Supplemental

Table of Contents

Material and methods 2
- Cell culture 2
- 4q35 Hi-C: 2
- 4q35 Hi-C analysis: 3
- Chromatin conformation Capture (3C): 3
- 3D-FISH: 3
- RT-PCR: 5
- RNA extraction and gene expression (ddPCR): 5
- Methylation assay by sodium bisulfite sequencing 5
- IHC: 6

Primers sequences used 7

Supplemental Figure Legends 9
- Supplemental S1. Establishment of the Long and Short telomere isogenic clones. 9
- Supplemental S2. Measure of telomere length 9
- Supplemental S3. Strategy for the high-resolution mapping of DNA interactions based on the unbiased Hi-C technique. 10
- Supplemental S4. Hi-C quality control. 10
- Supplemental S5. Locus specific Hi-C screen. 10
- Supplemental S6. Organization of the 4q35.2 locus in control and FSHD myoblasts with long or short telomeres. 11
- Supplemental S7. Organization of the 4q35.2 locus in control myoblasts with long (upper panel) or short telomeres (lower panel). 11
- Supplemental S8. 3C Validation and Normalization. 12
- Supplemental S9. 3D DNA-FISH of interactions at 1p and 4q loci. 12
- Supplemental Figure S10. Gene expression analysis of *SORBS2*. 13
- Supplemental S11. RT-PCR in Control myoblast. 14
- Supplemental S12. Telomere shortening in FSHD cells induces alternative splicing modifications of *SORBS2* in Deltoid muscle biopsies. 14
- Supplemental Figure S13. Gene expression analysis of 4q35 genes. 15

Supplemental Figures 15

Material and methods

Cell culture

CD56-positive isolated myoblasts from primary muscle biopsies from the Wellstone foundation were infected with retroviral constructs containing CDK4 and a floxable human TERT as described (Robin et al. 2014). Individual clones were then treated with Cre-recombinase at different times and cultivated for approximately same numbers of population doublings to generate isogenic subclones with different telomere lengths (Fig. 1). Telomere length was assayed by Telomere Restriction Fragment (TRF) analysis (Supplemental S1).

Briefly, human myoblasts were seeded in dishes coated with 0.1% pigskin gelatin in 4:1 Dulbecco modified Eagle medium/Medium 199 supplemented with 15% FBS, 0.02M HEPES, 1.4mg/l vitamin B12, 0.03mg/l $ZnSO_4$, 0.055mg/l dexamethasone, 2.5µg/l hepatocyte growth factor and 10µg/l basic fibroblast growth factor. Cultures were maintained in a 2-5% oxygen environment and passaged at ~60% confluency. Population doublings (PDs) were calculated as PD = ln[(final number of cells)/(initial number of cells)]/ln(2).

Myogenicity of the cells was verified by myotube formation following a change to differentiation media (2% horse serum in 4:1 Dulbecco modified Eagle medium: Medium 199) when 70-90% confluent.

4q35 Hi-C:

The Hi-C experiment was performed as described (van Berkum et al. 2010) on myoblasts with long telomeres, modified to enrich for interactions involving 4q35. The set of RNA probes for enrichment through pull-down was designed against the last 5Mb of chromosome 4, at the 4q35 locus. In brief, proteins within intact cells were crosslinked for 10min in a 1% formaldehyde (methanol free) solution. The reaction was then stopped with Glycine, and cells were lysed in solution completed with protease inhibitor (10mM Tris-HCl pH 8.0, 10mM NaCl, 0.2% Ige Cal CA 630 (NP-40) + 50µl protease inhibitor Sigma #P8340). DNA was then digested with the *Hind*III restriction enzyme (400 Units) over-night leaving an overhang that is filled in by including a biotinylated dCTP (Biotin-14-dCTP, Invitrogen). Blunt ends were then ligated under very dilute conditions (8ml final volume) that favor ligation of ends held in close proximity, i.e. DNA elements that are also in proximity in vivo, which has been maintained through the crosslinking step. Ligation products containing biotin-labels were selectively pulled-down using streptavidin. Subsequent addition of PCR primers and amplification produced the material of a full Hi-C library. The library was then hybridized to

RNA biotinylated probes (targeting the 4q35 locus), re-amplified after purification and then analyzed by paired-end sequencing (Illumina, HiSeq-150).

4q35 Hi-C analysis:

FastQC (Gentleman et al. 2004) was used to check the quality of raw reads. Reads were aligned individually to the UCSC hg19 assembly of the human reference genome using Bowtie2 (Langmead and Salzberg 2012). All subsequent analysis was performed using HOMER (Hypergeometric Optimization of Motif EnRichment), a suite of tools for analyzing genomic sequencing data including Hi-C data (Heinz et al. 2010). Aligned reads were paired and duplicates were removed. Parameters were set to filter out self-ligations, include only reads within 1000bp of a *Hind*III locus, and ensure a minimum distance of 20Kb between read pairs. A 100Kb resolution background model was generated using all paired reads and False Discovery Rate was controlled using the Benjamini correction. Significant interactions ($p<0.05$) were visualized with Circos using files generated with the Circos option in HOMER.

Chromatin conformation Capture (3C):

3C samples were prepared as described above (Hi-C). Quantification used the ddPCR technology (Biorad) following manufacturers' instructions with the following modifications in the PCR profile set-up: denaturation 95ºC – 10min, followed by 40 cycles of 94°C – 30 sec, 60°C – 30s annealing, 72°C – 30s extension. PCR was stopped with a final 98°C – 10 min step. Plates were then read on the ddPCR droplet reader. All primers were HPLC purified (Supplementary Table S1), all PCR pairs used were tested and amplicons were Topo-cloned and sequenced in order to validate the sequences amplified. To quantify samples and normalize data, series of 11 PCRs were done around the ENr313 region (ENCODE region 313) testing for proximity effect: an average of the 11 assays was made and used to normalize all future 3C target interactions. All experiments were done using biological triplicates, and ddPCR duplicates (total of 6 values per data point). Statistical analyses for all ddPCR measures were performed using Prism statistical tool and results analyzed using the 2 way ANOVA Holm sidak's comparison test with alpha set at 0.05.

3D-FISH:

Probes for FISH were produced following manufacturer's instructions (nick translation kit, Abbott molecular). Good nick translation reactions occurred when the DNA smear ranged around 300bp. The purified probes were re-suspended in a 50µl mixture of Hybridization solution B (Cytocell). Templates were chosen with the UCSC genome browser and ordered from the CHORI institute (SORBS2, RP11-1L10; proximal region of FRG1, G248P88732C10). BACS and Fosmids were verified by PCR and/or metaphase spreads

Slide preparation: 8000 cells were plated in 4 chamber-well slides (VWR) previously coated with 0.1% gelatin. The next day cells were washed once in 1X PBS, chambers were removed, and the slides incubated 5min in 1X PBS at RT and fixed in a 4% freshly made PFA for 10min at RT under a fume hood. After 3x5min 1x PBS washes, nuclei were permeabilized in 0.5% Triton X-100 in PBS for 20 min at RT. A first incubation in 20% glycerol, PBS for 30min at RT was done prior to 6 cycles of successive incubations between liquid nitrogen and glycerol. Following 6 cycles, slides were again incubated for 1min in 20% glycerol at RT, followed by 3x5min washes with 1X PBS and a HCl 0.1M incubation for 10min. Finally slides were incubated 5min in 2X SSC and stored at 4°C in 50% formamide pH 7 in 2X SSC in a Coplin jar wrapped in aluminum foil.

Slide hybridization: after a minimum of 5 days storage at 4ºC in formamide, slides were used for hybridization. Excess solution was removed from the slide and a mixture of 3µl of each probe with 4µl of Hybridization solution B (Cytocell) was used for each well (10µl total). The slides were covered with a coverglass (FisherBrand) and sealed with rubber cement. The slides were then co-denatured (with the probe mix) at 80ºC for 5min in a hybridization oven prior to an overnight incubation at 37ºC in a humidification chamber. The next day, rubber cement was removed and the slides were incubated for 5min at 70°C in 0.5X SSC, 0.1% SDS. Slides were then sequentially washed in PBS-Tween 20 0.001% for 5min at RT and then in 1X PBS for 10min at RT. Excess solution was removed and a 10µl DAPI (0.125µg/ml in Vectashield) solution was dropped on each slide.

Image processing: Images were acquired with a confocal scanning laser system (LSM 780 from Zeiss, Germany). A 40X plan-APOCHROMAT oil immersion objective was used to capture optical sections at intervals of 0.24µm. The pinhole was set to one Airy with optical slices in all wavelengths with identical thickness. Images were then generated into .lsm files and were processed through the Imaris software (Bitplane AG). After 3D reconstruction, at least 30 nuclei (60 alleles) were examined per condition, the volume of the probes and distance between their gravity centers (distance between the closest probes of each target) were determined and used for statistical analysis.

RT-PCR:

RT-PCRs were done using a 2µl cDNA mixture from a 1µg RNA RT reaction, split into 5 PCRs reactions of 20; 25; 30; 35 and 40 PCR cycles, respectively. Reactions were then separated by electrophoresis run on a 1.5% agarose gel. Amplicons of *SORBS2* transcripts were validated by Sanger sequencing.

RNA extraction and gene expression (ddPCR):

RNA was extracted from 1x10^6 cells using a column-based kit (RNeasy Plus, Qiagen) following the manufacturer's instructions. RNA was quantified using a Nanodrop and 500ng of total RNA was reverse transcribed with a first-strand cDNA synthesis kit (iScript, BioRad) using an optimized combination of random hexamers and oligo-dT. Following cDNA synthesis, the samples were diluted one to four in water and stored at -20ºC until further use. mRNA quantification for gene expression analysis was pursued using manufacturer's instructions (ddPCR, BioRad). In brief, 1µl of diluted cDNA was used in a 19µl reaction composed as follows: 10µl of ddPCR Master mix (2x), 250nM of each primers and 0.1µl of the chosen Universal Probe Library (UPL, Roche) (10µM), the mix was brought up to a final volume of 19µl with DNase free water. After droplet generation, the sealed plate was put in a thermo-cycler with the following cycles: 95ºC heat activation for 10 min followed by 40 cycles of 30 seconds at 95ºC for denaturation and 1 min at 60ºC for annealing-extension steps. The PCR reaction was stopped with a final deactivation step at 98ºC for 10 min. The plate was transferred and loaded into the ddPCR QX100 reader. Every assay was done in duplicate using RT triplicate. Statistical analyses for all ddPCR measures were performed using Prism statistical tool and results analyzed using the ANOVA Tukey's comparison test.

Methylation assay by sodium bisulfite sequencing

2µg of genomic DNA was denatured for 30 minutes at 37°C in NaOH 0.4N and incubated overnight in a solution of sodium bisulfite 3M pH5 and 10mM hydroquinone using previously described protocol (Gaillard et al. 2014). Converted DNA was then purified using the Wizard DNA CleanUp kit (Promega) following manufacturer's recommendation and precipitated by ethanol precipitation for 5 hours at -20°C. After centrifugation, DNA pellet was resuspended in 20µl of water and stored at -20°C until use. Converted DNA was then amplified using primer sets designed with the MethPrimer software (Li and Dahiya 2002) avoiding the presence of CpGs in the primer sequence in order to amplify methylated and unmethylated DNA with the same efficiency. Amplification was carried out using the High Fidelity Taq polymerase (Roche) according to manufacturers' instructions. After initial denaturation at 94°C for two minutes, amplification was done at 94°C for 20 seconds, 54°C for 30 seconds, 72°C for one minute for 10 cycles and at 94°C for 20 seconds, 54°C for 30 seconds, and with an amplification step of 4 minutes and 30 seconds for the first cycle and an implement of 30 seconds at each subsequent cycle for 25 cycles. At the end of the program, a final extension step at 72°C for 7 minutes was done.

PCR products were then purified using the Wizard SV gel and PCR Purification system (Promega), resuspended in 50µl of water and cloned using the pGEM®-T Easy Vector cloning kit (Promega). Randomly selected colonies were grown overnight at 37°C with ampicillin selection, and colonies were PCR amplified directly using T7 or SP6 primers. For each sample and region, at least ten randomly cloned PCR products were sequenced according to Sanger's method by Eurofins MWG Operon (Ebersberg, Germany) with either SP6 or T7 primers.

Sequences were analysed with the BiQ Analyser software (Bock et al. 2005) and the average methylation score was calculated as the number of methylated CpGs for the total number of CpGs in the reference sequence.

IHC:

Paraffin-embedded tissues were cut in to 5-µm sections. Sections were then deparaffinized, rehydrated, and then antigens retrieved with citrate buffer (10 mmol/l sodium citrate, pH 6.0; 0.05% Tween 20) in a pressure cooker. Endogenous peroxidase, biotin, and proteins were sequentially blocked with solutions of 3% hydrogen peroxide (Sigma-Aldrich), the Avidin/Biotin Blocking Kit (Vector Laboratories), and 10% bovine serum albumin (Vector Laboratories). Primary antibody (Epitomics, RB19714) was diluted in 1XTBST with 5% bovine serum albumin and then sections were incubated overnight at 4°C. After three washes with 1XTBST, secondary antibody and ABC reagent were applied using the VECTASTAIN ABC Kit (Vector Laboratories) following the manufacturer's protocol. Tissue sections were then incubated with ImmPACT DAB peroxidase substrate (Vector Laboratories), counterstained with methyl green, and then dried overnight before mounting coverslip. Images were captured with Axiovision software v4.6.3 on Axioskop 2 plus microscope mounted with a AxioCamHR color camera (Carl Zeiss Microscopy) using Plan-APOCHROM 20X and 40X objectives.

Primers sequences used

Primer LIST		Seq	UPL
control HiC			
HindIII -1	F	GTTCATCTTGCTGCCAGAAATGCCGAGCCTG	83
HindIII-2	R	ATCCCAGCTGTCTGTAGCTTTAGAAAGTGGG	
GAPDH1	F	GCCCAATACGACCAAATCTAA	
GAPDH2	R	ATTGTTGCCATCAATGACCC	
NcoI-1	F	ACCTGTTGTTTAATGAAGGGGCTCAGAAGC	
NcoI-2	R	GTTTGCAGTGTGCTGTGCAGCATGTGTGTA	
HiC library Primers			
PE Adaptors 1	F	P-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG	
PE Adaptors 2	R	ACACTCTTTCCCTACACGACGCTCTTCCGATCT	
PE PCR Primer 1.0	F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	
PE PCR Primer 2.0	R	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT	
ddPCR SORBS2			
3' UTR	F	CTCGGGTTTACCTGCTTCCT	37
	R	TGTTTGCTTTACAATCTCTGTCA	
Ex17-Ex18	F	CAAGATAAAAGATCACCAACTC	22
	R	TGTGTGACTGGTTATCAAAGAGC	
Ex30-Ex31	F	TGATCGCATAAATCCAGATGAC	
	R	TGAACATAGTCCAGAGGCTTTTT	
5'UTR	F	TTAAACACATTCCCTTGTCCTG	
	R	TTCTTCTTGAGGTCCCCACA	
Supplemental			
Ex8-Ex9	F	GGACATCTCGTTTTCAGTGAATC	61
	R	GTCTGCAAAGGGATGAAAATG	
Ex19-Ex20	F	GAACCGAGCCACTTTCTCAC	59
	R	CTGTTATCTGGCTCGGAAGG	
Ex21-Ex22	F	TAGAACGGACTTCTCATGAATGTTA	
	R	TGTTTCATTCAAATGGAAAAACAG	
Ex39-Ex40	F	TTAAGGAGGAACCTGCTGTCC	
	R	CCACCTTGAATATTTTCATGAGTAAAC	
ddPCR Gene Expression			
Tmem8C - Myomaker	F	GGAGAACAGAGGCCTCCAC	22
	R	AGACAAGCCGGGTCCATT	
MyoD	F	CACTACAGCGGCGACTCC	70
	R	TAGGCGCCTTCGTAGCAG	
Fam149A	F	TGCAAAGAAAACACCAGTGC	75
	R	ATGTGGTCCACGCCAGTC	
TRIML1	F	AGAGTGAAACTGTGCCAGGAA	68
	R	TTCCTGTTCTAGTAGCTGGAGTTG	
ZFP42	F	GGCCTTCACTCTAGTAGTGCTCA	62

	R	CTCCAGGCAGTAGCTGATCTGAGT	
FAT1	F	AGACCAAGTTCATTACTGCATCG	45
	R	TCCAGGTTCAGGGTTTTCAG	
FRG1	F	GGGAGCTCCACACAAAGAAGT	65
	R	CAGACTTCAGGGCGATTCTG	
PDLIM3	F	CCCAAACCTTTCGTGATCC	86
	R	CCACTGCCACAGTCAATCC	

3C Primers

FRG1-SORBS2 region

SORBS2#1	F	GGTGAGGTAGACATAAATTCAGCA
	R	ACCCTTGATTGGATTCAGGA
SORBS2#2	F	TCCCAAATGTCCTGTCAGTG
	R	GGCACCGTTTTAAGGGTATG
SORBS2#3	F	CATGAGTCTGTACAAACAAAACAAGA
	R	AAGAACAAAATATGGGCAATACA
SORBS2#4	F	GGAAGGCAGTCTCATTTCTCA
	R	TATTCTTTGGGGCGTTTGAC
SORBS2#5	F	AAGAATCACCCGGGAATTTG
	R	ATGCAGGCCAACGGTTCT
SORBS2#6	F	TTCTGACTAGCCGCATTGAG
	R	GGGCTCATAGGAATAATGCAC
SORBS2#7	F	TTCATTCTGTAGCTTATATATTTCGTG
	R	GCTTTTTCCAATAACATTAAGGAAGT
Prom region FRG1#Fix	F	gctaccttgccattatctcca
	R	caaagggaaagacgaacagg

† The Prom Region FRG1#F primer did not have a UPL in it. In order to detect it by ddPCR, the reverse complentary sequence of the #52 UPL (gggaggag) was added to the 5'end of the F primer

Enr313- 3C control region

Enr313#1	F	TGCTTTAAGTTCCTGGGGTAAA
	R	TCACTCACGTAGCCTGCACT
Enr313#2	F	AGGGAAGACTGGGAAAATCC
	R	GCATCCTAAAGGGTTGTAACTCA
Enr313#3	F	GAGACTACTGTGAACACCTCTAGGC
	R	TTACTGACTTAATTTCAGAGCTTGTT
Enr313#4	F	CCCACTTTTGCTTCAACACA
	R	TGCCTTTGTTTTACTCCCAAA
Enr313#5	F	CATCTGCCACGTTATGTAAAGC†
	R	TATCTTGCCACCATATTCAAGG
Enr313#6	F	GGGGACATGAGCTCCAGAA
	R	TCCTTGTGGGGAAATTGTGT
Enr313#7	F	TGTAGTTGCGATTTTTAAGAAACATT
	R	CAGTTTATGAACAAAGAGGGAACA
Enr313#8	F	GGCTGTGTACTACTCATAGATTCTGG
	R	TGGGAACAACCCAACTTTGT
Enr313#9	F	TTCCCCACATGAAGTTTGC
	R	CATTTTCAGTCTAACTGTCCTTTGA
Enr313#10	F	ACCTTTGTCATTGGCTCATGT

† The ENr313#5 F primer did not have a UPL in it. In order to detect it by ddPCR, the reverse complentary sequence of the #52 UPL (gggaggag) was added to the 5'end of the F primer

	R	TCATTTGTAAAACAAGAATCATCTCA
Enr313#11	F	TGTTTGTGCTATCTTGCTTAAAATG
	R	GTTGTGTGTGCATGTTTGGA
Enr313#12	F	GGTGCAGGCTGAGAATAAGA
	R	CCACAGATAGACTCTGGGGATCT

Supplemental Figure Legends

Supplemental S1. Establishment of the Long and Short telomere isogenic clones.

a) Isogenic clones with long and short telomeres were produced as previously reported (Stadler et al. 2013; Robin et al. 2014). Successive infections with retroviruses encoding CDK4 and a floxable hTERT cassette were done on primary cells derived from biopsies. Early excision of the hTERT was done to generate the ‘short’ telomere subclones and later excisions yielded subclones with long telomeres. For subsequent experiments, we used subclones with similar Population Doublings (PD). b) Examples of growth curves for isogenic clones from immortal populations are shown for FSHD and Healthy myoblasts. No differences in growth rates were detectable at the time point chosen for the experiments between the different conditions. PDs were registered and cells were analyzed for at least 30 PDs prior to senescence when growth rate decreased. c) We analyzed DNA damage response by counting the number of γH2AX foci per nucleus in cells with long and short telomeres (FSHD and control myoblasts). As positive control, we used cells subjected to UV radiation. No increase in DNA damage was seen in the short telomere cells. **** = $p<0.0001$, ANOVA.

Supplemental S2. Measure of telomere length

Telomere shortening was monitored as a function of population doublings and telomere length was measured by TRF. TRF shown has an example of early time points, and late replication, when cells are entering senescence. Population doublings (PDs) are reported as the number of cell divisions calculated since the myoblasts isolation form biopsy. To note, hTERT removal for the FSHD myoblast for the long telomere clones was done at PD 111.

Supplemental S3. Strategy for the high-resolution mapping of DNA interactions based on the unbiased Hi-C technique.

Briefly, purification of the biotinylated ligation products is followed by PCR amplification, creating the whole genome Hi-C library, which doesn't include biotin. Biotinylated probes specific to the 4q terminus were used to enrich the Hi-C library for interactions involving 4q35.2. The resulting library holds the same parameters as the Hi-C library but is essentially composed by sequences involving at least one end from 4q35.2 (Green DNA strand).

Supplemental S4. Hi-C quality control.

PCR ligation control (left) and digestion (right) to validate successful biotin integration. Hi-C preparation will result in the creation of a *Nhe*I restriction site instead of an *Hind*III site (due to the fill-in step). Primers were designed around a *Hind*III restriction site. PCR amplicons from one 3C (control) and 4 Hi-C preparations were digested by either *Hind*III or *Nhe*I. The 3C preparation does not include the fill-in step, consequently no *Nhe*I site is created.

Supplemental S5. Locus specific Hi-C screen.

Using biotinylated probes, the Hi-C technique can be used to investigate in an unbiased fashion, specific regions of the genome (in our case, 4q35). Plot showing the number of paired-end reads with one end mapped at chromosome 4 and its 4q35 locus representing the chromosome 4 sequences enrichment; the vast majority of the mapped reads are located on chromosome 4, among them the majority are in 4q35, demonstrating the efficiency of the capture.

Heat-map representation of the treated paired-end sequencing, mapped to the chromosome 4 (bottom left), and the last 5 Mbs of the chromosome (bottom left), using the HOMER software.

Supplemental S6. Organization of the 4q35.2 locus in control and FSHD myoblasts with long or short telomeres.

Interactions are represented with a Circos of the 4q-specific Hi-C experiment. The last 8 Mb of chromosome 4 is shown as a blue (or red) circle, and the location of known genes as segments in a green inner circle. Interactions are color coded (red lines >3MB, green 1-3 MB, blue 0.5-1MB, Orange 200-500kb, purple <200 kb). Black bars in the circles indicate the biotinylated bait sequences. This map was obtained using a filter against duplicate sequences and only takes into consideration interactions that aligned precisely to the 4q35 region without suspicion of other match (e.g to 10q).

Supplemental S7. Organization of the 4q35.2 locus in control myoblasts with long (upper panel) or short telomeres (lower panel).

As in figure S6, interactions are represented with a Circos of the 4q-specific Hi-C experiment. The last 8 Mb of chromosome 4 is shown in grey. Interactions are shown in a color coded fashion to represent the different degrees of significance (Blue $p<0.001$; Red $p<0.01$).

Maps were obtained using HOMER software and take into consideration bias introduced by the *HindIII* site localization, number of reads and 4q baits used. In short, maps are compared to a randomized generated 4q map using all baits sequences, and corrected regarding enrichment of restriction sites.

To note, the FSHD long telomere map is presented in figure 1, the FSHD short telomere map failed to pass our stringent filter when generating the final Circos map.

Supplemental S8. 3C Validation and Normalization.

a) *Fam149A* Interactions. A common primer adjacent to a *Hind*III site in *FRG1* was used with a variety of reverse primers adjacent to *Hind*III sites in *Fam149A* (7 primers) approximately 4 Mb away (left). Following crosslinking, digestion with *Hind*III, ligation and reversal of crosslinks, the DNA was amplified with the different primer pairs. Only FSHD sample with short telomere differs, with a decrease of 40.6% towards the center of interaction between *FRG1* and *Fam149A*.

b) Primers were picked surrounding successive *Hind*III restriction sites in ENr313, a region with no interaction (ENCODE region 313). One primer was constant. Distances in base pairs are represented relative to the first primer used. The interactions detected represent non-specific proximity effects in the absence of true looping. An average of the 11 points was made and used to normalize other 3C quantifications to allow sample comparison.

Supplemental S9. 3D DNA-FISH of interactions at 1p and 4q loci.

a) Quantification of the number of adjacent (A) and separated (S) signals from processed images of *In situ* hybridization using an ISG15 and 1p Subtelomeric probe. Images were processed with IMARIS for N~30 nuclei per condition. The mean of the distances measured between the gravity centers of each signal is reported in the left panel and a schematic representation of the chromosome on the right. Respective mean ± SEM of the adjacent distances are shown. No differences were seen between adjacent signals nor between separated signals; independent test between all conditions for binary choice, p <10E-7; t-test of the mean distance between centers, p=0.01. Myoblasts with short telomeres show separated signals, confirming the existence of a TPE-OLD regulation of *ISG15*, importantly no differences were seen between FSHD and control cells. b) Quantification of of the number of adjacent (A) and separated (S) signals from processed images of *In situ* hybridization using a ZFP42 and FRG1probe.

Images were processed with IMARIS for N~30 nuclei per condition. The mean of the distances measured between the gravity centers of each signal is reported in the left panel and a schematic representation of the chromosome on the right. Respective mean ± SEM of the adjacent distances are shown. No differences were seen between adjacent signals nor between separated signals; independent test between all conditions for binary choice, p <10E-7; t-test of the mean distance between centers, p=0.01. All cells showed adjacent signals, regardless of the telomeres and the D4Z4 length.

Supplemental Figure S10. Gene expression analysis of *SORBS2*.

a) Representation of the most prevalent *SORBS2* variants as predicted by Ecgene. b) RNA was isolated from 1x10^6 cells and cDNAs made in triplicate. Assays were designed with either a primer or probe spanning a unique exon-exon junction. Due to the complexity of the *SORBS2* gene, the primer pairs detect short sequences and quantify the abundance of exon junctions without being able to identify specific transcripts exclusively. For all four assays, we report the transcripts predicted from the NCBI database. For ddPCR, results are expressed as the number of molecules detected in a 12.5ng RT input; values shown as averages ± standard error of the mean (assay done in duplicate using 3 cohorts of matched siblings from 2 biopsies; a total of 12 measures per condition). Increase in all assays is observed upon differentiation (myoblast compared to myotubes). For all assays, no significant differences were detected when considering telomere length (ANOVA). Similarly, no significant differences were detected considering the disease status, however a trend is noticeable between control and FSHD myotubes, with an increased in *SORBS2* transcripts for the exons 39-40 assay. In the 5'UTR assay, in myoblast, we detect a significant increase in expression in FSHD myoblasts with short telomeres (p=0.0001). In myotubes, an overall significant increase (>3-fold increase, p=0.0001) of expression is observed when compared to myoblast. The increased expression is even more important in FSHD cells, but not correlated with telomere length. (control myotubes-long telomere vs. FSHD myotubes-long telomeres, adjusted p value=0.0001; control myotubes-short telomere vs. FSHD myotubes-short telomeres, adjusted p value=0.039).

Supplemental S11. RT-PCR in Control myoblast.

As observed in Fig. 2 and Supplemental S2; control cells do not exhibit the splicing variant of *SORBS2* upon telomere shortening. RT-PCR of 2µl cDNA from a 1µg RNA RT reaction were separated in 5 aliquots and amplified by PCR for 20; 25; 30; 35 or 40 cycles. *SORBS2* transcripts were analyzed from to additional control samples from two different cohorts (11 and 15). Transcripts containing exon 31 to 42 are present in controls regardless of telomere length. Transcripts containing exons 17 to 42 are not detectable in the control cells.

Supplemental S12. Telomere shortening in FSHD cells induces alternative splicing modifications of *SORBS2* in Deltoid muscle biopsies.

The biceps is more affected than the deltoid in most FSHD patients. Changes in *SORBS2* splicing (in biceps, text Figure 2) were also seen in cells derived from deltoid muscles. RT-PCR of 2µl cDNA from a 1µg RNA RT reaction were separated in 5 aliquots and amplified by PCR for 20; 25; 30; 35 or 40 cycles. *SORBS2* transcripts were analyzed from 8 different samples from two different cohorts. Transcripts containing exon 31 to 42 are present in controls and FSHD samples regardless of telomere length. Transcripts containing exons 17 to 42 are only present in FSHD cells with short telomeres (with the exception of small amounts of the shorter transcript in one control sample). The longer *SORBS2* transcript (star) was only observed in FSHD. Telomere length induces changes in the splicing of *SORBS2* transcripts.

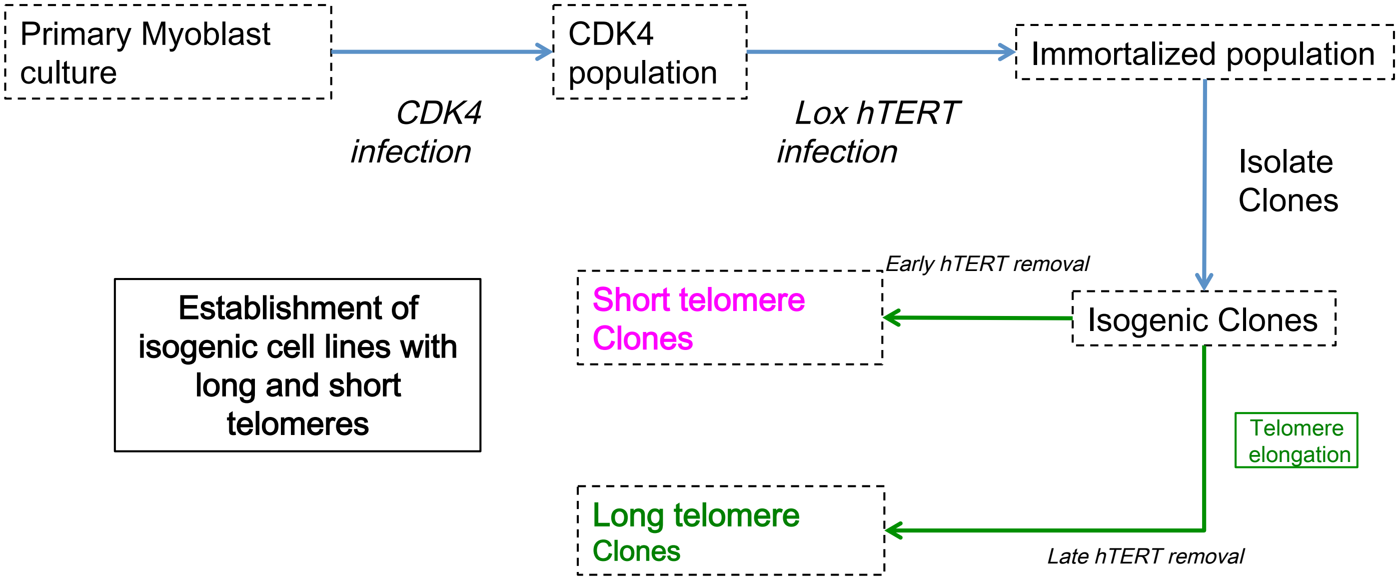
Supplemental Figure S13. Gene expression analysis of 4q35 genes.

RNA was isolated from 1million cells and cDNAs made in triplicate. ddPCR results are expressed as the number of molecules detected in a 12.5ng RT input; values shown as averages ± standard error of the mean (assay done in duplicate using 3 cohorts of matched siblings from 2 biopsies; total of 12 measures per condition). Gene expression results for Fam149A, ZFP42, FAT1, FRG1, TRIML1 and PDLIM3. With the exception of PDLIM3, no differences were detected in any conditions. PDLIM3 expression increases upon differentiation (Holm-Sidak's multiple comparisons test (alpha=0.05)).

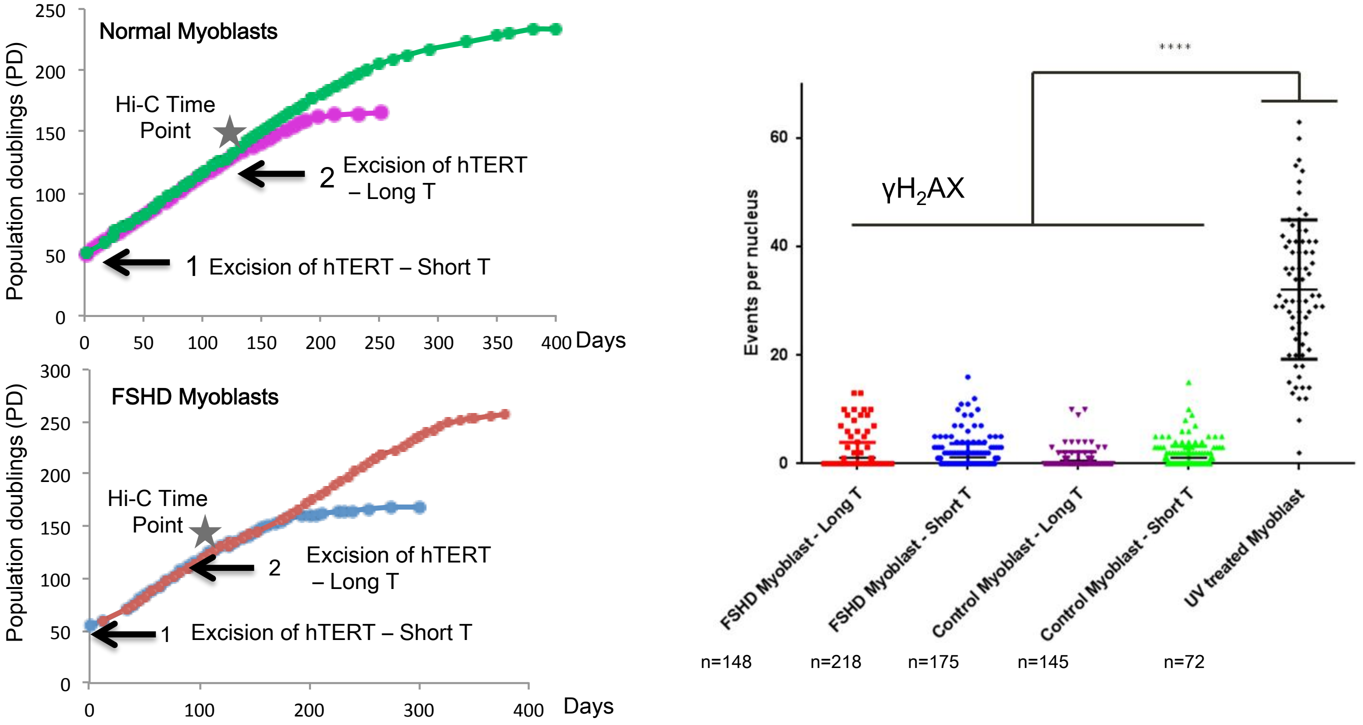
Supplemental Figures

Supplemental Figure S1.

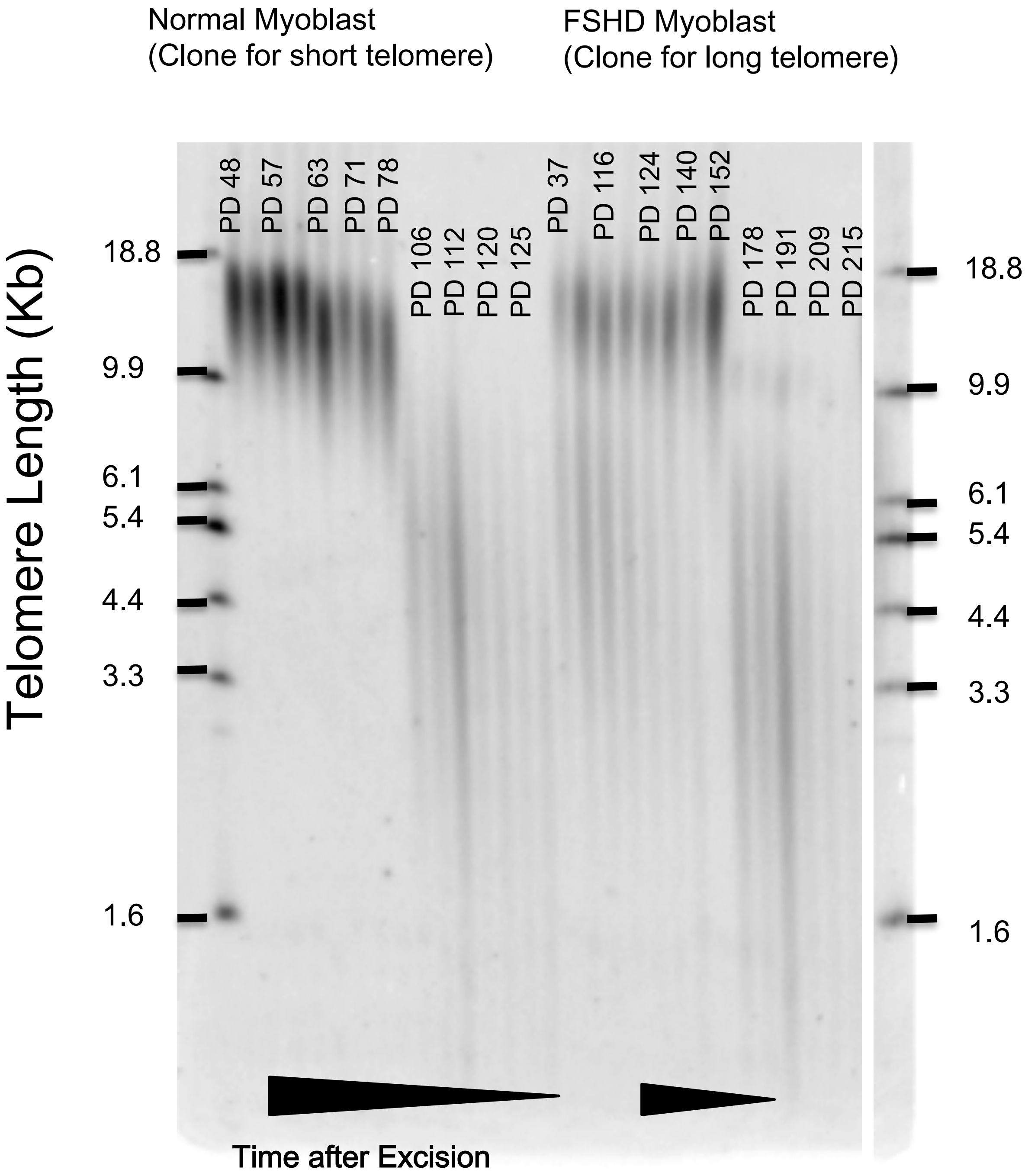
a)



b)

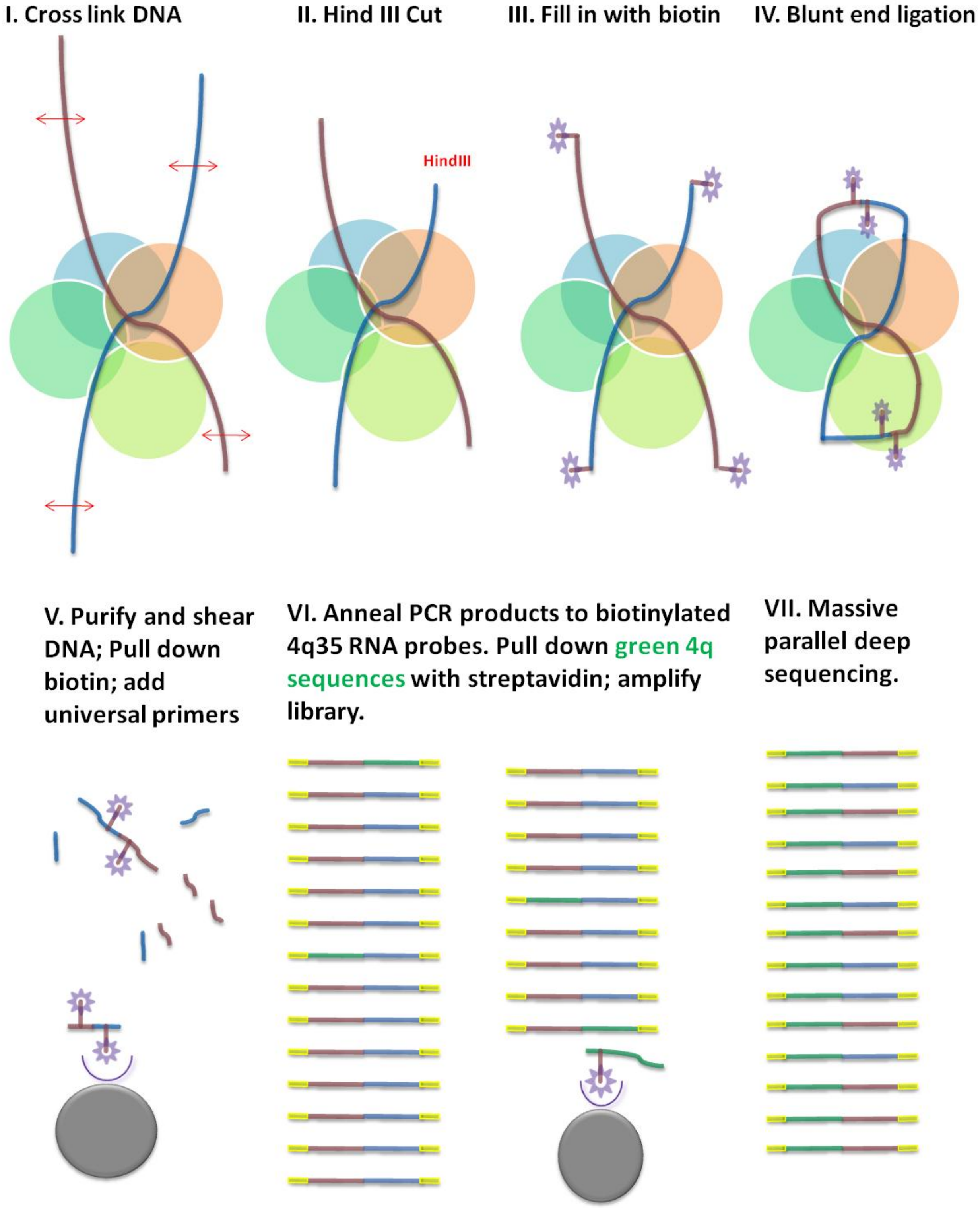


Supplemental Figure S2.



Supplemental Figure S3.

Locus specific Hi-C Strategy



I. Cross link DNA

II. Hind III Cut

III. Fill in with biotin

IV. Blunt end ligation

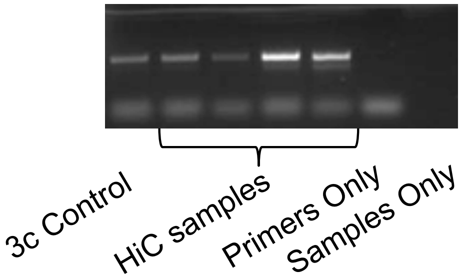
V. Purify and shear DNA; Pull down biotin; add universal primers

VI. Anneal PCR products to biotinylated 4q35 RNA probes. Pull down green 4q sequences with streptavidin; amplify library.

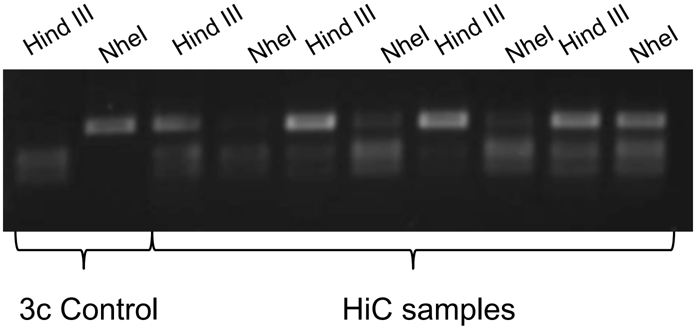
VII. Massive parallel deep sequencing.

Supplemental Figure S4.

Hi-C ligation control



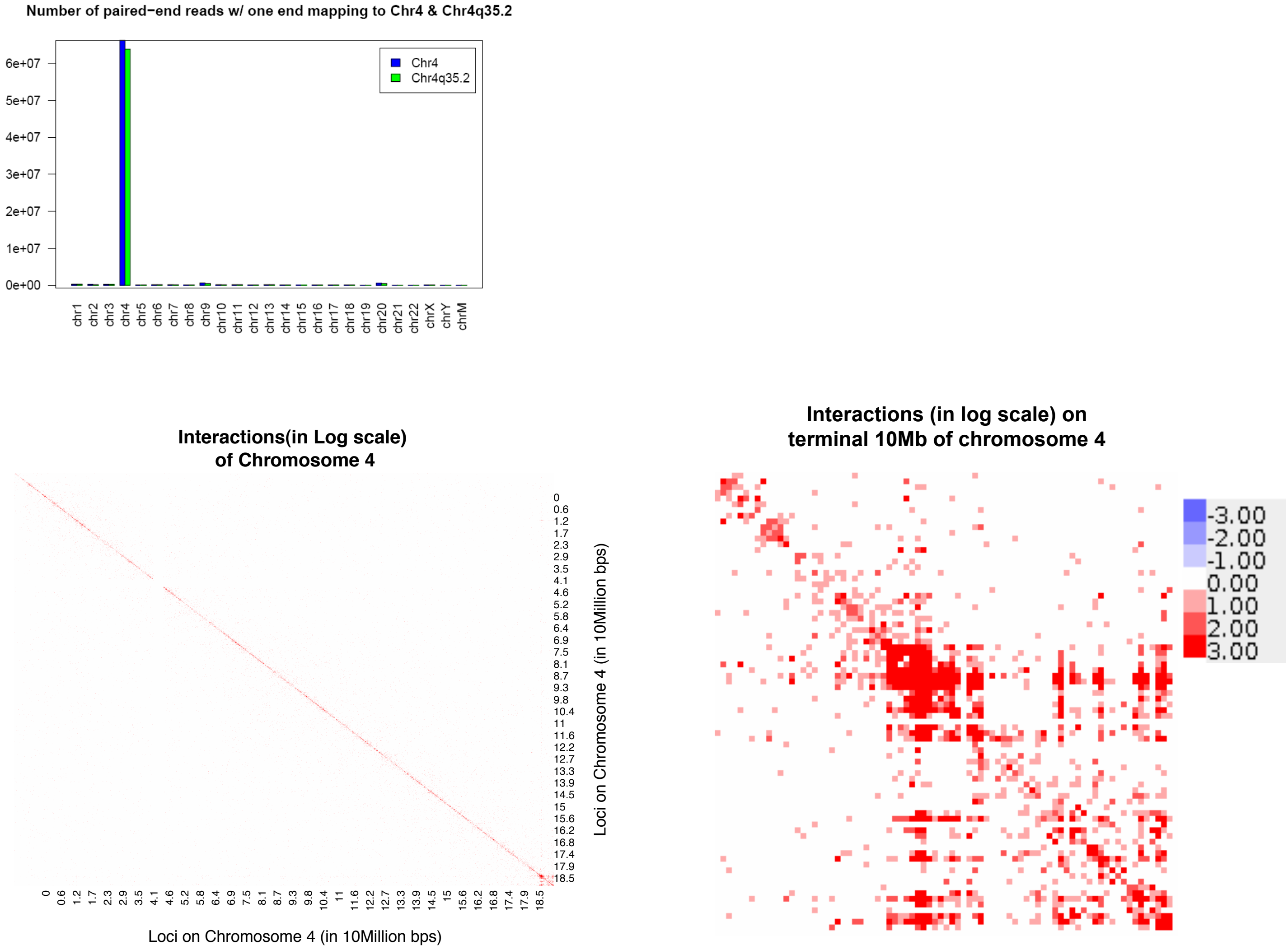
Pcr control



Pcr Digestion

- NheI/HindIII

Supplemental Figure S5.

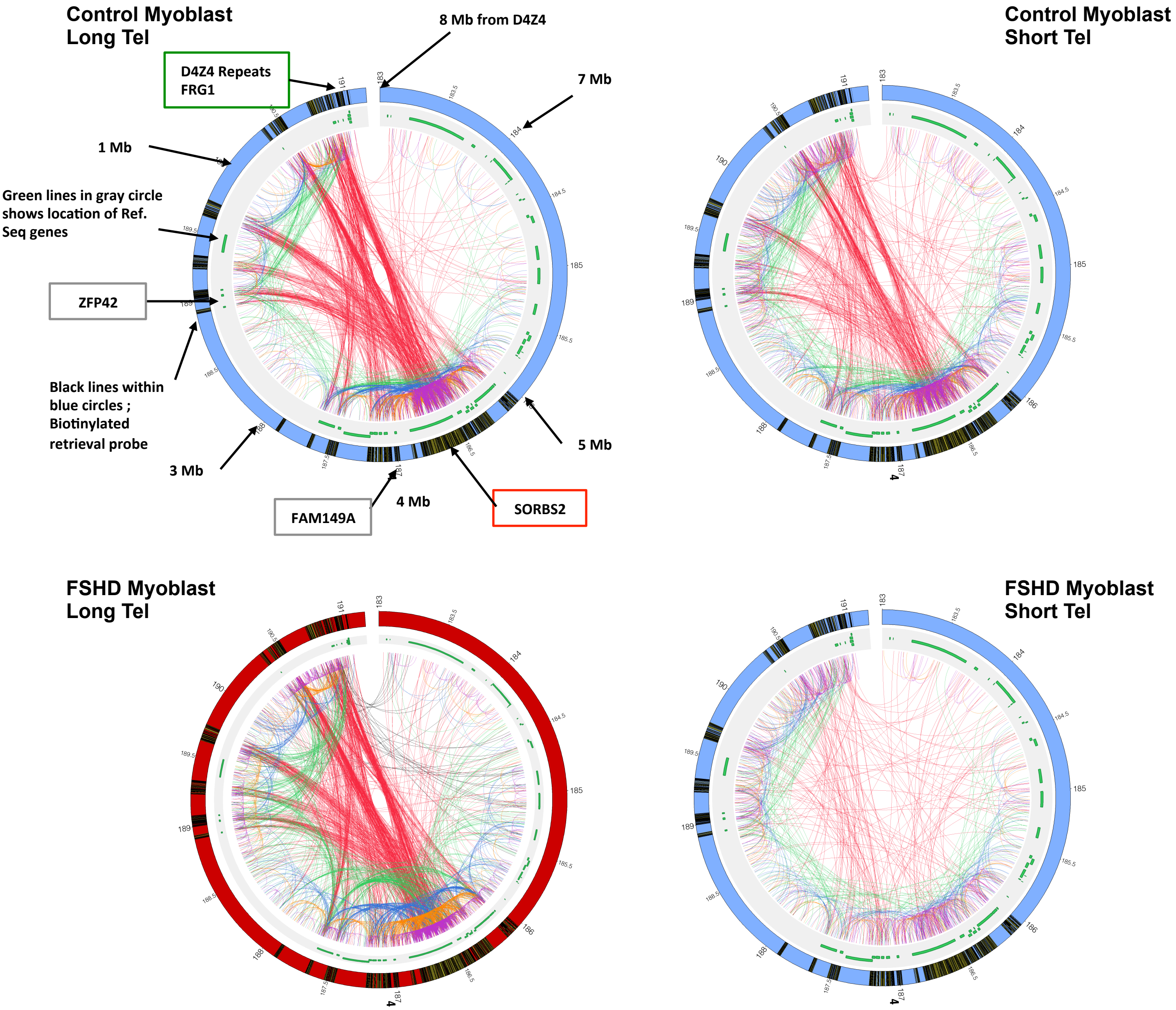


Number of paired−end reads w/ one end mapping to Chr4 & Chr4q35.2

Interactions(in Log scale) of Chromosome 4

Interactions (in log scale) on terminal 10Mb of chromosome 4

Supplemental Figure S6.



Control Myoblast Long Tel

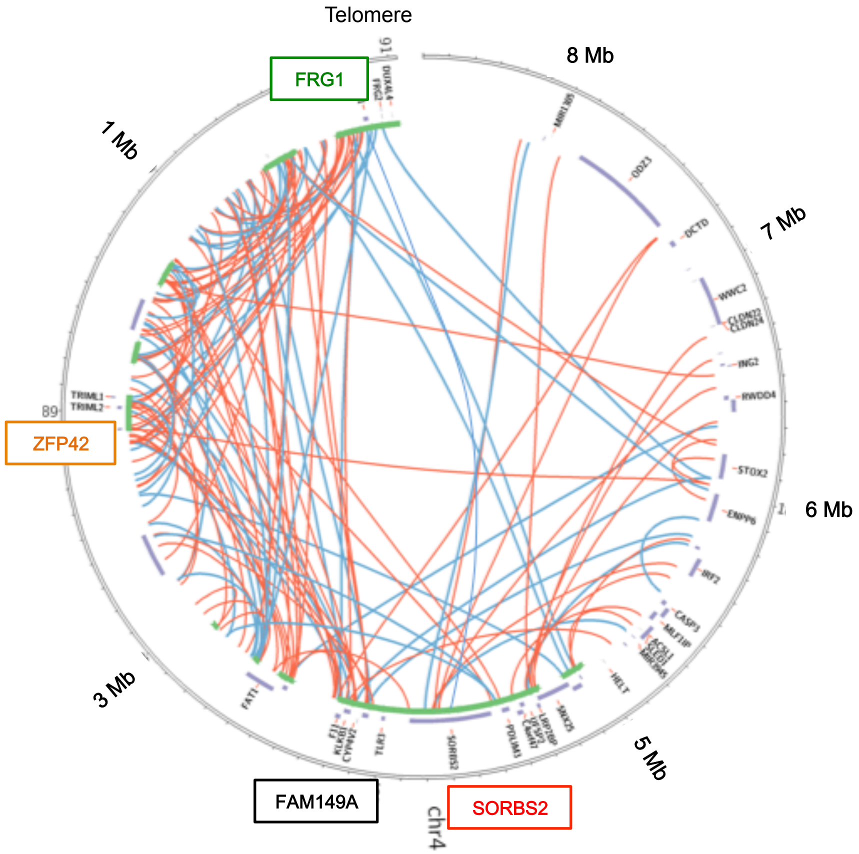
Control Myoblast Short Tel

FSHD Myoblast Long Tel

FSHD Myoblast Short Tel

Supplemental Figure S7.

Control Myoblast Long Telomere

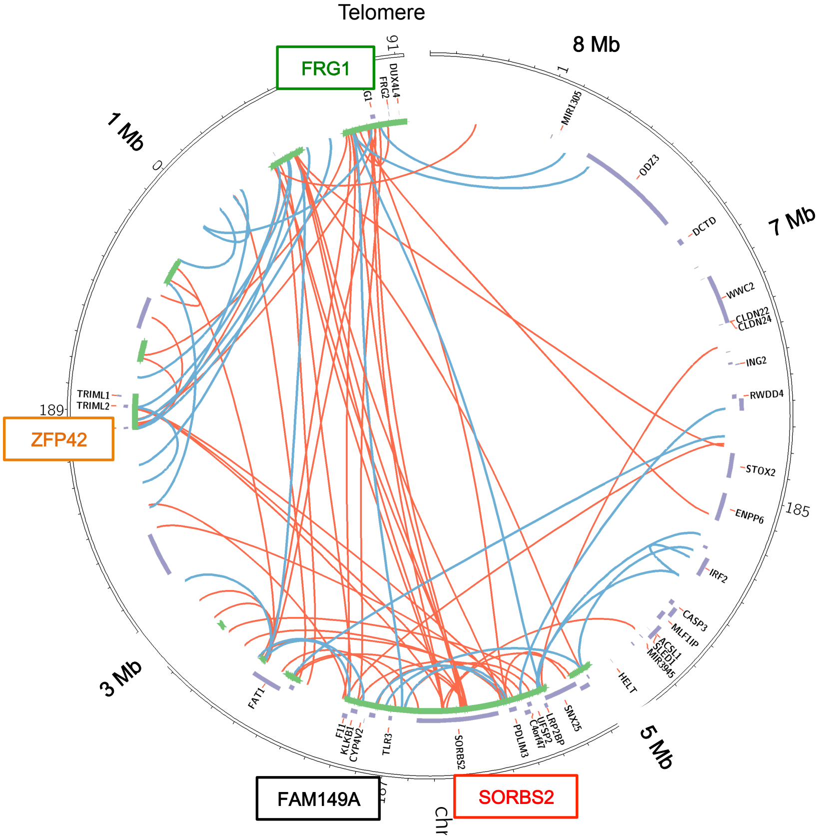


Legend

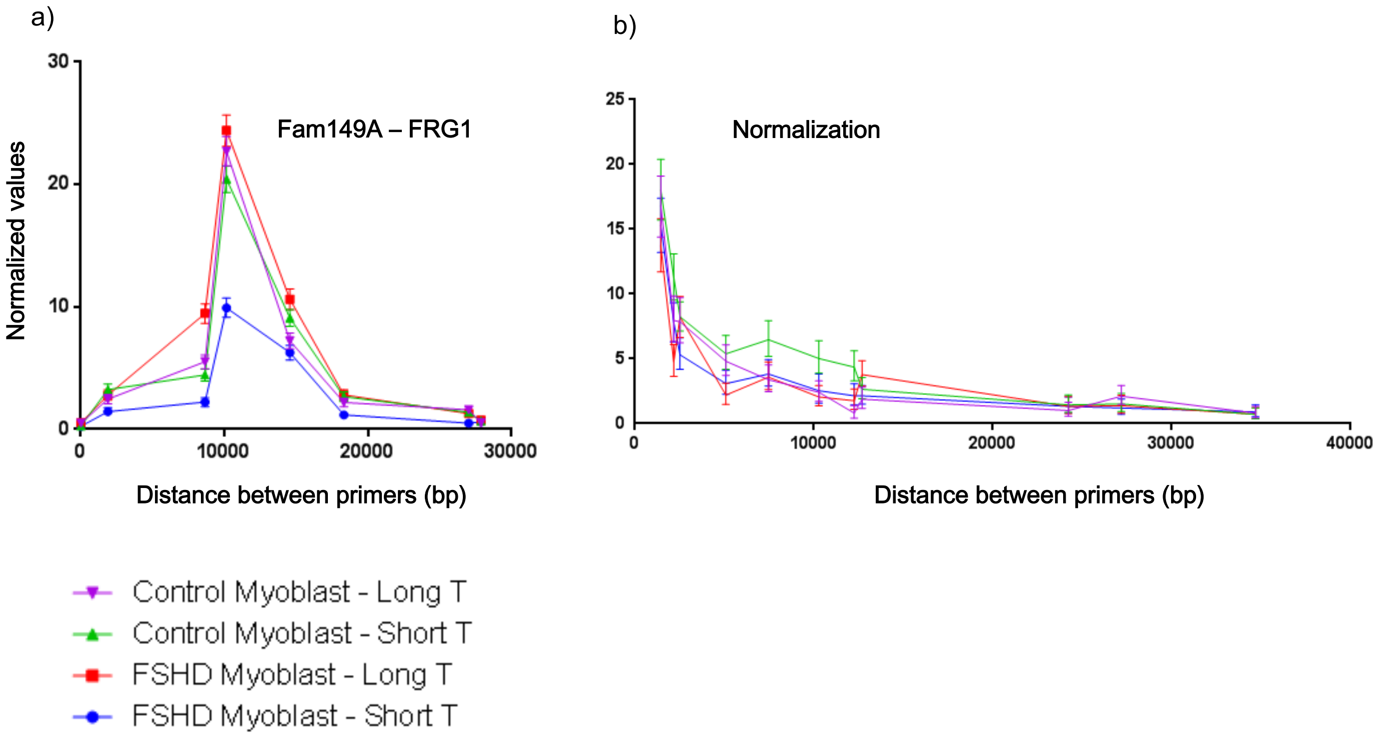
X Mb Distance from the Telomere
Genes
RNA Baits location

Statistical significance
P< 0.05
P<0.001

Control Myoblast Short Telomere

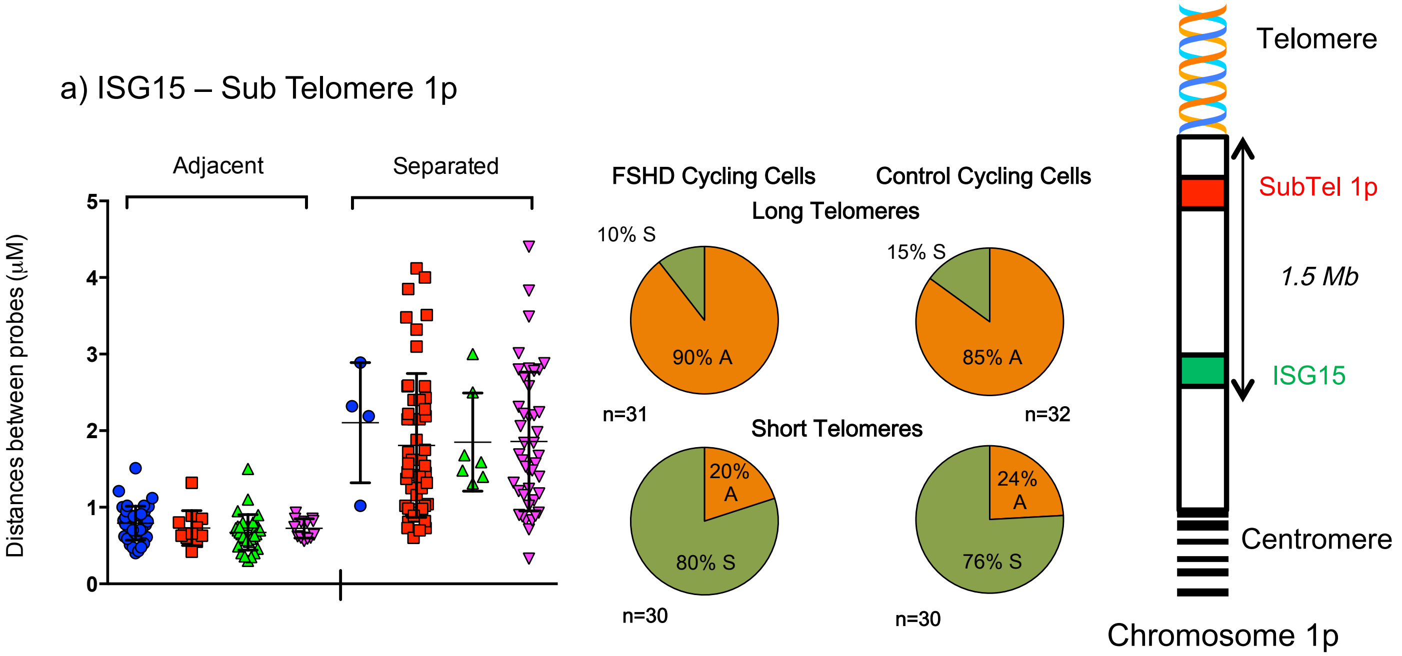


Supplemental Figure S8.

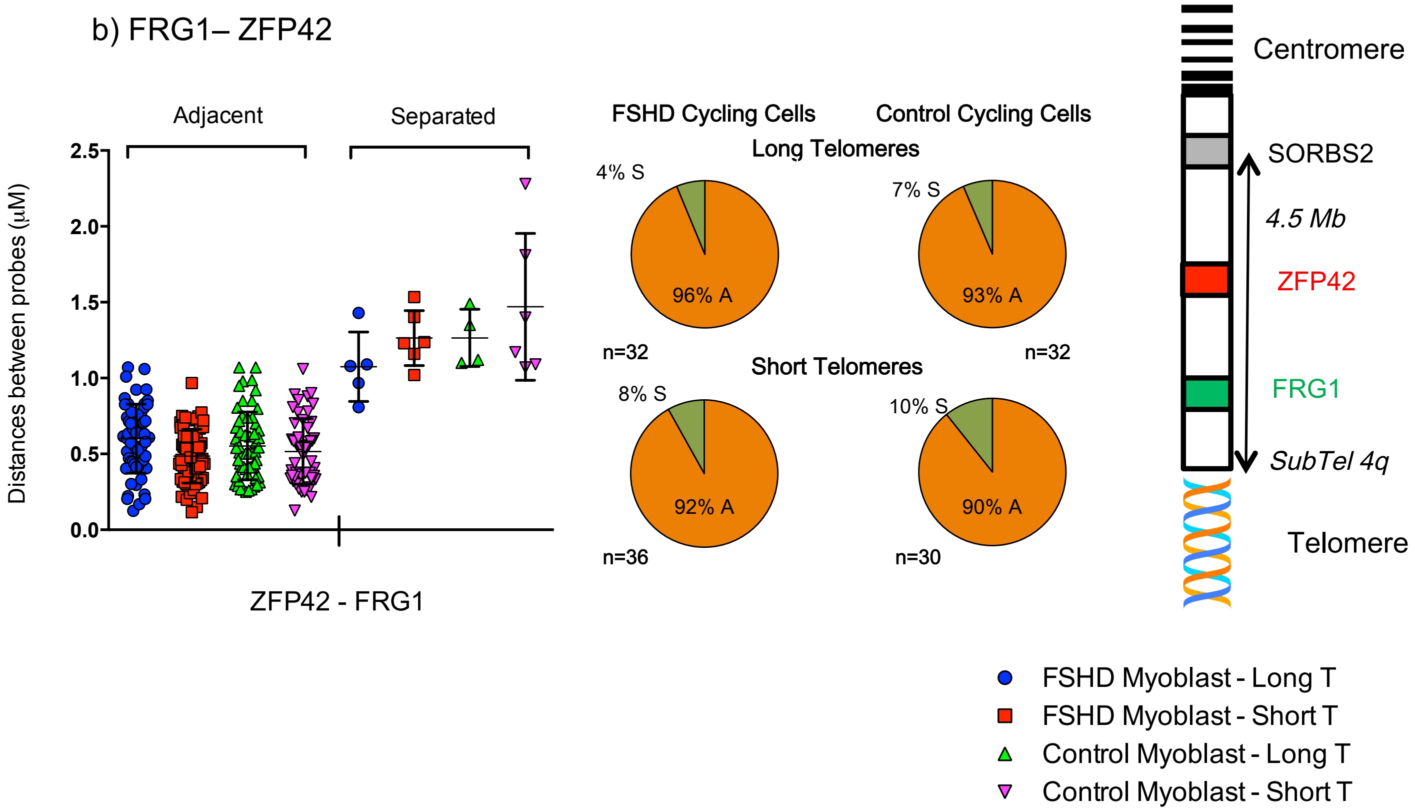


Supplemental Figure S9.

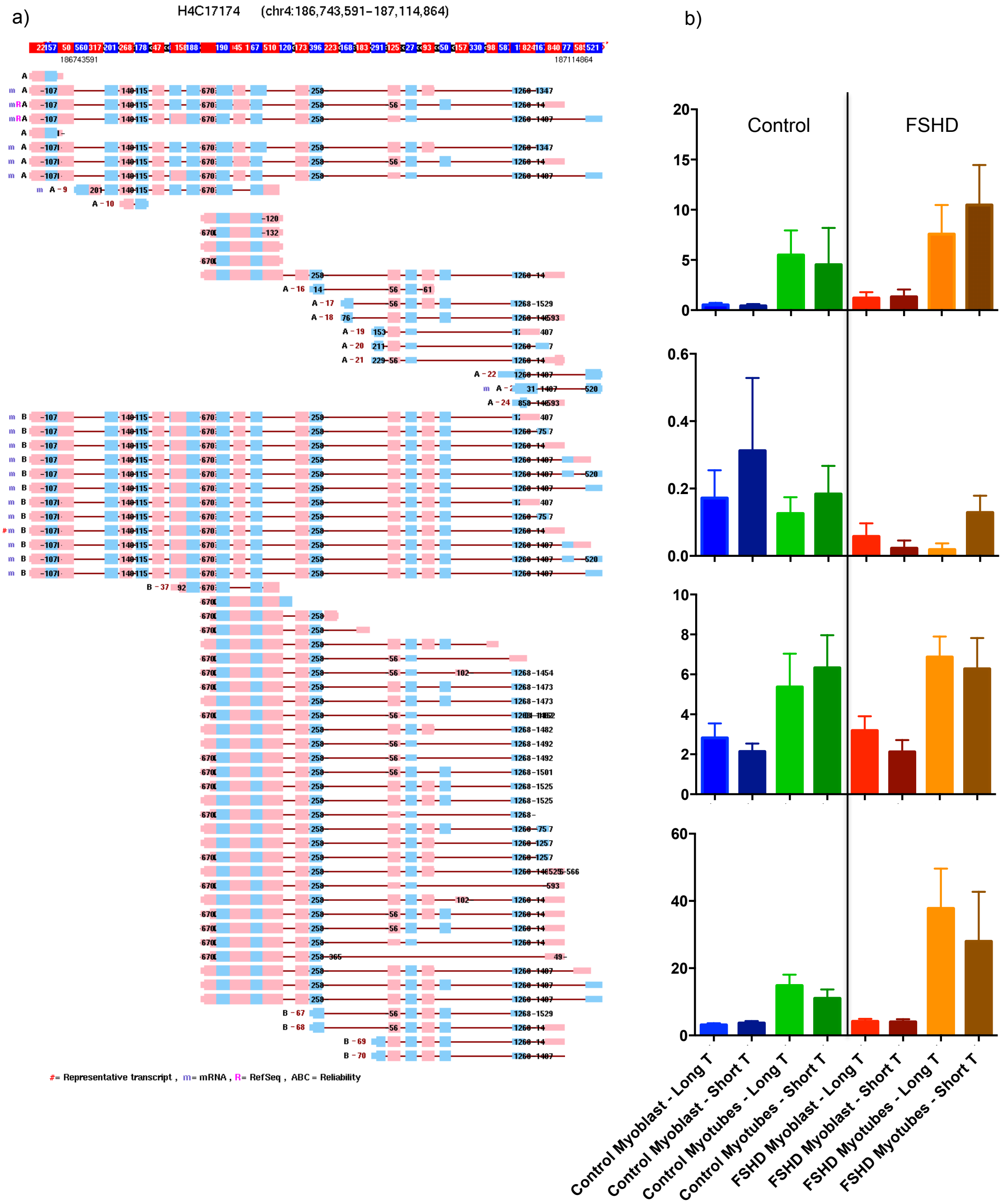
a) ISG15 – Sub Telomere 1p



b) FRG1– ZFP42



Supplemental Figure S10.



__Variants detected by the assay__

Exon 8-Exon 9 junction

No Transcript variant reported

Exon 19- Exon 20 Junction

Transcript variants X9; X12-13; X21-29; X30-33; X36-39; X65

Total detected: **19**

Exon 21- Exon 22 Junction

No Transcript variant reported

Exon 39- Exon 40 Junction

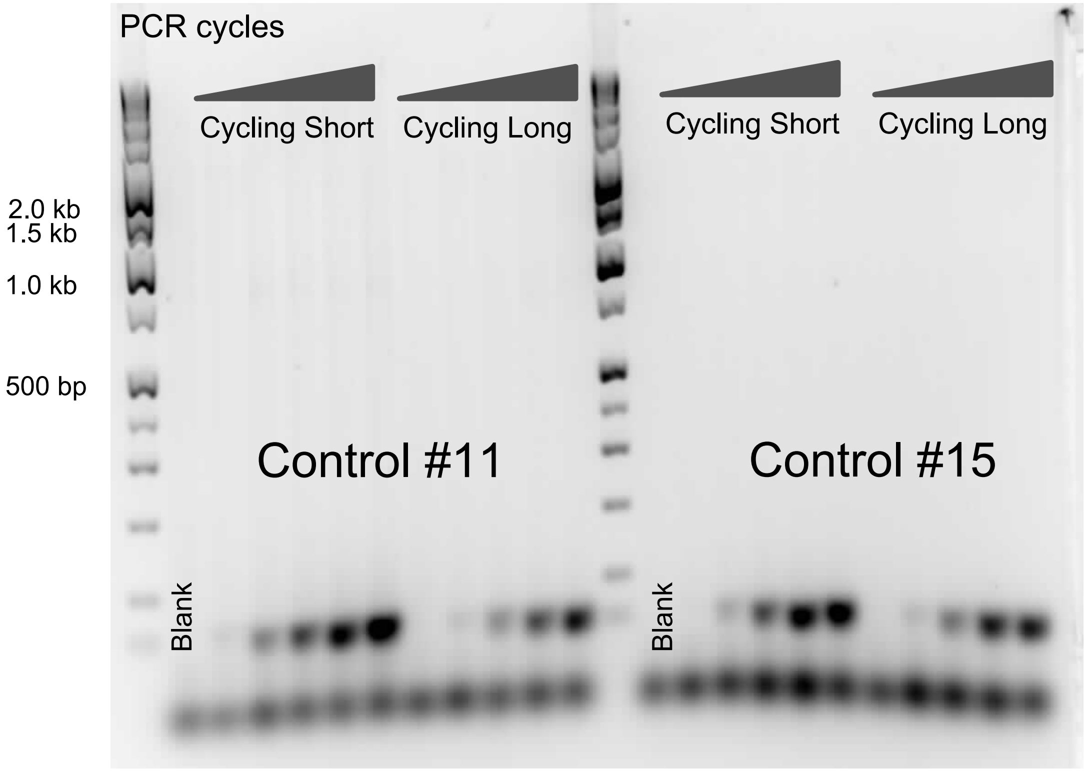
Transcript variants 1-9

Transcript variants X9; X12-15; X17-64; X66

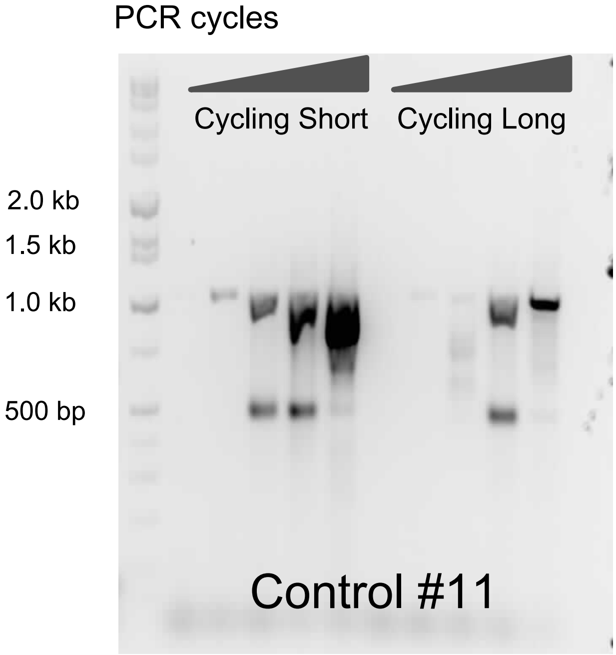
Total detected: **63**

Supplemental Figure S11.

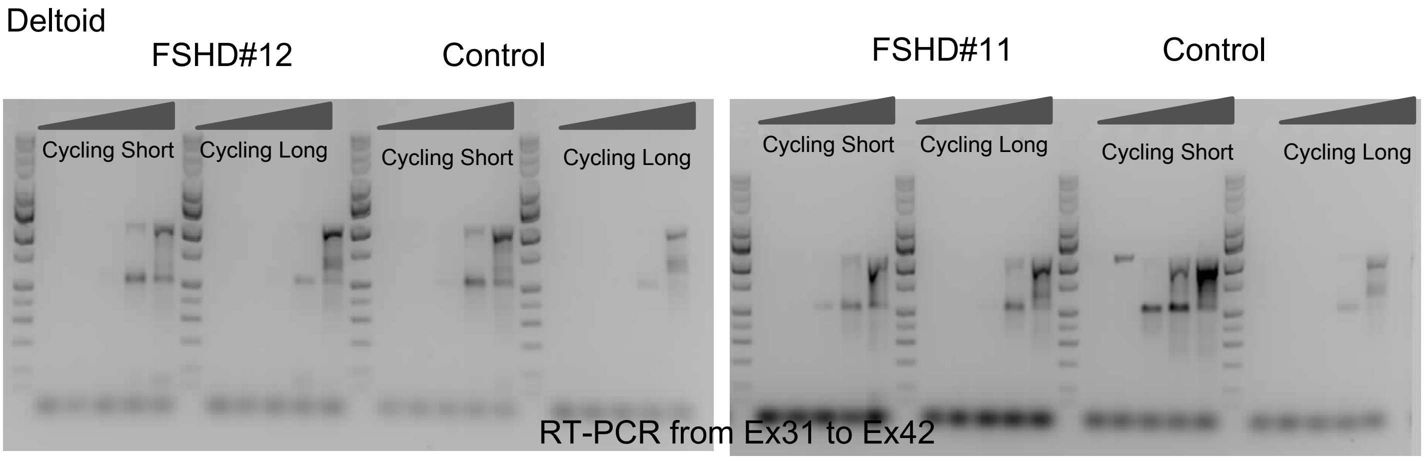
RT-PCR from Ex17 to 42



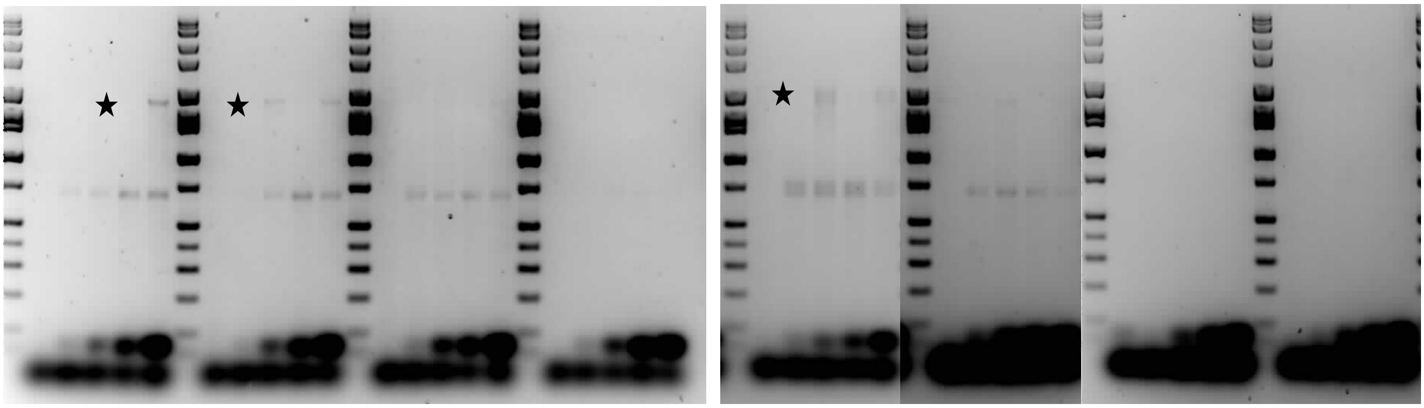
RT-PCR from Ex31 to 42



Supplemental Figure S12.



RT-PCR from Ex31 to Ex42



RT-PCR from Ex17 to Ex42

Supplemental Figure S13.