Supporting Information

Rahimov et al. 10.1073/pnas.1209508109

SI Materials and Methods

RNA Isolation and First-Strand cDNA Synthesis. RNA concentration was quantified with UV absorption at 260 nm using a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific), and the RNA integrity was assessed using the RNA 6000 Nano chip on the Agilent 2100 Bioanalyzer (Agilent Technologies). The majority of RNA samples had an RNA integrity number (RIN) of ≥8.0, as determined by Bioanalyzer. Five micrograms of total RNA was subjected to DNase I (Ambion) treatment for 15 min at 37 °C in the presence of RNaseOUT (Invitrogen). DNase I digestion reaction components were removed using the RNeasy Kit (Qiagen) with an additional on-column DNase treatment for 15 min. RNA was eluted with 60 μL of elution buffer (EB), and the volume was reduced with speed vacuum. cDNA was synthesized from the remaining ∼2–3 μg of total RNA using the RevertAid first-strand cDNA synthesis kit (Fermentas) using oligo$(dT)_{18}$ primers in 20-μL reactions, and the final RT reactions were diluted by adding 45 μL of nuclease-free water.

Microarray Analysis. Gene expression profiling was carried out using the Affymetrix GeneChip Human Gene 1.0 ST arrays. The current format of these arrays interrogates 28,869 annotated genes in the human genome with approximately twenty-six 25-mer oligonucleotide probes spread across the full length of the transcript. Biotin-labeled target for the microarray experiment was prepared using 100 ng of total RNA, and cDNA was synthesized using the GeneChip WT (Whole Transcript) Sense Target Labeling and Control Reagents kit as described by the manufacturer (Affymetrix). The sense cDNA was then fragmented by UDG (uracil DNA glycosylase) and APE 1 (apurinic/apyrimidic endonuclease 1) and biotin-labeled with TdT (terminal deoxynucleotidyl transferase) using the GeneChip WT Terminal labeling kit (Affymetrix). Hybridization was performed using 5 μg of biotinylated target, which was incubated with the GeneChip Human Gene 1.0 ST array (Affymetrix) at 45 °C for 16–20 h. Following hybridization, nonspecifically bound material was removed by washing and detection of specifically bound target was performed using the GeneChip Hybridization, Wash and Stain kit, and the GeneChip Fluidics Station 450 (Affymetrix). The arrays were scanned using the GeneChip Scanner 3000 7G (Affymetrix), and raw data were extracted from the scanned images and analyzed with the Affymetrix GeneChip Command Console Software (Affymetrix).

Microarray Data Analysis. The raw array data were preprocessed and normalized using the Robust Multichip Average (RMA) method (1). This procedure includes background correction and quantile normalization of the arrays at the probe level, followed by robust summarization of expression at the transcript level. The reported results are based on only those probesets annotated with Entrez gene IDs and, in cases of multiple probesets, with the same Entrez ID number on only the probeset with the largest interquartile range. The 50 microarrays were processed in 5 batches on different dates, and even after quantile normalization, the global expression patterns clustered strongly by batch. Within batches the arrays clustered weakly by cohort, a trend noted previously for qPCR data from cell cultures derived from a subset of these cultures (2). Because all arrays from each cohort were processed within the same batch, we controlled for both between-batch variability and between-cohort variability by including a factor for "cohort" in our linear model. Additionally, because the sex of the subjects was not balanced between the affected and unaffected subjects (3 male and 10 female vs. 1 male and 11 female), we also included a factor for "gender" to separate sex-related expression differences from disease-related differences. Because FSHD typically affected biceps more severely than deltoid, we used a model that includes interactions between disease class and muscle type. This allowed us to assess the differences between affected and unaffected subjects separately for biceps and deltoid muscle and the differences between biceps and deltoid separately for affected and unaffected subjects. We also assessed the differences between biceps and deltoid of the within-muscle disease-related changes. The full linear model used was "∼0 + class:muscle + cohort + gender." Because FSHD tends to have greater severity for patients with fewer D4Z4 repeats in 4q35, we also looked for genes with expression that was correlated with the number of D4Z4 repeat units, again controlling for cohort and sex and allowing for muscle-specific differences. For this, we used the linear model "∼0 + muscle + contraction:muscle + cohort + gender." Here, the continuous covariate "contraction" was defined as [35 – min (EcoRI, 35)]/(4 × 3.3), where EcoRI is the length in kilobases of the EcoRI/BlnI fragment. Fragment lengths were capped at 35 kb (corresponding to ∼9 D4Z4 units) for unaffected subjects, because we would expect the per-repeat-unit impact on expression levels to be much milder for long arrays than for short arrays. Because each D4Z4 repeat contributes 3.3 kb to the length of the EcoRI fragment, the scaling factor was chosen so that the coefficients represent the estimated change in expression associated with a loss of four D4Z4 repeat units. (The choice of scaling does not affect the t scores or P values.)

To check whether sets of genes with altered expression in previous FSHD-related studies are also altered in our data, we used the ROAST function in limma (3) to test whether the genes in these sets tended to be up-regulated, down-regulated, or differentially expressed without regard to direction. Gene IDs from published lists were mapped to HUGO gene symbols, and matched to microarray probeset IDs annotated with the same gene symbols using the R package hugene10stprobeset.db. P values were computed by random rotations of the data that are compatible with the factors in the linear model (4). We used 1000 random rotations, with the "mean50" function applied to the t-scores as the summary statistic.

Differentially expressed genes were further analyzed for biological relevance using the IPA tool (IPA 9.0) (Ingenuity Systems). A dataset containing Entrez gene identifiers and corresponding expression values of differentially expressed genes was uploaded into the application. Each identifier was mapped to its corresponding object in Ingenuity's Knowledge Base. These molecules, called Network Eligible molecules, were overlaid onto a global molecular network developed from information contained in Ingenuity's Knowledge Base. Networks of Network Eligible Molecules were then algorithmically generated based on their connectivity.

To check whether selected genes (e.g., *IDI2*) were altered in GEO profiles for other muscle disorders, we queried the website (www.ncbi.nlm.nih.gov/geoprofiles) with search strings of the form: "IDI2 [gene symbol] AND rank subset effect [flag type] AND (muscular [GDS text] OR muscle [GDS text]) AND (dystrophy [GDS text] OR disorder [GDS text] OR disease [GDS text])".

Quantitative Real-Time RT-PCR. All qPCR reactions were run in quadruplicate. Expression levels of the target genes were normalized relative to the geometric mean of two internal control

genes (5), which were selected using the GeNorm software (http://medgen.ugent.be/~jvdesomp/genorm/) from among 11 commonly used endogenous control genes (*ACTB*, *GAPDH*, *B2M*, *18S*, *PPIA*, *HPRT1*, *GUSB*, *TBP*, *RPLP0*, *TFRC*, and *PGK1*). In our samples, both affected and unaffected biceps and deltoid, *GUSB* and *PPIA* were the most stable internal control genes. After normalization, $\log_2$ (fold changes) and P values were computed using limma as described in the microarray methods, but with biceps and deltoid samples fit with separate linear models, and using ordinary t-statistics rather than empirical Bayes–moderated t statistics. (Sharing information on variances between genes and between muscle types for the microarrays tends to moderate the influence of outliers, but can also bias estimates, so was not done for the qPCR data.) The "cohort" factor was omitted from the linear models for the validation cohorts reported in Table 3, as not all of these included unaffected individuals. Alternately, modeling "cohort" as a random effect in a mixed-effect linear model (using the R package lme4) accommodates partial cohorts, and shrinks cohort effects toward a common mean. This gave results similar to those in Table 3 for most genes; P values were >twofold smaller than in Table 3 for *G0S2* and *SAMHD1* (biceps and deltoid) and *ACTA2* and *CAB39L* (just deltoid), although this may be attributable, in part, to the asymptotic χ^2 distribution giving "anticonservative" P values for the likelihood-ratio test (6).

To test whether a linear combination of normalized Ct values from the 15 genes in the qPCR panel (control genes excluded) could accurately discriminate FSHD from control biceps samples, we trained a classifier on only those samples used in the microarray study. This was done with the R package glmpath (7) using logistic regression with an L_1 regularization parameter λ selected by the Bayesian Information Criterion to control overfitting. The resulting classifier, which had nonzero coefficients for 9 of the 15 genes (see below), classified all of the training samples correctly, although this may still reflect some degree of overfitting. (Note that even assessing these samples with cross-validation would be biased given that the genes in the qPCR panel were selected based on microarray results for these samples.) The classifier had 91% accuracy (1 false positive, 1 false negative, 12 true positives, 8 true negatives) on the 22 validation biceps samples, which were used neither in selecting the qPCR genes nor in fitting the classifier. (This gives $P = 0.00006$ with a simple binomial test and $P = 0.001$ with a binomial test that accounts for imbalance in class sizes by setting the probability of success in the null-model to 13/22.) The analogous procedure for the deltoid samples produced a classifier with nonzero coefficients for 10 genes, which gave 81% classification accuracy (2 false positives, 3 false negatives, 13 true positives, 8 true negatives) on the 26 validation samples. (This gives $P = 0.001$ with a simple binomial test, and $P = 0.03$ when accounting for the imbalance in class sizes.) Results are shown in Fig. S3. The discriminant score for biceps was -4.1**ACTA2* -1.5**IL32* + 0.02**LBP* $-$ 4.0**MYH8* + 1.9**OXCT1* + 33.3**PFN2* $-$ 1.4**SAMHD*1 $-$ 13.6**SLC25A33* + 2.0**TECRL* + 122.2, and the discriminant score for deltoid was -0.0004**F2R* + 3.9**G0S2* + 8.1**GLT25D2* + 5.1**IDI2* + 2.2**LBP* $-$ 2.1**MYH8* + 8.8**OXCT1* $-$ 4.4**SAMHD1* $-$ 19.0 **SLC25A33* $-$ 6.9**TECRL* + 58.0. In these expressions, a gene symbol represents the normalized Ct value for that gene, and different choices of normalization genes will affect the constant offset term. Moreover, all of the coefficients depend on the efficiencies of the qPCR reactions, which may vary with reagents or experimental conditions. Positive scores correspond to predicted class "FSHD," and negative scores to predicted class "control." In interpreting the coefficients, note that Ct scores are higher when gene expression is lower and that there may be many other combinations that discriminate samples essentially as well; in particular, a coefficient of zero does not mean a gene is of no value for discriminating samples, because L_1 regulation tends to favor sparseness in the vector of regression coefficients.

1. Irizarry RA, et al. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31:e15.
2. Homma S, et al. (2012) A unique library of myogenic cells from facioscapulohumeral muscular dystrophy subjects and unaffected relatives: Family, disease and cell function. *Eur J Hum Genet* 20:404–410.
3. Wu D, et al. (2010) ROAST: Rotation gene set tests for complex microarray experiments. *Bioinformatics* 26:2176–2182.
4. Langsrud Ø (2005) Rotation tests. *Stat Comput* 15:53–60.
5. Vandesompele J, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3: RESEARCH0034.
6. Pinheiro JC, Bates DM (2000) *Mixed-Effects Models in S and S-Plus* (Springer, New York).
7. Park MY, Hastie T (2007) L1-regularization path algorithm for generalized linear models. *J R Stat Soc Series B Stat Methodol* 69:659–677.

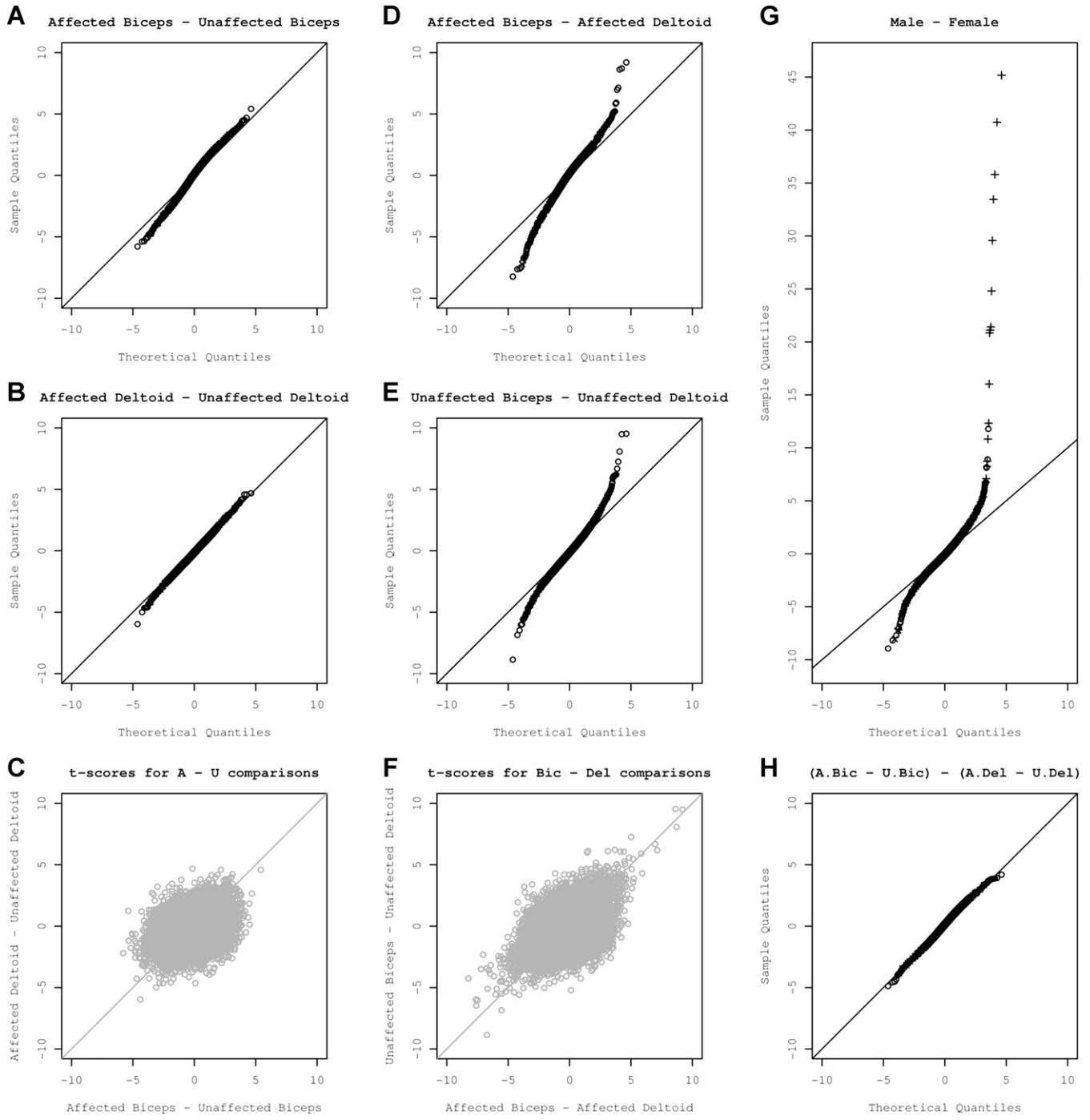

Fig. S1. Overview of gene expression changes for all comparisons. (*A*–*E* and *H*) Quantile–quantile plots showing the observed distribution of *t* scores for differential expression of each of ~20,000 genes (vertical axis) vs. the theoretical distribution of *t* scores one would expect if no genes were differentially expressed (horizontal axis). (The *t* score for each gene is its estimated log fold change between classes divided by the SE of this estimate.) (*C* and *F*) Plots of *t* scores for each gene in one comparison vs. *t* scores in another comparison. For comparisons between affected and unaffected biceps (*A*) and between affected and unaffected deltoid (*B*), the observed distribution of *t* scores closely follows the theoretic distribution, indicating a general lack of significantly altered genes. There is little correspondence between the *t* score for biceps and the *t* score for deltoid (*C*). For comparisons between affected biceps and affected deltoid (*D*), there are many genes with more extreme differences than predicted by the theoretic distribution. However, this is also evident for comparisons between unaffected biceps and deltoid (*E*). The genes differentially expressed between muscle types tend to be the same for both affected and unaffected samples (*F*), and the distribution of differences between the two cases closely follows the theoretical distribution (*H*). Finally, the male vs. female comparison (*G*) shows an extreme difference for ~20 genes on the Y chromosome (indicated by +), as well as a milder effect for many other genes, including several on the X chromosome (indicated by x).

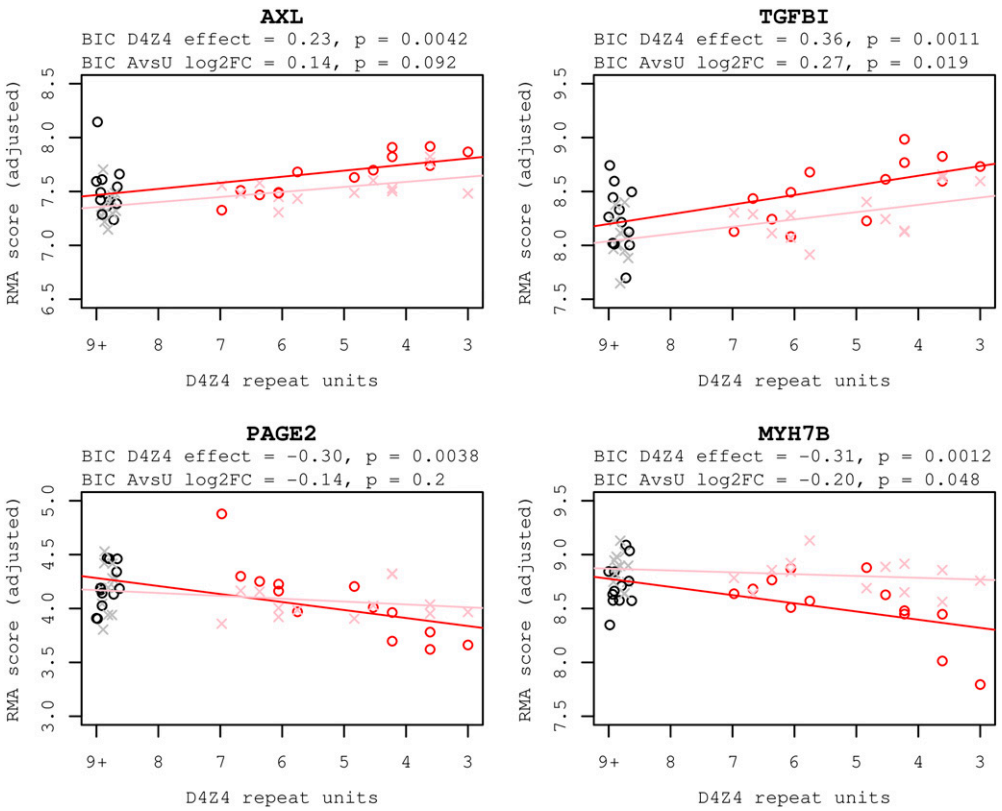

Fig. S2. Four examples of genes for which a D4Z4-length-dependent model gives a better fit to expression data than a simple affected vs. unaffected model. The horizontal axis gives D4Z4 repeat count, with unaffected samples capped at 9, and with slight jitter added to reduce overlap. The vertical axis gives RMA expression levels, with estimated cohort and sex effects removed. Red and black circles represent samples from affected and unaffected biceps, respectively. Pink and gray x symbols represent samples from affected and unaffected deltoid, respectively. Red and pink lines indicate D4Z4-dependent fits to expression levels in biceps and deltoid, respectively. The D4Z4 effect sizes listed above the plots give estimated changes in expression corresponding to a loss of four D4Z4 units. (Note that these effect sizes depend on this choice of scale, but the associated *P* values do not.) The $\log_2$ fold-changes and *P* values from simple affected vs. unaffected tests are included for comparison. AXL, AXL receptor tyrosine kinase; PAGE2, P antigen family, member 2 (prostate associated).

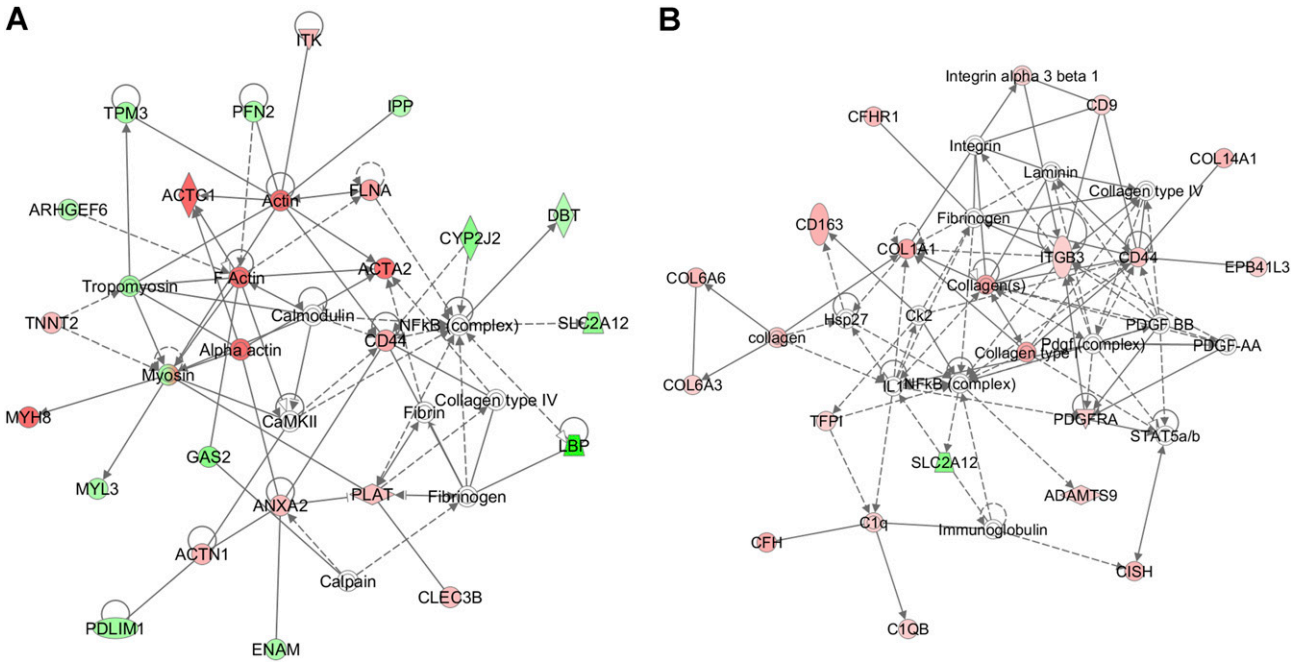

Fig. S3. Schematic diagram of molecular networks constructed from differentially expressed genes and their interacting partners in biceps (*A*) and deltoid (*B*) generated by IPA. Genes involved in tissue development, cardiovascular system development and function, and skeletal and muscular system development and function were altered in biceps. Deltoid showed changes in genes implicated in connective tissue disorders and cellular movement. Nodes represent genes and lines show the relationship between genes. The intensity of the node color indicates the degree of up-regulation (red) or down-regulation (green) of genes. Nodes are displayed using various shapes that represent the functional class of the gene product.

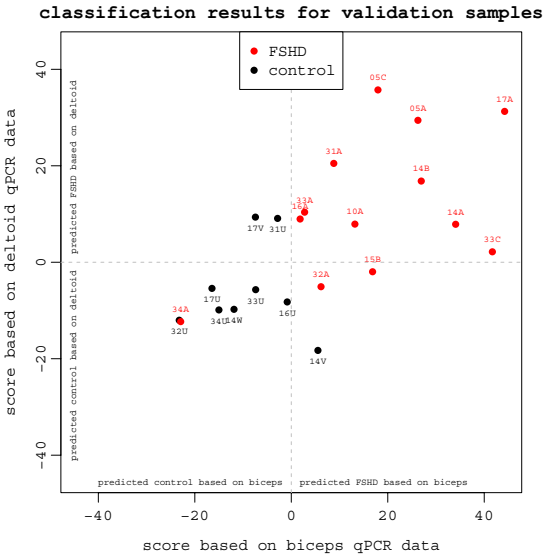

Fig. S4. Samples can be classified as FSHD vs. control based on genes in the qPCR panel. L_1-regularized logistic regression classifiers were trained using only the samples from the microarray study, with the Ct values of genes in the qPCR panel used as predictors. This was done separately for biceps and deltoid samples. Validation samples from 26 individuals not used in the microarray study were then predicted to be of class FSHD or control using these classifiers, which score samples using a linear combination of Ct values for genes in the qPCR panel (see *SI Materials and Methods*). The scores for the 22 validation individuals for which we have both biceps and deltoid data are plotted in the figure, with the biceps score on the vertical axis and deltoid score on the horizontal axis. In both cases positive scores correspond to predicted class "FSHD" and negative scores to predicted class "control." The true class is indicated by the color of the plotted symbol, red for FSHD and black for control. For the biceps predictions, there were one false positive (14V) and one false negative (34A), and for the deltoid predictions, there were two false positives (17V, 31U) and three false negatives (15B, 32A, 34A). (Results are not plotted for the four individuals for which we have only deltoid data; these were predicted correctly, with scores 5.1 for 05B, 12.8 for 23A, −12.8 for 23U, and 20.9 for 34B.)

Table S1. Demographics and clinical data from muscle biopsy donors

Cohort ID	Subject ID	Relationship	Sex	Age, y	EcoRI/BlnI allele (kb)	Biceps strength*	Deltoid Strength*
03	**03A**	Proband	F	40	20	4+/5	5/5
	03U	Sister of 03A	F	42	>40	5/5	5/5
05	05A	Proband	F	55	25	5/5	5/5
	05B	Son of 05A	M	19	25	4/5	4/5
	05C	Brother of 05A	M	49	25	5/5	5/5
07	**07A**	Proband	F	18	29	5/5	4+/5
	07U	Mother of 07A	F	49	34[†]	5/5	5/5
09	**09A**	Proband	F	31	25	4+/5	5/5
	09U	Mother of 09A	F	57	47	5/5	5/5
10	10A	Proband	F	48	24	5/5	5/5
12	**12A**	Daughter of 12B	F	22	18	4+/5	5/5
	12B[‡]	Proband	M	49	18	4+/5	5/5
	12U	Daughter of 12B	F	24	>112	5/5	5/5
	12V[‡]	Sister of 12B	F	45	>112	5/5	5/5
13	**13B**	Proband	F	42	16	4/5	5/5
	13U	Mother of 13B	F	63	57	5/5	5/5
14	14A	Proband	M	50	19	4/5	5/5
	14B	Brother of 14A	M	53	19	4+/5	4+/5
	14V	Sister of 14A	F	49	60	5/5	5/5
	14W	Brother of 14A	M	47	72	5/5	5/5
15	**15A**	Proband	M	66	28	4+/5	5/5
	15B	Brother of 15A	M	69	28	5/5	5/5
	15V	Sister or 15A	F	60	107	5/5	5/5
16	16A	Proband	F	56	20	4+/5	5/5
	16U	Sister of 16A	F	60	>59	5/5	5/5
17	17A	Proband	M	23	19	5/5	4/5
	17U	Brother of 17A	M	21	97	5/5	5/5
	17V	Father of 17A	M	50	90	5/5	5/5
18	**18A**	Proband	F	36	21	4+/5	5/5
	18U	Brother of 18A	M	37	57	5/5	5/5
19	**19A**	Proband	M	65	22	4+/5	4+/5
	19U	Daughter of 19A	F	41	79	5/5	5/5
20	**20A**	Proband	M	28	20	5/5	5/5
	20U	Mother of 20A	F	48	39	5/5	5/5
21	**21A**	Proband	F	82	26	4+/5	4+/5
	21B	Daughter of 21A	F	59	26	4+/5	5/5
	21U	Daughter of 21A	F	48	63	5/5	5/5
22	**22A**	Proband	F	71	27	4+/5	5/5
	22U	Daughter of 22A	F	43	60	5/5	5/5
23	23A	Proband	M	27	18	2+/5	5/5
	23U	Father of 23A	M	59	45	5/5	5/5
31	31A	Proband	F	31	18	5/5	5/5
	31U	Mother of 31A	F	63	57	5/5	5/5
32	32A	Proband	F	64	30	5/5	5/5
	32U	Daughter of 32A	F	39	147	5/5	5/5
33	33A	Proband	F	49	20	5/5	5/5
	33C	Brother of 33A	M	51	20	4/5	5/5
	33U	Sister of 33A	F	45	77	5/5	5/5
34	34A	Proband	F	70	15	5/5	5/5
	34B	Son of 34A	M	39	15	3/5	5/5
	34U	Daughter of 34A	F	34	97	5/5	5/5

Donors are designated by cohort (family) number (03, 05, etc.), followed by A, B, or C for the FSHD subjects or U, V, or W for the unaffected first-degree relative(s). FSHD was defined by presence of both clinically apparent muscle weakness and a shortened 4q D4Z4 repeat array identified by an EcoRI/BlnI restriction fragment of <35 kb. Cohorts and subjects analyzed by microarray and qPCR are highlighted in bold. Cohorts analyzed independently and by only qPCR are not highlighted in bold. F, female; M, male.

*Muscle strength of biopsied muscles is presented using a modified MRC scale, where 5/5 is full strength.

[†]The 4q35 deletion in this subject is allele type 4qB, which is not associated with FSHD.

[‡]Subjects 12B and 12V were originally named 11A and 11U for internal analyses but were renamed for publication to reflect familial relationship.

Table S2. Comparison with previously published lists of FSHD-related genes

	Biceps, *P*			Deltoid, *P*		
Study and gene list	Mixed	Up	Down	Mixed	Up	Down
Arashiro et al. (1)						
AvsC.Up	**0.010**	**0.004**	0.967	0.027	**0.012**	0.968
AvsC.Down	**0.002**	0.171	**0.001**	0.16	0.507	0.046
AvsU.Up	**0.009**	**0.004**	0.993	**0.003**	**0.002**	0.998
AvsU.Down	**0.001**	0.706	**0.001**	0.089	0.59	0.038
CvsU.Up	**0.006**	**0.004**	0.893	0.236	0.204	0.773
Wallace et al. (2)						
p53 targets.DUX4.Up	0.210	0.218	0.532	0.055	0.508	0.030
Apoptotic.DUX4.Up	0.066	0.027	0.853	0.044	0.079	0.368
Bosnakovski et al. (3)						
DUX4.4hr.Up	**0.005**	0.142	0.022	**0.012**	0.027	0.230
DUX4.12hr.Up	**0.005**	0.557	**0.006**	0.055	0.156	0.210
DUX4.4hr.Down	0.017	**0.008**	0.968	0.100	0.542	0.186
DUX4.12hr.Down	**0.009**	0.043	0.05	0.119	0.793	0.034
Kumar et al. (4)						
PAX3.PAX7.C2C12.Up	**0.003**	**0.005**	0.721	0.059	0.045	0.440
Winokur et al. (5)						
FSHD.Down	0.097	0.031	0.97	0.393	0.212	0.621
FSHD.Up	0.085	**0.008**	0.983	**0.019**	**0.004**	0.947
FSHD.BvsD.Down	0.325	0.107	0.940	0.302	0.108	0.983
FSHD.BvsD.Up	0.047	0.030	0.625	0.691	0.242	0.969
FSHD.MyoD.Down	0.216	0.076	0.949	0.555	0.252	0.723
FSHD.MyoD.Up	0.128	0.039	0.853	0.524	0.215	0.541
FSHD.Myog.Diff.Down	0.017	0.141	0.062	0.137	0.927	0.028
FSHD.Myog.Diff.Up	0.481	0.465	0.479	0.436	0.153	0.979
Dixit et al. (6)						
FSHD.Specific.Up	0.279	0.098	0.716	0.331	0.124	0.509
FSHD.Specific.Down	0.417	0.811	0.19	0.518	0.264	0.737
Osborne et al. (7)						
FSHD.Up.Vascular	**0.002**	**0.002**	0.832	**0.014**	**0.007**	0.987
FSHD.Up	**0.011**	**0.007**	0.936	**0.020**	**0.010**	0.950
FSHD.Down	**0.001**	0.785	**0.001**	0.041	0.325	0.167
FSHD.Specific.Up	**0.011**	**0.009**	0.446	**0.015**	**0.010**	0.662
Geng et al. (8)						
DUX4-fl.Up.3.fold	**0.003**	0.034	**0.006**	0.212	0.408	0.251
DUX4-fl.Down.3.fold	0.021	**0.009**	0.914	0.085	0.07	0.669
DUX4-s.Up.3.fold	**0.001**	0.022	0.258	0.363	0.173	0.776
DUX4-s.Down.3.fold	0.541	0.284	0.717	0.789	0.421	0.58
DUX4-fl.Up.Germ.Cell	**0.001**	**0.001**	0.988	0.519	0.402	0.545

Myog.Diff, myogenic differentiation; AvsC, affected vs. unaffected carriers; AvsU, affected vs. unaffected; Down, down-regulated genes; mixed, direction not specified; Up, up-regulated genes. *P* values of <0.01 are highlighted in bold.

1. Arashiro P, et al. (2009) Transcriptional regulation differs in affected facioscapulohumeral muscular dystrophy patients compared to asymptomatic related carriers. *Proc Natl Acad Sci USA* 106:6220–6225.
2. Wallace LM, et al. (2011) DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann Neurol* 69:540–552.
3. Bosnakovski D, et al. (2008) An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J* 27:2766–2779.
4. Kumar D, Shadrach JL, Wagers AJ, Lassar AB (2009) Id3 is a direct transcriptional target of Pax7 in quiescent satellite cells. *Mol Biol Cell* 20:3170–3177.
5. Winokur ST, et al. (2003) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum Mol Genet* 12:2895–2907.
6. Dixit M, et al. (2007) DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc Natl Acad Sci USA* 104:18157–18162.
7. Osborne RJ, Welle S, Venance SL, Thornton CA, Tawil R (2007) Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology* 68:569–577.
8. Geng LN, et al. (2012) DUX4 activates germline genes, retroelements, and immune mediators: Implications for facioscapulohumeral dystrophy. *Dev Cell* 22:38–51.

Other Supporting Information Files

Dataset S1 (XLS)
Dataset S2 (XLS)