Supplementary Figures S1 to S5, Figure Legends, and Tables S1 and S2

Figure S1. Subtelomeric 4q has the lowest density of gene sequences of all the subtelomeric regions of q-arms. The cumulative lengths of annotated (RefSeq) genes in the terminal 3 Mb of the q arms of all the human chromosomes are shown (hg18, http://genome.ucsc.edu/). The base-pair coverage of the 3-Mb subtelomeric region of 4q for RefSeq genes (shown above) is only 2% and for UCSC Genes (including RefSeq genes and predicted genes based upon Genbank, CCDS and UniProt, http://genome.ucsc.edu/; data not shown) is 9%. In comparison, the genome-wide base-pair coverage for Refseq genes in hg18 is 36% (1.14 Mb). For calculation of genes in this region, six copies of *DUX4* were included, as in hg18. The number of copies of the D4Z4 repeat with its *DUX4* subregion ranges between 1 and 100 but expression of *DUX4* is very low, not increased by a higher copy number, and thought to be mostly from the terminal copy, the only one with a distal polyA signal (1).

Figure S2. Distribution of DH sites in 4q35.2 relative to G+C content, GC isochores, interspersed DNA repeats, and sequence conservation among vertebrates

The average data from three FSHD and three normal control myoblast cultures are shown. Green dots, unique DH sites seen in all six myoblast cultures; orange star, DH272, seen preferentially in the three FSHD myoblast cell strains; gray dots, DH peaks centered on simple tandem repeats (STRs). The % (G+C) for 5-bp windows, 17-way vertebrate sequence conservation, chicken/human sequence conservation, segmental duplications, short interspersed sequences (SINEs), long interspersed sequences (LINEs), long terminal repeats (LTRs), and RefSeq gene tracks are from the UCSC genome browser (http://genome.ucsc.edu/, hg18) and isochores from Isofinder (2).

Figure S3. DH site mapping in six myoblast cell cultures: near the proximal end of 4q35.2

The maps of DH sites in the vicinity of *FAT1* and *MTNR1A* are shown for each of the myoblast cell strains. Green dots, unique-sequence peaks seen in all myoblast cultures; gray dot, a DH site that overlapped a simple tandem repeat and was seen in all myoblast cultures. The coordinates of the repeat mask region were downloaded from the UCSC genome browser (hg18, rmsk3.27) and are displayed at the top as a dense tool to illustrate the regions where probes were missing from the array.

Figure S4. DH site mapping in six myoblast cell cultures: the proximal end of 4q35.2

The DH sites around *FAM149A* and *CYP4V2* are mapped and designated as in Figure S3. No DH sites were found in the vicinity of blood coagulation-related genes *KLKB1* and *F11*.

Figure S5. Tested amplicons around DH272

Amplicons around DH272 were chosen for qRT-PCR using the following gene prediction programs: EvoFold (3); FirstEF, first exon finder (4); RNAz, noncoding RNAs (5); and geneid (http://genome.imim.es/software/geneid/index.html). At the bottom is shown sequence conservation among 17 vertebrate species (hg18, http://genome.ucsc.edu/). The locations of the amplicons with the highest levels of transcripts (indicated by filled diamonds) suggest that these transcripts are not derived from protein-encoding RNAs, and so are ncRNAs. Note that the indicated gene predicted by geneid would have a coding sequence spanning 63 kb and that the end of the predicted gene that is shown is the 3' end. The distant, predicted 5' end has no DH site in its vicinity.

References

1. 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) *Hum Mol Genet*, **18,** 2414-2430.
2. Oliver, J.L., Carpena, P., Hackenberg, M. and Bernaola-Galvan, P. (2004) *Nucleic Acids Res*, **32,** W287-292.
3. Pedersen, J.S., Bejerano, G., Siepel, A., Rosenbloom, K., Lindblad-Toh, K., Lander, E.S., Kent, J., Miller, W. and Haussler, D. (2006) *PLoS Comput Biol*, **2,** e33.
4. Davuluri, R.V., Grosse, I. and Zhang, M.Q. (2001) *Nat Genet*, **29,** 412-417.
5. Washietl, S., Hofacker, I.L. and Stadler, P.F. (2005) *Proc Natl Acad Sci U S A*, **102,** 2454-2459.

Figure S1. RefSeq genes in the terminal 3 Mb of each q arm: base coverage

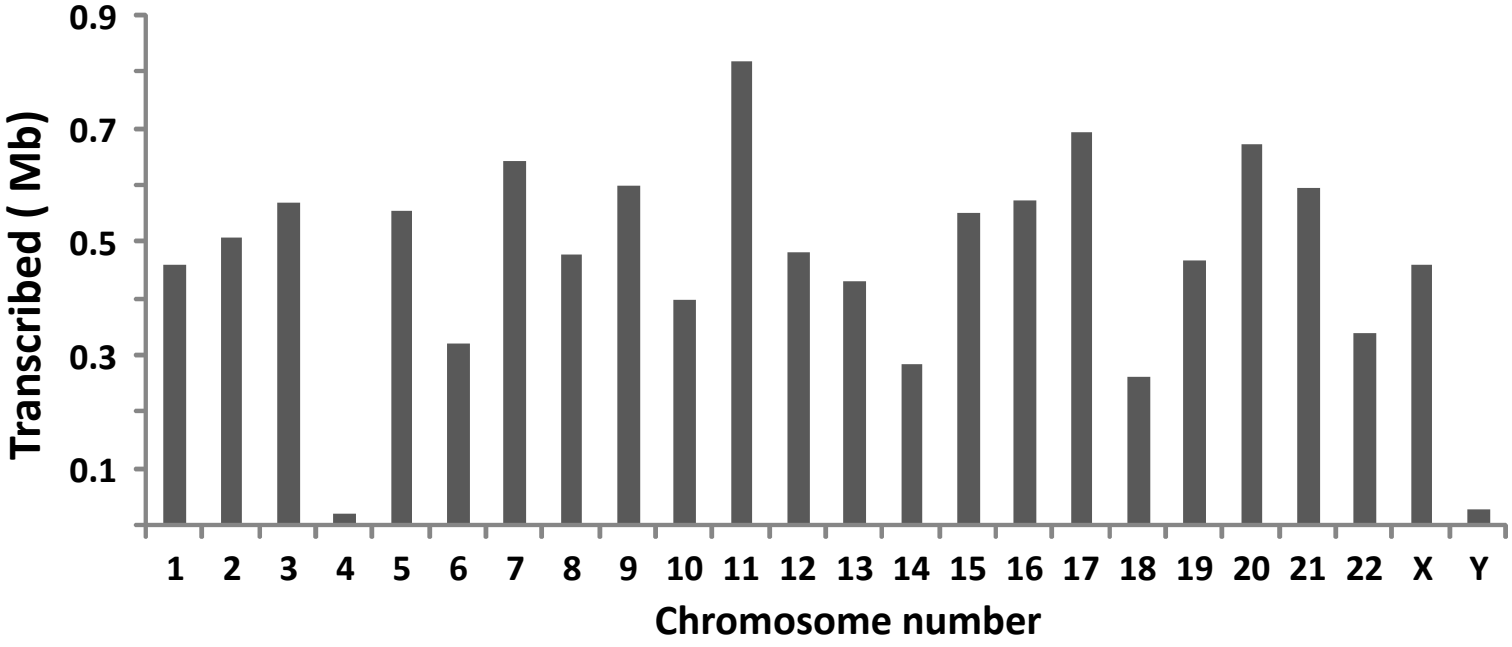

Figure S2. Distribution of DH sites in 4q35.2 relative to G+C content, interspersed DNA repeats, and sequence conservation among vertebrates

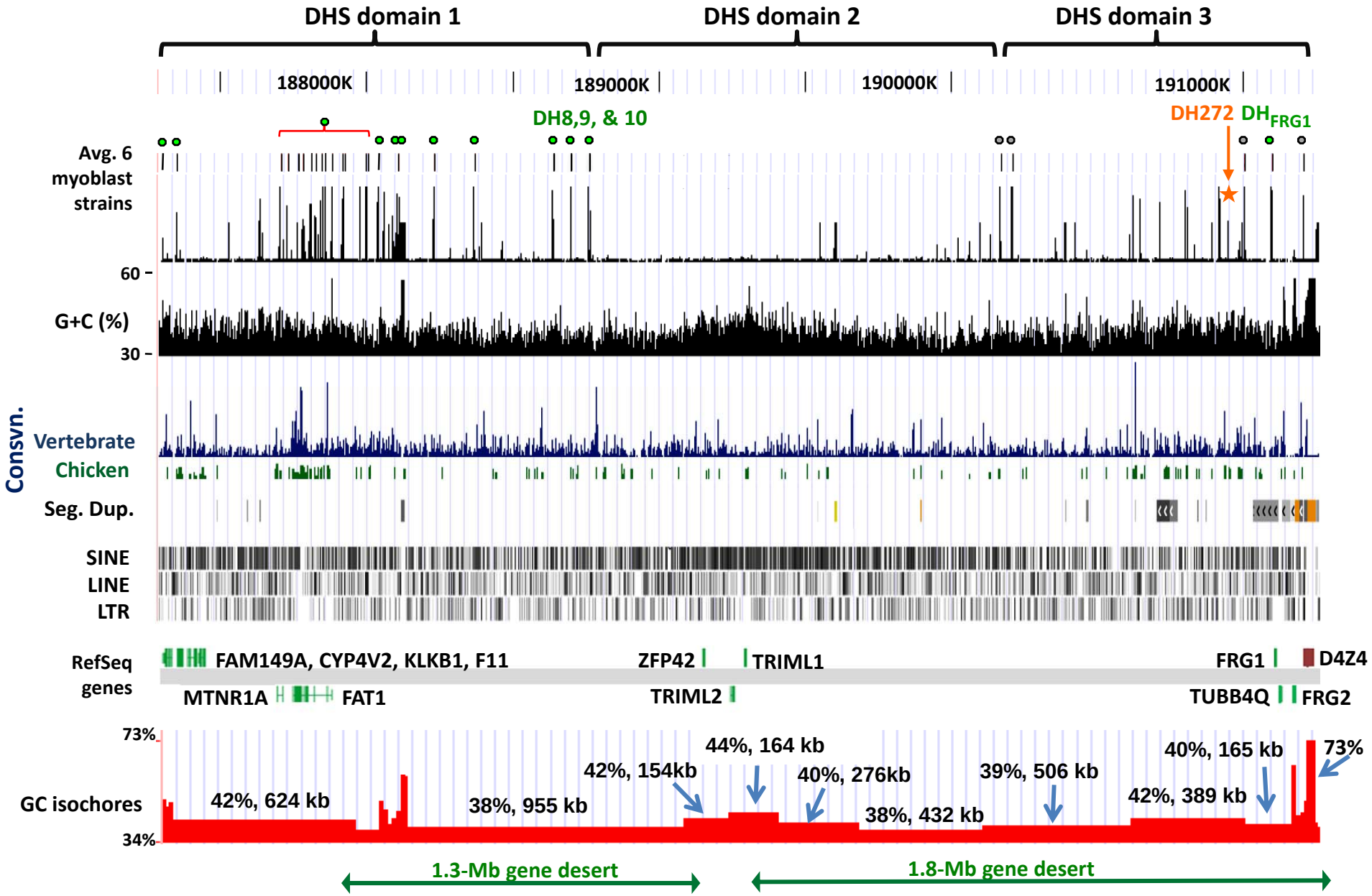

Figure S3. DH site mapping in six myoblast cell cultures: near the proximal end of 4q35.2

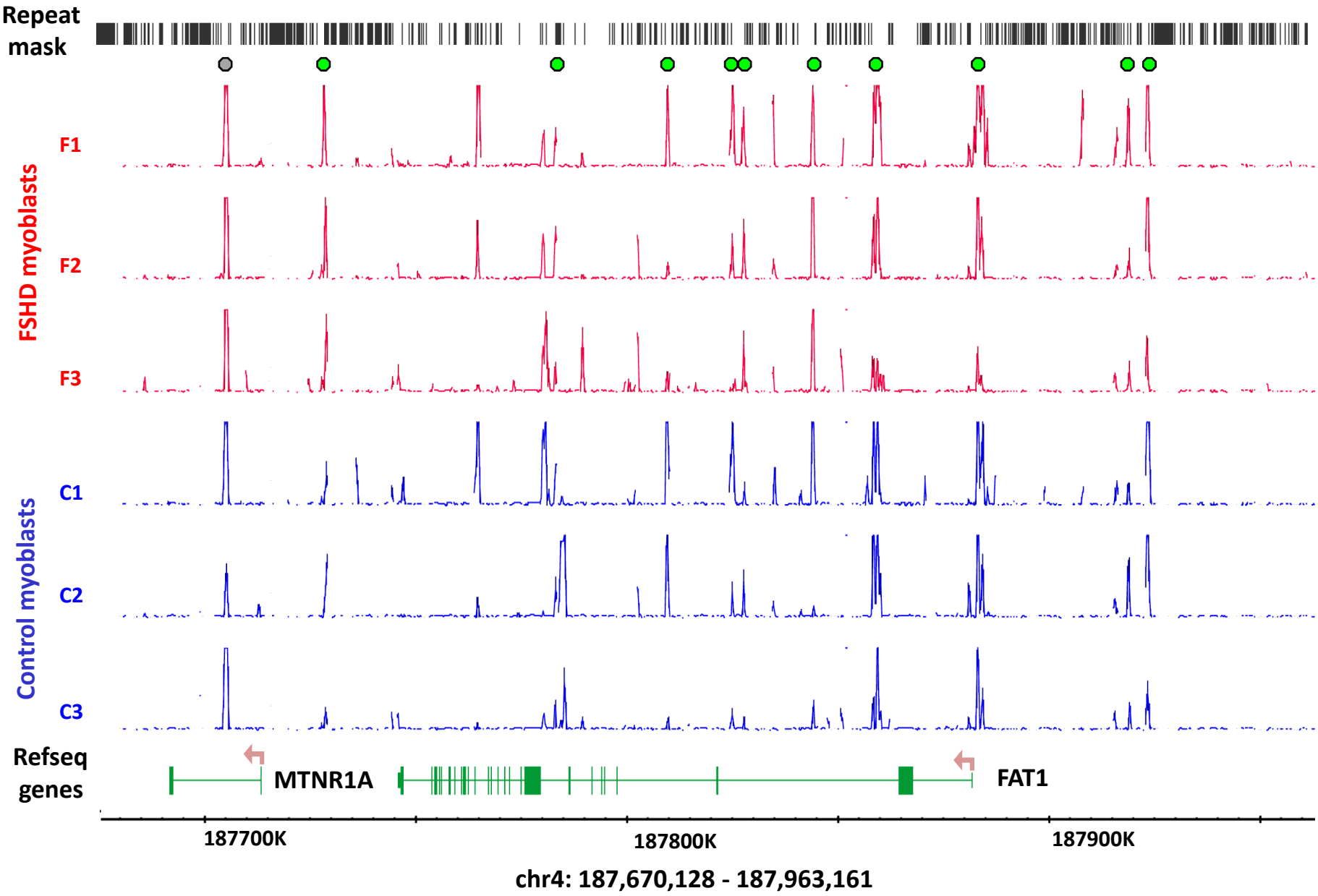

Figure S4. DH site mapping in six myoblast cell cultures: the proximal end of 4q35.2

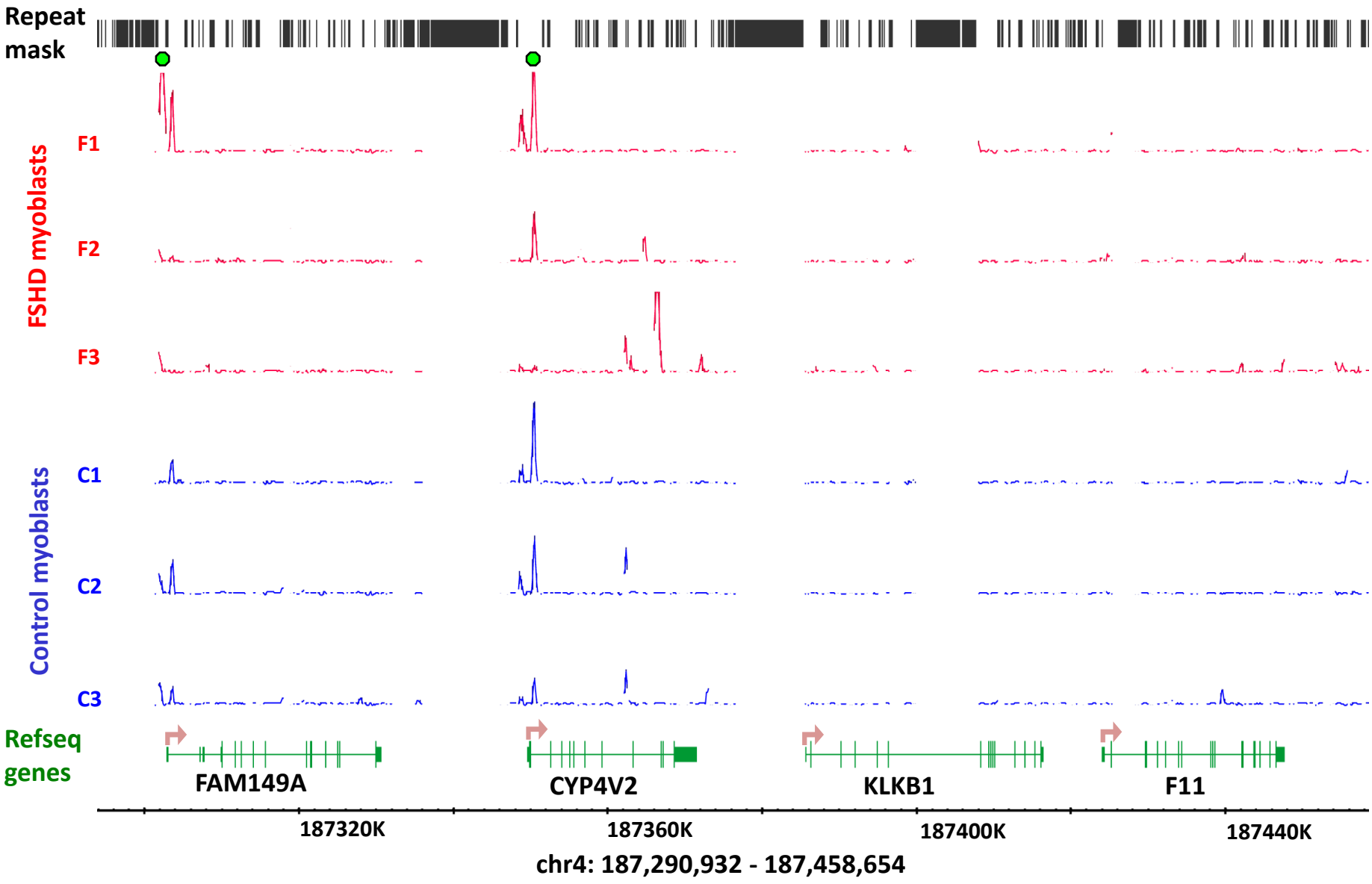

Figure S5. Tested amplicons around DH272:

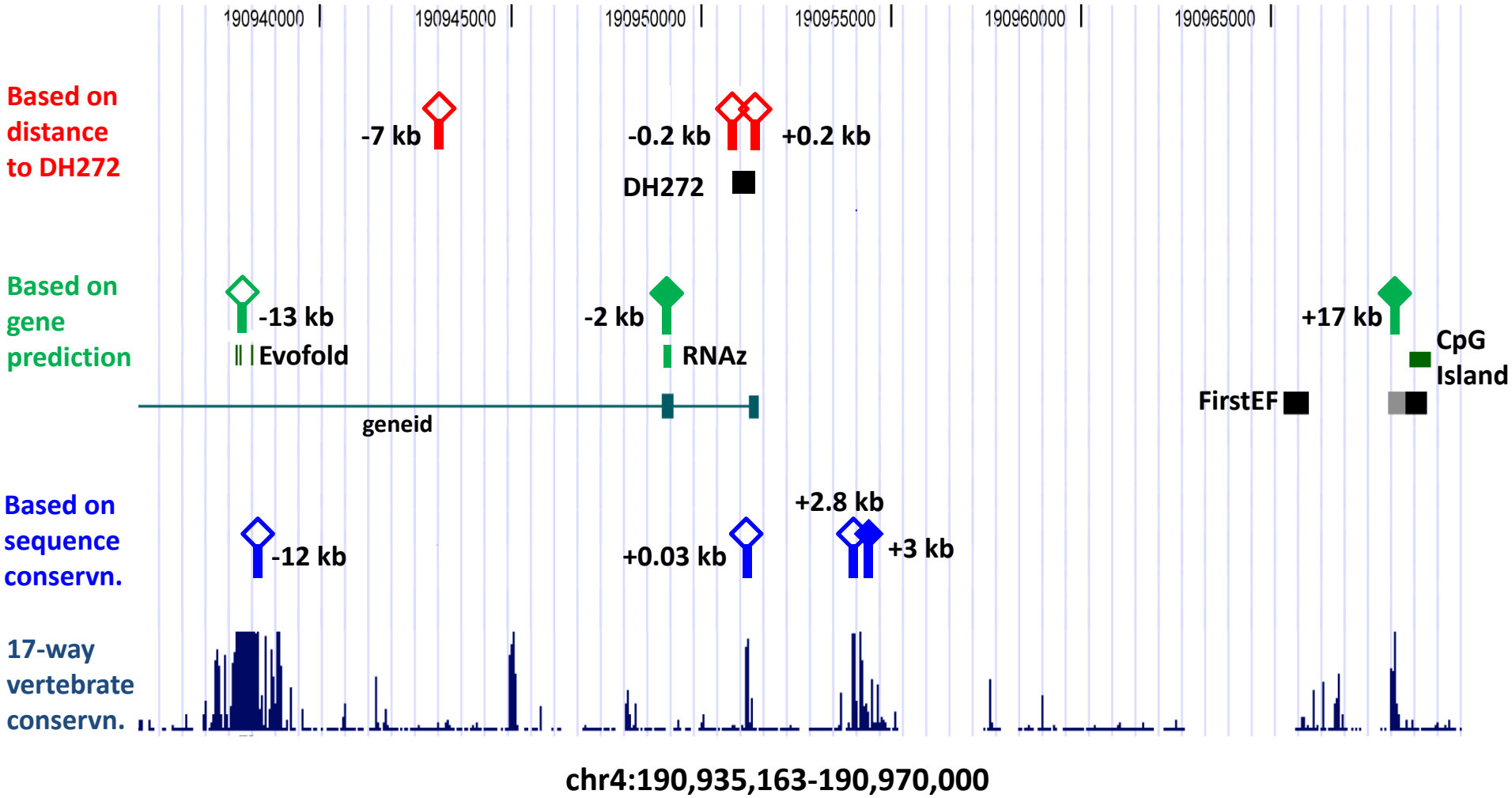

Table S1. Positions, length, and GC content of DH sites in 4q35.2 in myoblast cell strains[1]

A. Unique-sequence DH sites located >20 kb from a RefSeq gene[2]

	DH site name	DH site coordinates		G+C (%)	Length (bp)	Distance from the closest gene	Distance to D4Z4 (Mb)
chr4	DH2.2	187918708	187919336	44	628	37 kb from FAT1	3.30
chr4	DH2.1	187922992	187924054	43	1062	41 Kb from FAT1	3.30
chr4	DH3	187997838	187998472	48	634	116 kb from FAT1	3.23
chr4	DH4	188002298	188003940	43	1642	120 kb from FAT1	3.22
chr4	DH4.2	188062058	188062553	35	495	180 kb from FAT1	3.16
chr4	DH5	188106569	188107300	46	731	225 kb from FAT1	3.12
chr4	DH5.2	188111813	188112563	47	750	230 kb from FAT1	3.11
chr4	DH6	188230158	188230684	44	526	348 kb from FAT1	2.99
chr4	DH7	188367665	188368548	41	883	486 kb from FAT1	2.86
chr4	DH8	188636930	188637268	45	338	517 kb to ZFP42	2.59
chr4	DH9	188699311	188700213	42	902	454 kb to ZFP42	2.52
chr4	DH10	188760478	188761641	42	1163	392 kb to ZFP42	2.46
chr4	DH272	190950856	190951416	45	560	148 kb to FRG1	0.27

B. Unique-sequence DH sites inside or within 20 kb of a RefSeq gene

	DH site name	DH site coordinates		G+C (%)	Length (bp)	Position in gene region	Distance to D4Z4 (Mb)
chr4[3]	**DH0**	**187302067**	**187303834**	**63**	**1767**	**5' region of FAM149A**	**3.92**
chr4	**DH1**	**187350357**	**187350725**	**51**	**368**	**1st Intron of CYP4V2**	**3.87**
chr4	DH2.5_0	187728443	187729227	45	784	15 kb 5' to MTNR1A	3.49
chr4	DH2.5_1	187782947	187783450	37	503	interior intron of FAT1	3.44
chr4	DH2.5_2	187809329	187810057	51	728	interior intron of FAT1	3.41
chr4	DH2.5_3	187824363	187825476	44	1113	interior intron of FAT1	3.40
chr4	DH2.5_4	187827360	187828109	46	749	interior intron of FAT1	3.40
chr4	DH2.5_5	187843853	187844382	41	529	interior intron of FAT1	3.38
chr4	DH2.4	187858206	187860259	51	2053	interior intron of FAT1	3.36
chr4	**DH2.3**	**187883010**	**187884536**	**49**	**1526**	**promoter of FAT1**	**3.34**
chr4	**DH_FRG1**	**191098015**	**191099317**	**60**	**1302**	**promoter of FRG1**	**0.12**

C. DH sites that overlapped an Simple Tandem Repeat (STR)[4]

	DH site name	DH site coordinates		G+C (%)	Length (bp)	Distance from the closest gene	Distance to D4Z4 (Mb)	STR consensus (bp)	No. of tandem copies	STR consensus sequence
chr4	**DH2.6**	**187704727**	**187705595**	**60**	**868**	**1st Intron, MTNR1A**	**3.52**	32	23	TCCACACCACACCCTGTTCCTGGGACACACCG
chr4	DH11	190168604	190170157	51	1553	863 kb from TRIML1	1.05	54	26	TCCAGCTGCCCATTCCCTCTAAAGTGACGTCTGTATCACTCCTGCTCGAAAGTG
chr4	DH12	190209068	190210621	50	1553	888 kb to FRG1	1.01	27	59	GGTGTCTCACTGTAAAAGCTGCACTCG
chr4	DH13	191004109	191005407	62	1298	93 kb to FRG1	0.22	32	38	TCCGTGCCGTGTCCCTCGGCTCTCTCCCATTC
chr4	DH15	191207299	191208180	58	881	15 kb to DUX4	0.015	142	3	CCCTGAGCCTGGTGCATGCTGGGATTGCAGTGCTGCAGCCCTGTGACCAAAGG GCTGGGAGTGTTTATGAGACTGCATCTCCCAGCAAGACCAGCGAGAGGCGCG GAGCCTCGTCCCTTCCTCCAGTGATTAGCGCACTCT

[1]The displayed DH sites were observed in all six examined myoblast cultures (FSHD and control) except for DH272, which was seen preferentially in FSHD myoblasts (Figure S3).

[2]These DH sites did not overlap repeat sequences. The distance is shown for the closest end of a given gene.

[3]The 5 DH sites seen at or near the 5' ends of genes are shown in boldface.

[4]Of the DH sites overlapping STRs, all except DH15 were centered on the STR such that there was >80% overlap. For DH15 the overlap was about 50%.

Table S2. Primers used for qRT-PCR

Name of amplicon[1]	Forward primer	Reverse primer
-13kb	TTGTACAATCTAGCTTTAATGAAGTGCAA	TTGTAGTTCTTAAATGGAGTGTTTCAGA
-12kb	GGGGGACCATAAATACGTGTTAC	AAAACTTTTCTCAAAAAGCCATT
-7kb	GCATCATATCCAGGTCCCATCA	TGGATTCCCCACTTGCTAGAA
-2kb	TCCTCCTCACACGGCTCCCTG	GATGGGTGTTGCCATCAAGGT
-0.2kb	TTTTTCTTTCTGCCTTTCGAGAAC	TCTTTCCGCCTCTCCAATCA
+0.03kb	GGGTCAGATCTTCAGAGAAGTTT	CATAAAATGTATTGTGAGGGCAG
+0.2kb	TTCCTGCACCGTCTTTACCG	AGGGTCCATGACCCAGATGA
+2.8kb	ATCACAAAGATACTTCACTCTCAACTT	CGGCCCAGTCTAATTACATCAA
+3kb	TGGTGTGTTTGTTTCCTTGGG	TGCAAATGCCTGGTACTGCTG
+17kb	TCCAGTTCGCAATCAGTTGAAA	CCAGCTGGTTAATTTAGGGTGCT

[1]These amplicons are described in Figures 6 and S5 and were the indicated distance from DH272