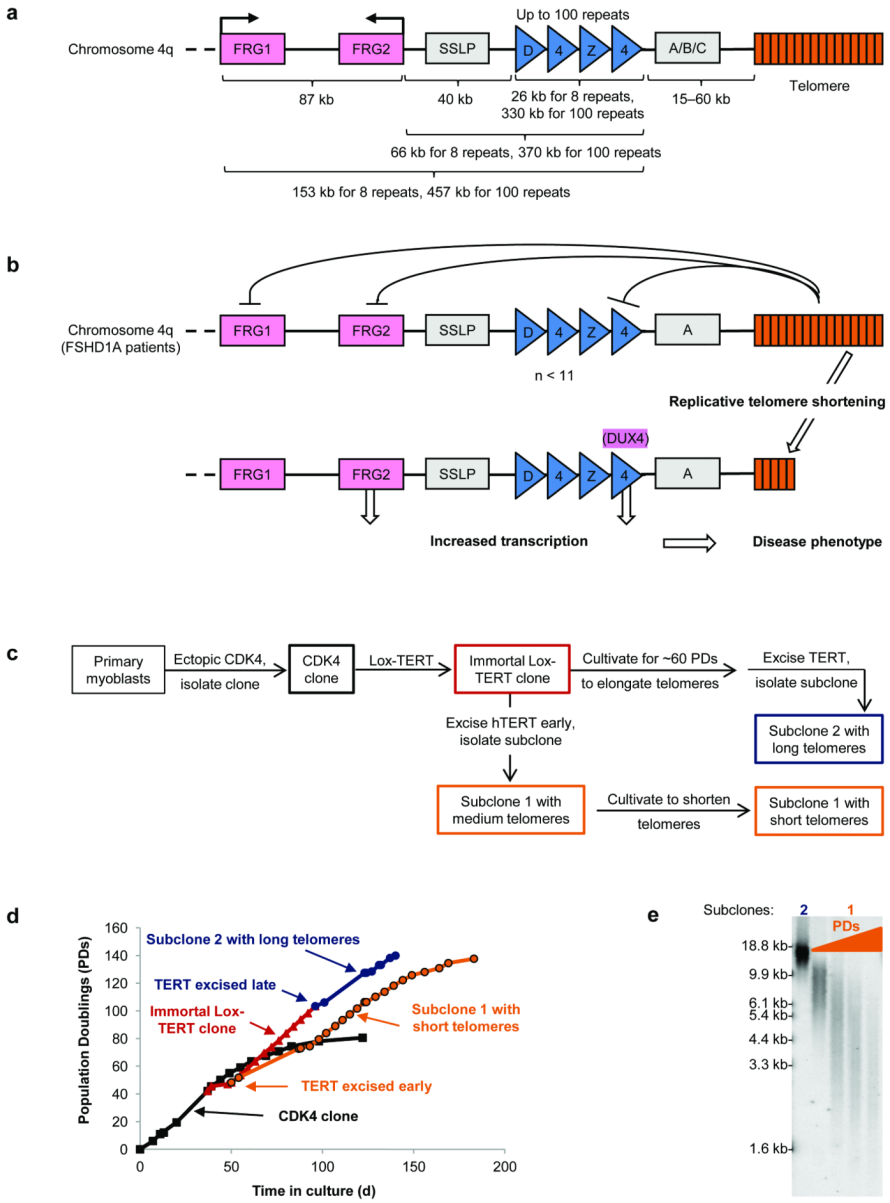

Figure 1.

Telomere position effect and FSHD.

(a) Schematic depiction of chromosome 4q. The normal D4Z4 repeat array consists of ~100 units, each 3.3 kb in length. Borderline FSHD1A patients carry <11 repeats, and <8 is considered diagnostic. Each D4Z4 (blue triangles) encodes a DUX4 homeobox protein. Only the A-haplotype (not B or C), which includes a polyA signal (PAS) for the most telomeric *DUX4*, is associated with FSHD. Several simple sequence length polymorphisms (SSLP) are permissive for FSHD. Distances betweenD4Z4 repeats and the promoters of *FRG1* and *FRG2* are shown for contracted versus normal D4Z4 numbers. (b) We hypothesize transcription is promoted by both short D4Z4 repeat numbers and telomere shortening. (c) Experimental design for the derivation of isogenic myogenic subclonal cell lines with long and short telomeres. (d) Growth curves of a clonal cell line from patient 15Abic (family 15, FSHD patient A, biceps) and the derivation of different subclones. Code: *CDK4* only (black squares), after introducing *TERT* (red triangles), early *TERT* excision

(orange line with black circles), late *TERT* excision (dark blue circles). Solid lines represent PDs during cloning when no analysis was possible. (e) Telomere restriction fragment analysis gel showing telomere lengths of subclone 2 at PD 137 and subclone 1 at PDs 80, 114, 115, and 131.

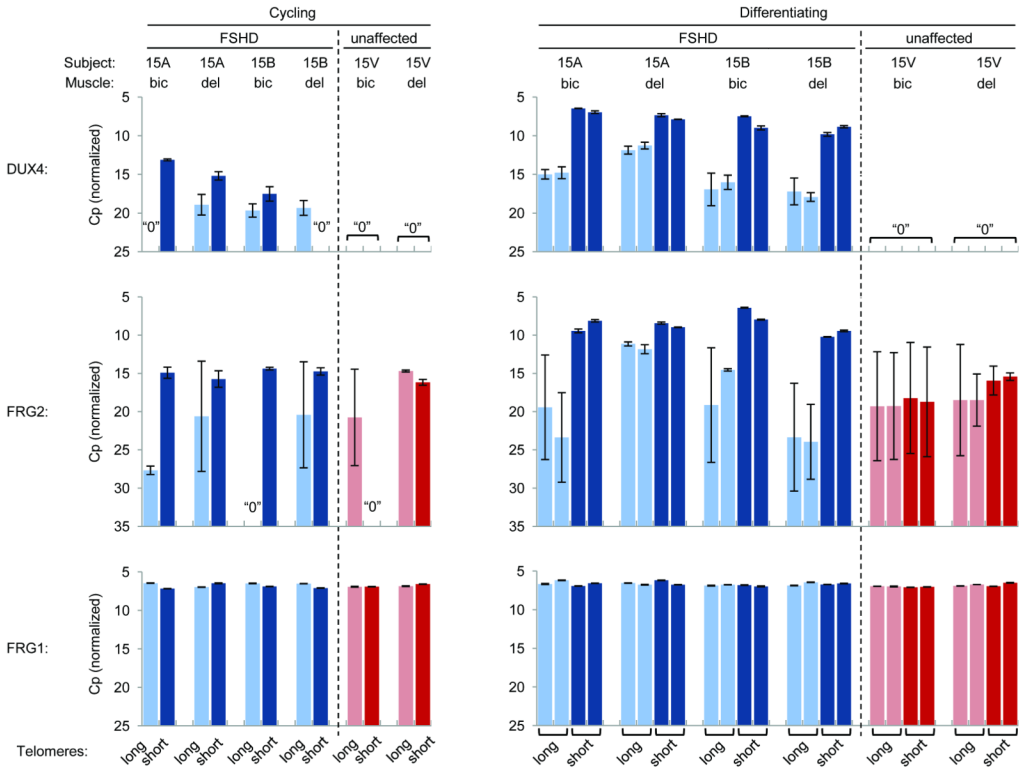

Figure 2.

DUX4, *FRG2* and *FRG1* expression in cells with short and long telomeres.

qRT-PCR was performed on subclones from two muscles (biceps and deltoid) of two FSHD siblings (A and B) from family 15 (represented in blue: 15Abic, 15Adel, 15Bbic, 15Bdel) and subclones from the same two muscles from their sibling (V, in red: 15Vbic, 15Vdel) who had normal numbers of D4Z4 repeats, all with long (light colors) and short (dark colors) telomeres. Each cell line was analyzed under growth conditions ("Cycling", left panel), and after 4 and 7 days in differentiation medium ("Differentiating", right panel: first bar of each cell line represents day 4,second bar represents day 7 in differentiation medium). Samples in which the Cp values were equal to or more than a water-only control were designated as "0", to indicate that they had been analyzed and under these RT-PCR conditions there were no detectable transcripts. *FRG2* RT-PCR was optimized to specifically detect transcripts from chromosome 4q (Supplementary Fig.3b). Shown are the mean Cp values and standard deviation (error bars) of 3 technical replicates of a representative experiment; each subclone has been independently analyzed 2-5 times and all values are shown in Supplementary Fig. 2a and b. The Y-axis is plotted as decreasing values so that an increase in mRNA appears as a taller bar. Each Cp value represents a doubling of the input, so the difference between (for example) 15 and 20 represents a 2^5=32-fold difference.

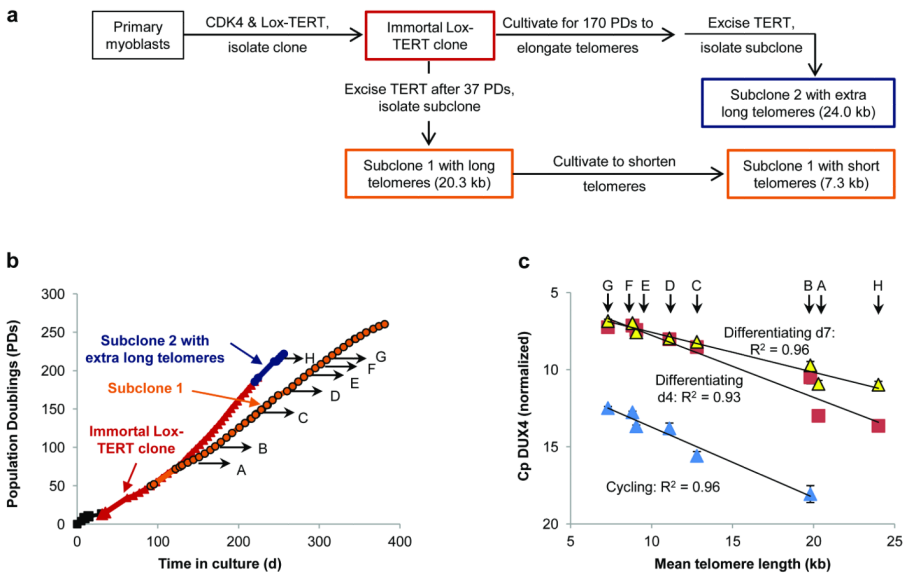

Figure 3.

Linear correlation of *DUX4* expression and telomere length.

(a) Generating isogenic subclones with various telomere lengths. (b) Growth curves for 01Adel (family 01, affected patient A, deltoid) and subclones. *CDK4* only (black squares), after introducing *TERT* (red triangles), early excision of *TERT* (orange line with black circles), late excision of *TERT* (dark blue circles). Solid lines represent PDs during cloning and expansion when no analysis was possible. Subclone 1 analyzed at PDs 80 (A), 100 (B), 146 (C), 172 (D), 196 (E), 206 (F), and 216 (G), with mean telomere lengths of 20.3, 19.8, 12.8, 11.1, 9.1, 8.8, and 7.3 kb. Subclone 2 analyzed at PD216 (H, mean telomere length 24.0 kb). (c) *DUX4* Cp values versus telomere length for subclone 1 (A-G) and subclone 2 (H) of 01Adel at the PDs of Figure 3b. *DUX4* for cells differentiating for 7 days (yellow triangles), differentiating for 4 days (red squares), and undifferentiated cycling myoblasts (light blue triangles). Correlation coefficients R^2 were calculated with Excel.

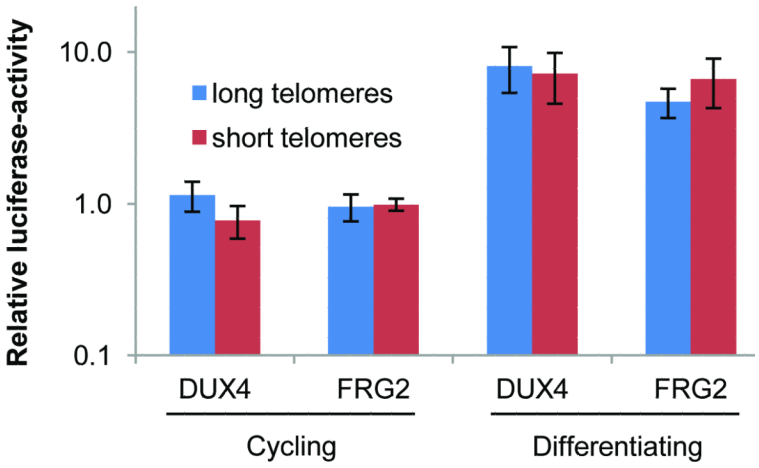

Figure 4.

Trans-acting factors do not change with telomere length.

DUX4 and *FRG2* promoter constructs were transiently transfected into 15Bdel FSHD cells with long and short telomeres. Relative luciferase activity of cell lysates show a change with differentiation but not telomere length. +/− SD of triplicate measurements of two independent experiments.

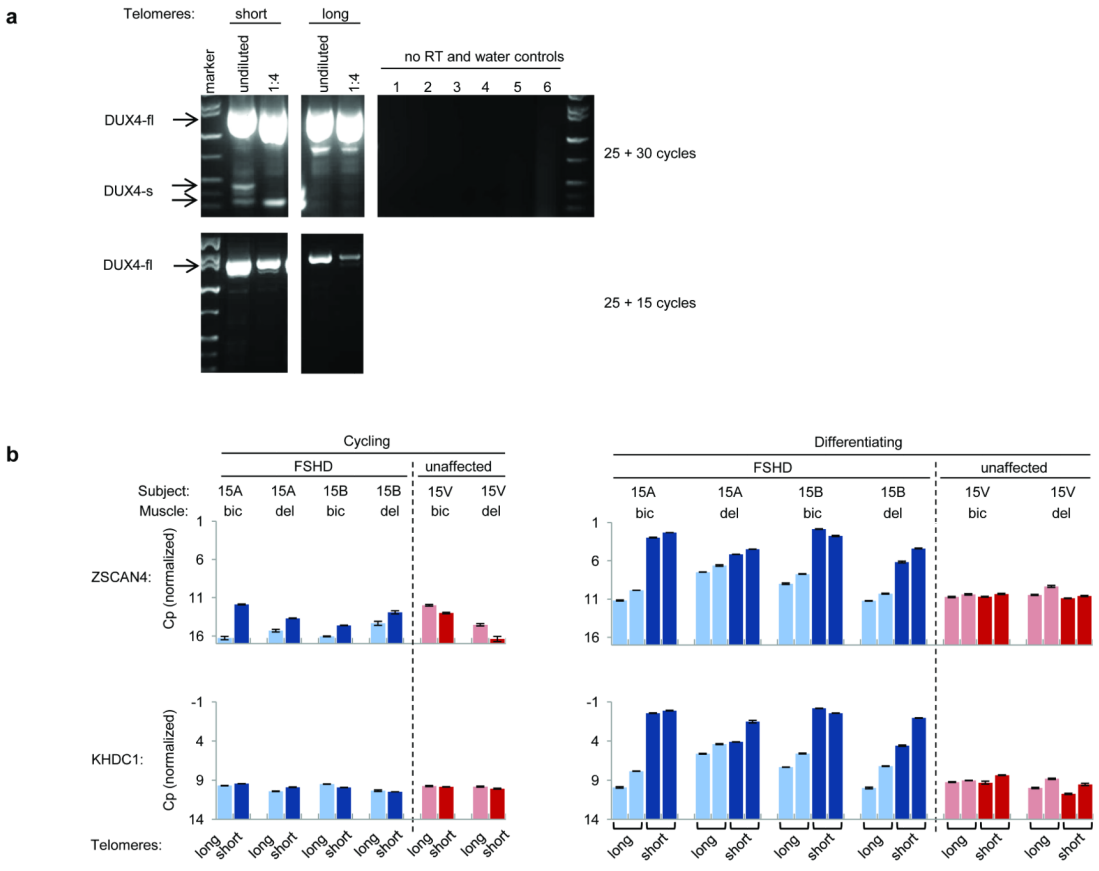

Figure 5.

Functional DUX4-fl is upregulated in cells with short telomeres.

(a) Agarose gel image after RT-PCR for *DUX4* splice variants in 01Adel with short or long telomeres after 7 days in differentiation medium. The identity of the bands indicated by arrows has been confirmed by sequencing. (b) qRT-PCR for *DUX4* targets *ZSCAN4* and *KHDC1* in subclones of cohort 15 with long and short telomeres (same samples as Fig. 2). +/− SD of three technical replicates.

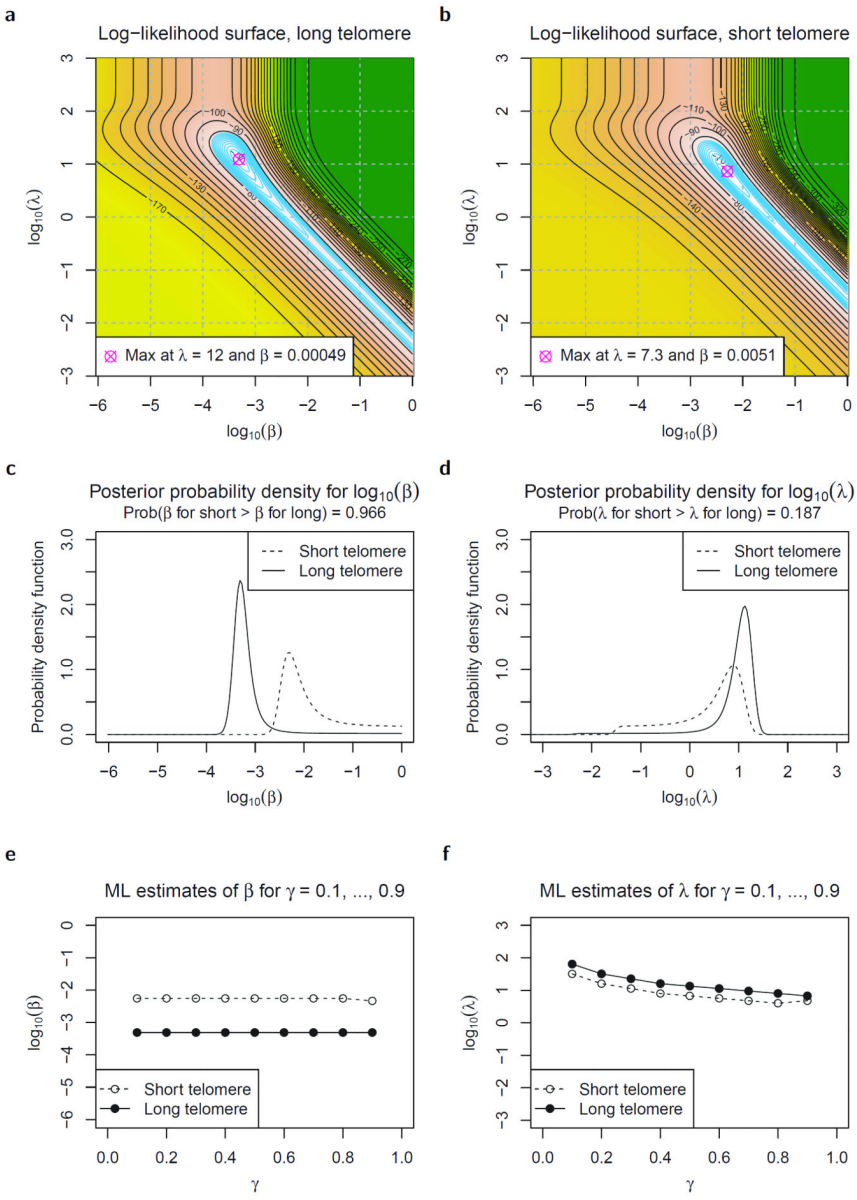

Figure 6.

Estimating the fraction of cells expressing *DUX4*, and counts per cell that do express *DUX4*.

(a) and (b) show the log-likelihood surfaces (base 2) of parameters β and λ for a zero-inflated Poisson (ZIP) model, for long and short telomeres respectively, with γ (probability of detecting one transcript per reaction) fixed at 0.5.

A magenta symbol (crossed circle) indicates the maximum likelihood values of λ and β. Black contours are spaced 10 units apart, and cyan contours 1 unit apart. (The computations omit certain combinatorial factors that are independent of the parameters, so these log-likelihoods are defined up to an additive constant). Note that each line of slope −1 in the ($\log_{10}(\beta)$,$\log_{10}(\lambda)$)-plane corresponds to the ZIP models with fixed product $\beta\lambda$, which represents the mean number of *DUX4* transcripts per cell when including both the Poisson and zero subpopulations. (c) and (d) show the marginal posterior probability densities for β and λ, respectively, with γ fixed at 0.5. Solid lines represent densities for long telomeres

and dashed lines for short telomeres. The densities are well-separated for β (c) but have substantial overlap for λ (d), suggesting a difference in the fraction of cells expressing *DUX4* rather than a difference in the amount of *DUX4* per cell that does express *DUX4*. (e) and (f) show that the same holds for other values of γ in the range from 0.1 to 0.9, although the estimated value of λ increases as γ decreases