Transcriptional profiling in facioscapulohumeral muscular dystrophy to identify candidate biomarkers

Fedik Rahimov[a,b,1], Oliver D. King[b,c,1], Doris G. Leung[d,e], Genila M. Bibat[d], Charles P. Emerson, Jr.[b,c], Louis M. Kunkel[a,b,f,2], and Kathryn R. Wagner[d,e,g,2]

[a]Program in Genomics, Division of Genetics, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115; [b]The Senator Paul D. Wellstone Muscular Dystrophy Cooperative Research Center and [c]Boston Biomedical Research Institute, Watertown, MA 02472; [d]Hugo W. Moser Research Institute at Kennedy Krieger Institute, Baltimore, MD 21205; Departments of [e]Neurology and [g]Neuroscience, The Johns Hopkins School of Medicine, Baltimore, MD 21205; and [f]The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA 02115

Contributed by Louis M. Kunkel, June 4, 2012 (sent for review May 24, 2012)

Facioscapulohumeral muscular dystrophy (FSHD) is a progressive neuromuscular disorder caused by contractions of repetitive elements within the macrosatellite D4Z4 on chromosome 4q35. The pathophysiology of FSHD is unknown and, as a result, there is currently no effective treatment available for this disease. To better understand the pathophysiology of FSHD and develop mRNA-based biomarkers of affected muscles, we compared global analysis of gene expression in two distinct muscles obtained from a large number of FSHD subjects and their unaffected first-degree relatives. Gene expression in two muscle types was analyzed using GeneChip Gene 1.0 ST arrays: biceps, which typically shows an early and severe disease involvement; and deltoid, which is relatively uninvolved. For both muscle types, the expression differences were mild: using relaxed cutoffs for differential expression (fold change ≥1.2; nominal *P* value <0.01), we identified 191 and 110 genes differentially expressed between affected and control samples of biceps and deltoid muscle tissues, respectively, with 29 genes in common. Controlling for a false-discovery rate of <0.25 reduced the number of differentially expressed genes in biceps to 188 and in deltoid to 7. Expression levels of 15 genes altered in this study were used as a "molecular signature" in a validation study of an additional 26 subjects and predicted them as FSHD or control with 90% accuracy based on biceps and 80% accuracy based on deltoids.

skeletal muscle | microarray

Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant muscular dystrophy affecting ∼1 in 7,000–20,000 individuals. It is characterized by progressive weakness and wasting of facial, shoulder girdle, and upper-arm muscles from which the disease takes its name, and also trunk, hip, and leg muscles in some patients (1). One of the hallmarks of FSHD is asymmetrical and selective degeneration of skeletal muscles. For example, the biceps muscle is involved early and severely, whereas the more proximal deltoid muscle is relatively spared. The underlying mechanism of this distinctive sparing of certain muscle types is unknown. In addition to muscle degeneration, abnormalities in retinal vasculature and hearing loss are observed in up to 49% and 64%, respectively, in some populations (2). Although possible underlying causes of extramuscular involvement and characteristic facial weakness have been speculated (3), the mechanism of these aspects of the disease is also unclear.

FSHD is caused by partial deletion of a critical number of repeat elements within the highly polymorphic macrosatellite D4Z4 on the subtelomeric region of chromosome 4q (4, 5). In unaffected individuals, the D4Z4 array usually consists of 11–100 repeats (corresponding to EcoRI fragments of 41–350 kb), whereas FSHD patients carry 1–10 repeats (corresponding to EcoRI fragments of 10–35 kb) (6). A small repeat size is associated with greater disease severity and an earlier age of onset (7, 8). The disease-causing D4Z4 deletions must, moreover, occur on chromosomal allele 4qA, because deletions on the equally common 4qB allele do not result in FSHD.

Several studies have demonstrated the myopathic potential of double homeobox 4 (*DUX4*), a gene located within each repeat element, in skeletal muscle cells. Overexpression of *DUX4*, as a result of chromatin relaxation within D4Z4, was initially proposed by Kowaljow et al. (9) to induce toxicity to muscle cells, potentially leading to muscle degeneration in vivo. Subsequent studies demonstrated further evidence to support this finding (10). Recently, genetic analysis of families carrying rare translocations between 4q and 10q chromosomes identified single-nucleotide polymorphisms (SNPs) in the pLAM region adjacent to the distal D4Z4 repeat that segregate with FSHD. These SNPs create a canonical polyadenylation signal on the permissive chromosomal allele, whereas the nonpermissive alleles lack these SNPs. *DUX4* transcripts expressed from the distal-most repeat extend into the pLAM sequence and are polyadenylated when the poly(A) signal SNPs are incorporated into the transcripts, thus increasing their intracellular stability (11). DUX4, a double-homeodomain–containing protein, shares similarities with transcription factors paired box 3 (PAX3) and paired box 7 (PAX7) and is proposed to act as a transcriptional activator (10, 12).

There is currently no pharmacological treatment available for FSHD, and clinical trials with novel therapeutics have been discouraged by the lack of a recognized mouse model. Clinical trials have also been discouraged by the fact that FSHD is a highly variable and slowly progressive disease, whereas the efficacy of therapeutic interventions is ideally established over short periods of time. Therefore, molecular biomarkers of FSHD that could be used to assay responsiveness to therapy would greatly facilitate FSHD therapeutic development and clinical research. High-density oligonucleotide arrays reliably quantify the expression levels of thousands of genes simultaneously and enable identification of such biomarkers (13). To identify mRNA-based biomarkers, we performed gene expression analysis in two distinct muscles from individuals with FSHD and their unaffected first-degree relatives. Using this unique approach, we identified a "molecular signature" of FSHD muscles, and by comparing gene expression patterns between differentially affected muscles, we gained insight into genes and pathways involved in disease progression that will increase our understanding of pathogenesis in this condition.

Author contributions: F.R., O.D.K., C.P.E., L.M.K., and K.R.W. designed research; F.R., O.D.K., D.G.L., and G.M.B. performed research; D.G.L., G.M.B., and K.R.W. contributed new reagents/analytic tools; F.R., O.D.K., C.P.E., L.M.K., and K.R.W. analyzed data; and F.R., O.D.K., L.M.K., and K.R.W. wrote the paper.

The authors declare no conflict of interest.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE36398).

[1]F.R. and O.D.K. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: kunkel@enders.tch.harvard.edu or wagnerk@kennedykrieger.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1209508109/-/DCSupplemental.

Results

Clinical Observations and Muscle Biopsy Collection. In a protocol approved by The Johns Hopkins Medicine Institutional Review Board, families with FSHD were recruited and provided written informed consent to participate. Skeletal muscle was biopsied in 29 FSHD subjects and 22 unaffected first-degree relatives from 21 independent families (or cohorts), several of which have been introduced previously (14), with D4Z4 repeat size determined in all individuals as described previously (15) (Table S1). This particular study design allows pairwise comparison between affected and unaffected family members, which can potentially reduce the effects of gene expression variability caused by genetic differences between individuals from different familial backgrounds. Muscles with advanced degeneration are often infiltrated with fat and fibrosis, and this can skew the expression patterns specific to muscle. To diminish potential bias caused by fat and fibrosis, only mildly affected muscles were included in our study. Furthermore, because the biceps muscle of FSHD patients is generally more severely affected than deltoid, we obtained biopsies from both biceps and deltoid muscles from all individuals with the goal of comparing gene expression signatures to help pinpoint mechanisms of selective muscle degeneration. Genome-wide expression profiling was performed on 11 cohorts (13 affected and 12 unaffected individuals), and the remaining 10 cohorts (16 affected and 10 unaffected individuals) were used for independent evaluation of candidate biomarker genes.

Genome-Wide Gene Expression Profiling of Biceps and Deltoid Muscle Biopsies. Using GeneChip Human Gene 1.0 ST arrays, we determined genome-wide gene expression profiles of mRNA isolated from FSHD and control biceps and deltoid muscle tissues. Applying empirical Bayes-moderated t tests, we identified 191 and 110 genes differentially expressed (using fold change ≥ 1.2 and $P < 0.01$ as criteria for differential gene expression) between FSHD and control biceps and deltoid tissues, respectively (Fig. 1 and Dataset S1). Of these genes, 79 and 83 were up-regulated, and 112 and 27 genes were down-regulated in FSHD biceps and deltoid muscles, respectively. After adjusting for multiple hypotheses testing using the false-discovery rate (FDR) method allowing for up to an estimated 25% of false-positive hits, expression differences of 188 transcripts in biceps remained statistically significant, whereas in deltoids, only 7 genes passed this FDR cutoff value. In contrast to previous studies conducted using oligonucleotide arrays to assess gene expression in FSHD muscles and cells, expression differences observed in our cohorts were not as dramatic. In biceps, no gene showed more than a twofold difference between affected and unaffected samples, and in deltoid, only a single gene did: HMGCS2 (3-hydroxy-3-methylglutaryl-CoA synthase 2, mitochondrial). Taken together, these results demonstrate the impact of D4Z4 deletions on global gene expression patterns in differentially affected skeletal muscles and provide us with a list of potential disease biomarkers.

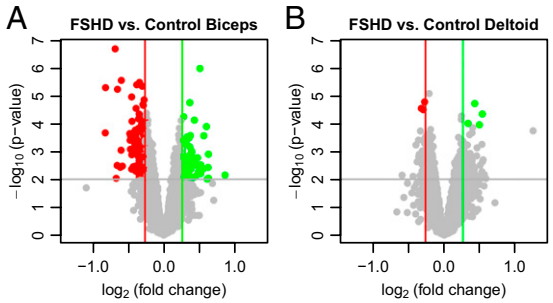

Fig. 1. Volcano plots of expression differences between FSHD subjects and controls in biceps (*A*) and deltoid muscles (*B*). The horizontal axis shows $\log_2$ (fold change), with vertical lines indicating cutoffs of 1.2-fold change either down (red) or up (green) relative to controls. The vertical axis shows $-\log_{10}$ (P value), with the horizontal gray line indicating the cutoff of nominal $P = 0.01$. Genes that passed both cutoffs and also satisfied FDR < 0.25 are colored red (down in FSHD) or green (up in FSHD).

Muscle-Type–Specific Patterns of Global Gene Expression. Because of differential muscle involvement within the same individual, comparison of gene expression patterns between biceps and deltoid allows us to potentially examine the underlying mechanisms of disease progression at the molecular level. Out of 191 and 110 differentially expressed genes in biceps and deltoids, 29 genes were dysregulated in both affected muscle types relative to their unaffected counterparts. These 29 genes showed similar fold changes in both muscle types. Overall, affected deltoid had 42% fewer differentially expressed genes compared with affected biceps, which could be suggestive of a relationship to the state of muscle pathology or to pathogenesis. Comparing affected biceps directly to affected deltoid gave many more significant differences, with 359 genes differentially expressed even when controlling for FDR < 0.25. However, there were also 265 significant differences (FDR < 0.25) between biceps and deltoid even among unaffected samples, which included 129 of the 359 genes from the affected biceps versus affected deltoid comparisons. We also tested directly for genes with changes in expression between biceps and deltoid that were significantly different between affected and unaffected samples. (Note that this is equivalent to testing for genes whose change in expression between affected and unaffected samples was significantly different between biceps and deltoid.) There were 125 genes that satisfied the relaxed cutoff of fold change ≥ 1.2 and nominal $P < 0.01$, but none was significant at FDR < 0.25. Fig. S1 gives an overview of the changes observed in the between-disease-state and between-muscle-type comparisons. Overall, these findings suggest that differences in gene expression patterns between biceps and deltoid might make biceps muscles more susceptible than deltoid muscle to damage caused by D4Z4 deletions. These results are also in concordance with earlier work that demonstrated inherent differences in global gene expression between distinct types of skeletal muscle in the human (16).

Expression Changes Associated with Residual D4Z4 Repeat Length. Because FSHD tends to be more severe for patients with fewer D4Z4 units, we tested for genes with expression levels that tended to either increase or decrease with decreasing D4Z4 repeat length. If there is such an effect, this test should be more sensitive than one that treats all affected subjects uniformly. Our FSHD samples had a range from 3 to 7 D4Z4 units estimated from EcoRI/BlnI restriction-fragment size (Table S1), and for these tests, we capped the D4Z4 units for unaffected subjects at 9 (described in *SI Materials and Methods*). We found 199 genes in biceps and 146 genes (Dataset S2) in deltoid for which the expression level was associated with the D4Z4 repeat length using cutoffs $P < 0.01$ and effect size > 1.2. Here, the effect size was the estimated fold change corresponding to a loss of four D4Z4 repeats. For most genes passing the relaxed cutoffs, the simpler affected vs. unaffected model fit the expression data essentially as well. Several genes for which the D4Z4 length-dependent model gave the most pronounced improvement in fit are shown in Fig. S2. Controlling for multiple hypotheses testing using FDR < 0.25 retained 182 and 13 genes statistically significant in biceps and deltoid, respectively. Thus, similar to disease phenotype, the observed molecular signatures also appear to reflect D4Z4 residual size-dependent changes in gene expression in FSHD muscles.

Biological Pathways Disrupted in FSHD Muscles. To explore biological pathways disrupted in FSHD, we performed Ingenuity Pathway Analysis (IPA) of the genes differentially expressed more than 1.2-fold ($P < 0.01$) in muscles between affected and

Table 1. Canonical pathways overrepresented among genes differentially expressed (fold-change, ≥1.2; nominal *P* value, <0.01) in FSHD biceps compared with control muscles according to IPA

Pathway	Ratio*	*P*
Actin cytoskeleton signaling	10/238	2.05×10^{-5}
ILK signaling	8/193	2.48×10^{-4}
Hepatic fibrosis/hepatic stellate cell activation	7/147	2.53×10^{-4}
Atherosclerosis signaling	5/107	1.43×10^{-3}
Leukocyte extravasation signaling	7/199	1.55×10^{-3}
Intrinsic prothrombin activation pathway	3/34	2.34×10^{-3}
Complement system	3/35	3.08×10^{-3}
Synthesis and degradation of ketone bodies	2/17	3.36×10^{-3}
Coagulation system	3/38	3.64×10^{-3}
Calcium signaling	6/207	4.61×10^{-3}

*Ratio, no. of differentially expressed genes from microarray out of total no. of genes associated with the canonical pathway according to IPA.

unaffected individuals. The most significantly altered 10 canonical pathways in biceps and deltoid are listed in Tables 1 and 2, respectively. Furthermore, molecular networks altered in muscles from affected individuals are depicted in Fig. S3. The expression levels of 10 genes {*MYH8*, *ACTA2*, *ARHGEF6* [Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6], *PFN2*, *LBP*, *ACTC1* (actin, alpha, cardiac muscle 1), *SLC9A1* [solute carrier family 9 (sodium/hydrogen exchanger), member 1], *MYL3* (myosin, light chain 3), *ACTN1* (actinin, alpha 1), and *FGF6*} known to play a role in actin cytoskeleton were altered in affected vs. unaffected biceps, making this pathway, the members of which have also been demonstrated previously to be altered in FSHD muscles (17), a potential biomarker for FSHD. Note that several of these genes are altered in other muscle disorders including Duchenne muscular dystrophy (DMD) (18, 19), α-sarcoglycan deficiency (αSGD) (18), and polymyositis (20). Another significantly changed pathway in biceps was the tissue fibrosis pathway {*ACTA2*, *CCL5* [chemokine (C-C motif) ligand 5], *COL1A1* (collagen, type I, alpha 1), *COL3A1* (collagen, type III, alpha 1), *LBP*, *MYH8*, and *MYL3*}, perhaps reflecting fibrotic infiltration of muscle tissue, a commonly observed reaction to chronic myopathic processes (18–21). The expression levels of two collagen genes, *COL1A1* and *COL3A1*, have been shown to be elevated in affected vastus lateralis (17) and biceps (22) muscles in previous studies as well. Interestingly, this pathway was not changed in the spared deltoid muscle. Pathways perturbed in deltoid were slightly different from those affected in biceps and typically included fewer differentially expressed genes than the pathways for biceps. The complement system pathway was the most significantly altered pathway in deltoid, with four overexpressed genes [*C6* (complement component 6), *C1QB* (complement component 1, q subcomponent, B chain), *CFH* (complement factor H), and *C3AR1* (complement component 3a receptor 1)]. C1QB and CFH levels are also elevated in DMD (18, 19). The complement system is involved in opsonization, lysis of foreign cells, clearance of immune complexes and apoptotic cells, activation of inflammation, and augmenting the antibody response.

Table 2. Canonical pathways affected in FSHD deltoid compared with unaffected deltoid according to IPA

Pathway	Ratio*	*P*
Complement system	4/35	3.16×10^{-5}
Urea cycle and metabolism of amino groups	3/78	7.21×10^{-4}
Synthesis and degradation of ketone bodies	2/17	1.32×10^{-3}
Extrinsic prothrombin activation pathway	2/20	3.45×10^{-3}
Intrinsic prothrombin activation pathway	2/34	1.19×10^{-2}
Coagulation system	2/38	1.60×10^{-2}
Alanine and aspartate metabolism	2/82	1.96×10^{-2}
Glioma invasiveness signaling	2/60	3.96×10^{-2}
Glutamate receptor signaling	2/69	3.96×10^{-2}
Butanoate metabolism	2/127	4.09×10^{-2}

*Ratio, no. of differentially expressed genes from microarray out of total no. of genes associated with the canonical pathway according to IPA.

Comparisons with Previous Studies. Several microarray studies related to FSHD have been reported previously, including studies of FSHD biopsies and cells (17, 22, 23) and studies of expression changes caused by overexpression of FSHD-related genes in model organisms or cell cultures (10, 24, 25). Many of these studies reported stronger expression changes than our study. There is the possibility that the lists of genes identified as differentially expressed in the earlier studies are, as a whole, subtly altered in our data, even if no single gene reached significance. To investigate this possibility, we used ROAST (rotation gene set testing) (26) to test whether the lists of genes tended to have higher-than-average up-regulation or down-regulation in our data and also whether there was higher-than-average misregulation in either direction. These tests were performed separately for biceps and deltoid. Significant correlation ($P < 0.01$) between both up- and down-regulated genes determined in our study and in studies of Arashiro et al. (23) and Osborne et al. (17) were observed for both muscle types. To test the hypothesis of whether overexpression of DUX4 interferes with the regulation of PAX3/7 target genes, we also compared genes differentially expressed in FSHD with putative PAX3/7 target genes determined by Kumar et al. (27) and found a significant tendency for up-regulation of these genes in FSHD biceps. We also tested whether the genes that were differentially expressed upon DUX4 overexpression in the study of Geng et al. (28) were altered in our data. The full sets of DUX4 targets did have a tendency for misregulation in our biceps data but without strong consistency in the direction of the changes. The subset of DUX4 target genes involved in germ cells and early development did, however, show a highly significant tendency for up-regulation in our FSHD biceps data, consistent with the findings of Geng et al. Table S2 summarizes the results of these tests.

Validation of Array Results with Real-Time Quantitative PCR. The dynamic range of differential gene expression detection by microarrays is inherently lower than real-time quantitative (q) PCR. To verify the accuracy and performance of the microarrays, as well as the algorithms we used to analyze the array data, we measured the expression levels of 15 genes using a real-time qPCR approach. In particular, we focused our analysis on potential biomarker candidates that were selected to include both strongly up-regulated and down-regulated genes with significant *P* values. Comparison of fold-change values determined by arrays and qPCR for these genes is listed in Table 3. With the exception of the perinatal myosin heave chain gene (*MYH8*), fold-change values determined by arrays and qPCR were in the same range, with qPCR showing slightly higher values. *MYH8* appeared to be more dramatically up-regulated in FSHD muscles, as determined by qPCR compared with microarrays, likely owing to its low abundance in adult skeletal muscles and the limited capacity of arrays to measure low abundance mRNA. Overall, there was strong correlation between microarray and qPCR results, with a correlation coefficient of 0.89 and 0.90 in biceps and deltoid samples, respectively (Fig. 2).

To further evaluate the consistency of the biomarker candidate genes identified in the first 11 cohorts, we assessed their expression levels in an independent sample set comprising 25 individuals from 10 additional cohorts and one additional member of cohort 15. The correlation coefficient between fold-change values of the

Table 3. Validation of microarray results with real-time qPCR

		FSHD vs. control biceps						FSHD vs. control deltoid					
		Array		qPCR-1		qPCR-2		Array		qPCR-1		qPCR-2	
Gene symbol	Gene name	FC	*P*	FC	*P*	FC	*P*	FC	*P*	FC	*P*	FC	*P*
LBP	Lipopolysaccharide binding protein	−1.8	0.0002	−2.4	0.03	−1.5	0.32	−1.4	0.03	−2.4	0.009	−1.1	0.78
IDI2	Isopentenyl-diphosphate delta isomerase 2	−1.6	0.0000002	−1.6	0.02	−1.9	0.10	−1.2	0.02	−1.6	0.008	−1.9	0.18
OXCT1	3-Oxoacid CoA transferase 1	−1.6	0.003	−1.3	0.22	−1.1	0.73	−1.5	0.007	−1.7	0.03	−1.1	0.50
TECRL	*Trans*-2,3-enoyl-CoA reductase-like	−1.5	0.000003	−1.4	0.004	−1.6	0.20	−1.1	0.45	−1.3	0.09	−1.7	0.24
G0S2	G_0/G_1 switch 2	−1.3	0.06	−1.6	0.11	−2.5	0.01	−1.3	0.02	−2.0	0.05	−2.6	0.006
GLT25D2	Glycosyltransferase 25 domain containing 2	−1.3	0.000004	−1.4	0.02	−1.5	0.04	−1.1	0.08	−1.5	0.009	−1.5	0.19
EXTL1	Exostoses (multiple)-like 1	−1.3	0.0002	−1.7	0.005	−1.5	0.06	−1.1	0.04	−1.5	0.01	−1.7	0.12
PFN2	Profilin 2	−1.2	0.0002	−1.2	0.05	−1.2	0.14	−1.1	0.06	−1.3	0.03	−1.4	0.21
F2R	Coagulation factor II (thrombin) receptor	+1.2	0.002	+1.7	0.08	+1.0	0.97	+1.1	0.06	+1.1	0.74	−1.1	0.61
IL32	Interleukin 32	+1.1	0.45	+1.4	0.33	−1.1	0.70	+1.3	0.05	+1.4	0.31	−1.1	0.81
SLC25A33	Solute carrier family 25, member 33	+1.2	0.0004	+1.3	0.08	+1.0	0.90	+1.1	0.02	+1.3	0.08	+1.0	0.83
SAMHD1	SAM domain and HD domain 1	+1.3	0.001	+1.5	0.009	+1.2	0.23	+1.3	0.0025	+1.1	0.32	+1.4	0.04
MYH8	Myosin, heavy chain 8, skeletal muscle, perinatal	+1.5	0.01	+6.1	0.01	+11.5	0.00002	+1.5	0.02	+4.0	0.04	+2.3	0.15
ACTA2	Actin, α2, smooth muscle, aorta	+1.5	0.0001	+1.9	0.0005	−1.1	0.54	+1.2	0.1	+1.3	0.23	+1.5	0.09
CAB39L	Calcium-binding protein 39–like	+1.2	0.07	+1.6	0.03	+1.1	0.51	+1.2	0.05	+1.2	0.17	+1.5	0.02

FC, fold change; qPCR-1, qPCR data of cohorts analyzed with microarrays; qPCR-2, independent validation qPCR.

second qPCR data and arrays in biceps and deltoid samples was 0.62 and 0.74, respectively. For some genes, the difference in expression levels between affected and unaffected samples did not reach statistical significance, perhaps attributable, in part, to increased variability from samples with increased fat infiltration or fibrosis. By combining the Ct values from the 15 genes in the qPCR panel into a single score using weights chosen based only on the 11 original cohorts, we were able to classify the validation samples as FSHD or control with ∼90% accuracy for biceps and ∼80% accuracy for deltoid (Fig. S4).

Discussion

Although D4Z4 deletions on chromosome 4q were discovered two decades ago (4), the molecular pathogenesis of FSHD is still arguably the least well-understood of the muscular dystrophies, a fact that is particularly unsatisfying given its relatively high prevalence in the population. Evidence for abnormal overexpression of the putative transcription factor DUX4 in FSHD muscles and myoblasts (11, 29, 30) invokes the hypothesis that dysregulation of its target gene expression might result in muscle degeneration. In the present study, we used a global transcriptomic approach to characterize gene expression patterns in muscle tissues from individuals affected with FSHD. In addition to identifying gene expression differences between affected and unaffected individuals, we also report expression signatures of differentially affected biceps and deltoid muscles.

One of the most salient features of this study using first-degree relative controls is that there are very few differences between normal and diseased muscle. Most of the expression changes between disease and control were subtle. After adjusting for multiple hypotheses testing, expression differences of only 7 genes in deltoid and 188 genes in biceps remained statistically significant. Importantly, none of these was a gene in the vicinity of the 4q35 locus such as *FRG1* (FSHD region gene 1), *FRG2* (FSHD region gene 2), or *SLC25A4* [solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4], suggesting that derepression of genes in *cis* is not a sequel of D4Z4 contraction. Changes in the expression of *DUX4* itself were not detected, as expected because of its extremely low expression level. Indeed, *DUX4* typically requires specialized nested PCR assays for detection (11), and its expression in the current cohorts was recently reported in a separate study (31).

In addition to our study, expression profiling of FSHD muscles have been assessed previously by three independent studies. Based on observed expression changes in both muscles and cells, Winokur et al. (22) concluded that dysregulation of genes involved in myogenic differentiation and genes essential for buffering oxidative stress results in diminished myoblast differentiation coupled with vulnerability to oxidative stress leading to muscle degeneration. Although similar changes in the same pathways were confirmed by Celegato et al. (32) using mRNA and protein expression analyses, Osborne et al. (17) did not observe differences in these pathways. They, on the other hand, identified genes expressed in vascular smooth muscle and endothelium and genes that promote angiogenesis, potentially explaining some of the extramuscular symptoms such as retinal vasculopathy commonly observed in FSHD patients.

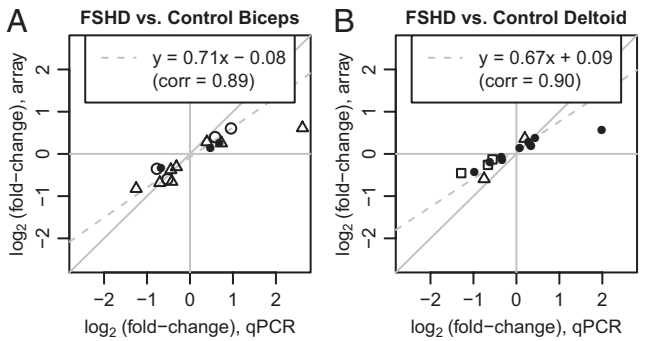

Fig. 2. Correlation between $\log_2$ fold-change values of 15 validation genes from microarrays and qPCR for FSHD subjects vs. controls comparisons in biceps (*A*) and deltoid muscles (*B*). The horizontal axis represents the $\log_2$ fold change between FSHD and control samples from qPCR and the vertical axis represents the $\log_2$ fold change from microarrays. Symbols in each scatter plot represent individual genes, with the shape of a symbol indicating the statistical significance of the difference between FSHD and control samples for that gene (square, $P < 0.01$ in qPCR; triangle, $P < 0.01$ in array; open circle, $P < 0.01$ in both; black dot, $P < 0.01$ in neither). The solid gray line with slope 1 indicates perfect agreement, and the dashed gray line indicates the line of best fit to the data (with slope, intercept, and correlation coefficient as indicated in the plots).

Our study, by design, examined mildly affected muscle, in an attempt to reveal FSHD-specific expression changes while minimizing expression changes related to infiltration by fibrotic or adipose tissues, which are common in more severely affected muscle. Thus, the discrepancy in the number of significantly altered genes between our and earlier studies might be the result of disease severity of profiled muscles. Moreover, by controlling for familial relationship when comparing expression levels in cases and controls, we have attempted to diminish the effects of interindividual variation on gene expression, which has already been proven by numerous studies to be under the control of genetic factors (33, 34). Therefore, the expression differences observed between FSHD and control muscles would more likely reflect true pathogenic gene expression profiles suitable for developing into disease biomarkers. Notably, these candidate biomarkers are not, for the most part, genes that were identified as misregulated upon overexpression of DUX4 in cell cultures. We do, however, see evidence of misregulation of some of the DUX4 target genes, particularly a significant up-regulation of DUX4 targets involved in germ cells and early development (12).

Among the most significantly altered genes in biceps, *IDI2* (isopentenyl-diphosphate delta isomerase 2) encodes a skeletal muscle-specific isozyme of isopentenyl diphosphate isomerase located in peroxisomes and involved in the mevalonate pathway, which is essential for cholesterol biosynthesis (35, 36). Its expression is reduced in affected biceps by 1.6-fold ($P = 0.000001$), whereas its expression was not significantly altered in deltoid. Interestingly, copy-number variation within the chromosomal locus harboring the *IDI2* gene and its homolog *IDI1* is associated with amyotrophic lateral sclerosis (37), a fatal neurodegenerative motor neuron disorder accompanied by severe skeletal muscle wasting. Note that *IDI2* is not reported to be altered in earlier studies of other muscular dystrophies (18–21) or in a search of all GEO profiles (38) related to muscular disorders. One other gene involved in cholesterol biosynthesis (Gene Ontology process GO:000695), hydroxymethylglutaryl CoA synthase 2 (*HMGCS2*) (39), showed strongly altered expression in FSHD biceps (1.8-fold up; $P = 0.007$), as well as deltoid (2.4-fold up; $P = 0.0002$). Interestingly, several other genes involved in cholesterol biosynthesis were found by Cheli et al. (40) to be deregulated in cell cultures for FSHD and also for FSHD-2, a form of the disease that is not associated with D4Z4 contractions in 4q.

Actin cytoskeleton plays an important role in cell motility, axon guidance, cytokinesis, and phagocytosis. Genes encoding members of the actin cytoskeleton were enriched among differentially regulated genes in biceps. In particular, the gene encoding the α2 isoform of actin (*ACTA2*) is up-regulated by 1.5-fold ($P = 0.0001$) in biceps muscles. Osborne et al. (17) also reported 2.5-fold increased expression of *ACTA2* in the vastus lateralis muscle of FSHD patients. Given its consistent overexpression in two independent analyses of FSHD muscles, the *ACTA2* gene could serve as a potential biomarker of FSHD muscles and perhaps to assess the degree of disease severity, because its expression levels in deltoid muscles were not altered to the same extent (1.2-fold; $P = 0.1$). Mutations in the actin-binding domain of the perinatal myosin heavy chain gene *MYH8* are associated with trismus–pseudocamptodactyly syndrome, characterized by abnormally short muscle–tendon units (41). This gene was found to be significantly overexpressed in FSHD muscles, as determined by both arrays and qPCR. Although *MYH8* is not normally expressed in healthy muscles beyond the perinatal period, where it is expressed primarily in the skeletal muscles of the limbs and face, it is reexpressed in dystrophic and regenerating muscles (42).

The complement system, which is abnormally regulated in FSHD deltoid in this study, is a component of the innate immune system that assists the acquired immune system in fighting infectious diseases or responding to tissue damage. Activation of the complement system has been observed in muscles affected with various types of muscular dystrophies and myopathies (43), including in nonnecrotic muscle fibers in patients with FSHD (44). The membrane attack complex generated by the members of the complement system recruits macrophages and neutrophils to the damaged myofibers leading to tissue necrosis. Microarray analysis of affected muscles in *mdx* mice, the animal model for Duchenne muscular dystrophy, showed up-regulation of genes of the complement system (45). Three (*C1QB*, *C3AR1*, *CFH*) out of four genes that were changed in deltoid were also up-regulated in biceps. Therefore, up-regulation of complement system genes in dystrophic muscles appears to be a common feature of muscular dystrophies, including FSHD.

Overexpression of *DUX4* induces caspase 3/7 activity, consequently leading to cell death in culture (9). Wallace et al. (25) recently demonstrated using transgenic zebrafish and mouse models overexpressing the human full-length *DUX4* transcript, that p53 pathway–dependent apoptosis genes were overexpressed in skeletal muscles of these organisms. We examined the expression levels of several p53 pathway–dependent apoptosis genes [*CASP3* (caspase 3, apoptosis-related cysteine peptidase), *BIRC5* (baculoviral IAP repeat containing 5), *BAX* (BCL2-associated X protein), *CASP1* (caspase 1, apoptosis-related cysteine peptidase), *APAF1* (apoptotic peptidase activating factor 1), *TRP63* (transformation-related protein 63), *BID* (BH3 interacting domain death agonist), *CASP9* (caspase 9, apoptosis-related cysteine peptidase), *BAK1* (BCL2-antagonist/killer 1), *TRP53* (transformation-related protein 53), *BAD* (BCL2-associated agonist of cell death), and *CASP7* (caspase 7, apoptosis-related cysteine peptidase)] highlighted in the study by Wallace et al. as the most significantly up-regulated genes but did not observe significant differences between FSHD and control muscle biopsies. Although we do not rule out the role of these genes in FSHD pathogenesis, the expression levels of these genes changed with forced overexpression of *DUX4*, whereas the endogenous expression level of *DUX4* in FSHD myoblasts appears to be very low (29).

In addition to identifying genes differentially expressed in diseased versus healthy muscles, we also identified a large number of genes differentially expressed between two different muscles types, biceps and deltoid, both dystrophic and healthy. These genes and pathways could be important targets to investigate the developmental mechanisms of distinct muscle types, as well as disease progression in FSHD patients.

The set of 15 genes identified in this study as differentially expressed in FSHD muscles and validated in an independent sample set can be considered a molecular signature of the disease. Development of such an FSHD biomarker will be helpful in evaluating efficacy of potential therapeutics for treating the disease in the future.

Materials and Methods

Sample Collection. Biopsy samples were collected from biceps and deltoid muscles from affected and unaffected family members. All biopsies were from muscles with at least 4 out of 5 strength on the MRC scale. Histological evaluation of biopsy samples demonstrated that all were primarily composed of myofibers with minimal fibrosis or fatty infiltration. Microarray and the initial qPCR validation data were obtained from 12 FSHD subjects and 13 first-degree unaffected relatives (Table S1, highlighted in bold). Altogether, 50 total RNA samples were hybridized to arrays. As an independent replication sample set, we evaluated 25 individuals from 10 additional cohorts, and 1 additional member of cohort 15, using qPCR analysis. DNA-based molecular diagnosis was confirmed by the University of Iowa Diagnostic Laboratories. A portion of each biopsy sample (ranging between 50 to 200 mg) was snap-frozen in liquid nitrogen immediately after procurement and stored at −80 °C for RNA extraction. All participants provided written informed consent, and this study was approved by The Johns Hopkins Medicine Institutional Review Board.

RNA Isolation and First-Strand cDNA Synthesis. Total RNA was isolated from frozen muscle tissues using TRIzol reagent (Invitrogen). More details are provided in *SI Materials and Methods*.

Microarray Analysis. Gene expression profiling was carried out using the Affymetrix GeneChip Human Gene 1.0 ST arrays. Microarray data were collected at Expression Analysis (www.expressionanalysis.com). Further details on array processing are provided in *SI Materials and Methods*.

Microarray Data Analysis. The raw array data were preprocessed and normalized using the Robust Multichip Average (RMA) method (46). Differential expression between classes was calculated using linear models with the limma package (47). To adjust for multiple hypothesis testing we applied FDR (48). The microarray CEL files and normalized probeset-intensity values were deposited into the GEO database (www.ncbi.nlm.nih.gov/geo) and are available under accession no. GSE36398. More details on data analysis are provided in *SI Materials and Methods*.

Real-Time Quantitative RT-PCR. High-throughput real-time qPCR was performed on the BioMark 96.96 Dynamic Array (Fluidigm) with TaqMan Gene Expression Assays (Applied Biosystems). The linear combinations of Ct values to use for sample classification were selected using logistic regression with L_1 regularization. Details are provided in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank all of the families who participated in this study; Daniel Perez and the FSH Society for assistance with patient recruitment and support of the Boston Biomedical Research Institute (BBRI) Senator Paul D. Wellstone Muscular Dystrophy Cooperative Research Center (MDCRC) for FSHD; Leigh Warsing Johnson, Naili Liu, and Jennifer Chen for assistance with sample handling; and Patrick Reed, Peter Jones, and Jeff Miller for helpful suggestions and comments on the manuscript. This work was supported by the Eunice Kennedy Shriver National Institute Of Child Health & Human Development of the National Institutes of Health (NIH) Grant 5U54HD060848 to the BBRI Senator Paul D. Wellstone MDCRC for FSHD. The Molecular Genetics Core Facility at Boston Children's Hospital, where we performed qPCR experiments on the BioMark instrument, is supported by NIH Grants P50-NS-40828 and P30-HD-18655. F.R. is supported by Muscular Dystrophy Association Grant MDA202863.

1. Tawil R, Van Der Maarel SM (2006) Facioscapulohumeral muscular dystrophy. *Muscle Nerve* 34:1–15.
2. Padberg GW, et al. (1995) On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2:S73–S80.
3. Fitzsimons RB (2011) Retinal vascular disease and the pathogenesis of facioscapulohumeral muscular dystrophy. A signalling message from Wnt? *Neuromuscul Disord* 21:263–271.
4. Wijmenga C, et al. (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 2:26–30.
5. van Deutekom JC, et al. (1993) FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum Mol Genet* 2: 2037–2042.
6. Mostacciuolo ML, et al. (2009) Facioscapulohumeral muscular dystrophy: Epidemiological and molecular study in a north-east Italian population sample. *Clin Genet* 75: 550–555.
7. Tawil R, et al.; The FSH-DY Group (1996) Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular dystrophy. *Ann Neurol* 39:744–748.
8. Lunt PW, et al. (1995) Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). *Hum Mol Genet* 4:951–958.
9. Kowaljow V, et al. (2007) The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord* 17:611–623.
10. Bosnakovski D, et al. (2008) An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J* 27:2766–2779.
11. Lemmers RJ, et al. (2010) A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329:1650–1653.
12. Geng LN, et al. (2012) DUX4 activates germline genes, retroelements, and immune mediators: Implications for facioscapulohumeral dystrophy. *Dev Cell* 22:38–51.
13. Ilyin SE, Belkowski SM, Plata-Salamán CR (2004) Biomarker discovery and validation: Technologies and integrative approaches. *Trends Biotechnol* 22:411–416.
14. 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.
15. Orrell RW, et al. (1999) Definitive molecular diagnosis of facioscapulohumeral dystrophy. *Neurology* 52:1822–1826.
16. Kang PB, et al. (2005) Variations in gene expression among different types of human skeletal muscle. *Muscle Nerve* 32:483–491.
17. 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.
18. Chen YW, Zhao P, Borup R, Hoffman EP (2000) Expression profiling in the muscular dystrophies: Identification of novel aspects of molecular pathophysiology. *J Cell Biol* 151:1321–1336.
19. Pescatori M, et al. (2007) Gene expression profiling in the early phases of DMD: A constant molecular signature characterizes DMD muscle from early postnatal life throughout disease progression. *FASEB J* 21:1210–1226.
20. Greenberg SA, et al. (2002) Molecular profiles of inflammatory myopathies. *Neurology* 59:1170–1182.
21. Sáenz A, et al. (2008) Gene expression profiling in limb-girdle muscular dystrophy 2A. *PLoS ONE* 3:e3750.
22. 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.
23. 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.
24. 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.
25. Wallace LM, et al. (2011) DUX4, a candidate gene for facioscapulohumeral muscular dystrophy, causes p53-dependent myopathy in vivo. *Ann Neurol* 69:540–552.
26. Wu D, et al. (2010) ROAST: Rotation gene set tests for complex microarray experiments. *Bioinformatics* 26:2176–2182.
27. 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.
28. Geng LN, et al. (2011) DUX4 activates germline genes, retroelements, and immune mediators: Implications for facioscapulohumeral dystrophy. *Dev Cell* 22:38–51.
29. Snider L, et al. (2010) Facioscapulohumeral dystrophy: Incomplete suppression of a retrotransposed gene. *PLoS Genet* 6:e1001181.
30. Snider L, et al. (2009) RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: New candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum Mol Genet* 18:2414–2430.
31. Jones TI, et al. (2012) Facioscapulohumeral muscular dystrophy family studies of DUX4 expression: Evidence for disease modifiers and a quantitative model of pathogenesis. *Hum Mol Genet*, 10.1093/hmg/dds284.
32. Celegato B, et al. (2006) Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre differentiation and a general deregulation of MyoD-dependent genes. *Proteomics* 6:5303–5321.
33. Morley M, et al. (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature* 430:743–747.
34. Cheung VG, Spielman RS (2009) Genetics of human gene expression: Mapping DNA variants that influence gene expression. *Nat Rev Genet* 10:595–604.
35. Clizbe DB, Owens ML, Masuda KR, Shackelford JE, Krisans SK (2007) IDI2, a second isopentenyl diphosphate isomerase in mammals. *J Biol Chem* 282:6668–6676.
36. Ramos-Valdivia AC, van der Heijden R, Verpoorte R (1997) Isopentenyl diphosphate isomerase: A core enzyme in isoprenoid biosynthesis. A review of its biochemistry and function. *Nat Prod Rep* 14:591–603.
37. Kato T, et al. (2010) Segmental copy-number gain within the region of isopentenyl diphosphate isomerase genes in sporadic amyotrophic lateral sclerosis. *Biochem Biophys Res Commun* 402:438–442.
38. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210.
39. Vilà-Brau A, De Sousa-Coelho AL, Mayordomo C, Haro D, Marrero PF (2011) Human HMGCS2 regulates mitochondrial fatty acid oxidation and FGF21 expression in HepG2 cell line. *J Biol Chem* 286:20423–20430.
40. Cheli S, et al. (2011) Expression profiling of FSHD-1 and FSHD-2 cells during myogenic differentiation evidences common and distinctive gene dysregulation patterns. *PLoS ONE* 6:e20966.
41. Toydemir RM, et al. (2006) Trismus-pseudocamptodactyly syndrome is caused by recurrent mutation of MYH8. *Am J Med Genet A* 140:2387–2393.
42. Bandman E (1985) Continued expression of neonatal myosin heavy chain in adult dystrophic skeletal muscle. *Science* 227:780–782.
43. Engel AG, Biesecker G (1982) Complement activation in muscle fiber necrosis: Demonstration of the membrane attack complex of complement in necrotic fibers. *Ann Neurol* 12:289–296.
44. Spuler S, Engel AG (1998) Unexpected sarcolemmal complement membrane attack complex deposits on nonnecrotic muscle fibers in muscular dystrophies. *Neurology* 50: 41–46.
45. Porter JD, et al. (2002) A chronic inflammatory response dominates the skeletal muscle molecular signature in dystrophin-deficient mdx mice. *Hum Mol Genet* 11: 263–272.
46. Irizarry RA, et al. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31:e15.
47. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:Article3.
48. Reiner A, Yekutieli D, Benjamini Y (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19:368–375.