

## ARTICLE

# Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei

Peter S Masny<sup>1,2,8</sup>, On Ying A Chan<sup>3,4,8</sup>, Jessica C de Greef<sup>3</sup>, Ulla Bengtsson<sup>1</sup>, Melanie Ehrlich<sup>5</sup>, Rabi Tawil<sup>6</sup>, Leslie F Lock<sup>1</sup>, Jane E Hewitt<sup>7</sup>, Jennifer Stocksdales<sup>1</sup>, Jorge H Martin<sup>1</sup>, Silvere M van der Maarel<sup>3</sup> and Sara T Winokur<sup>\*,1</sup>

Autosomal dominant facioscapulohumeral muscular dystrophy (FSHD) is likely caused by epigenetic alterations in chromatin involving contraction of the D4Z4 repeat array near the telomere of chromosome 4q. The precise mechanism by which deletions of D4Z4 influence gene expression in FSHD is not yet resolved. Regulatory models include a *cis* effect on proximal gene transcription (position effect), DNA looping, non-coding RNA, nuclear localization and *trans*-effects. To directly test whether deletions of D4Z4 affect gene expression in *cis*, nascent RNA was examined in single myonuclei so that transcription from each allele could be measured independently. FSHD and control myotubes (differentiated myoblasts) were subjected to sequential RNA-DNA FISH. A total of 16 genes in the FSHD region (*FRG2*, *TUBB4Q*, *FRG1*, *FAT1*, *F11*, *KLKB1*, *CYP4V2*, *TLR3*, *SORBS2*, *PDLIM3* (*ALP*), *LRP2BP*, *ING2*, *SNX25*, *SLC25A4* (*ANT1*), *HELT* and *IRF2*) were examined for interallelic variation in RNA expression within individual myonuclei. Sequential DNA hybridization with a unique 4q35 chromosome probe was then applied to confirm the localization of nascent RNA to 4q. A D4Z4 probe, labeled with a third fluorochrome, distinguished between the deleted and normal allele in FSHD nuclei. Our data do not support an FSHD model in which contracted D4Z4 arrays induce altered transcription in *cis* from 4q35 genes, even for those genes (*FRG1*, *FRG2* and *SLC25A4* (*ANT1*)) for which such an effect has been proposed.

*European Journal of Human Genetics* (2010) 18, 448–456; doi:10.1038/ejhg.2009.183; published online 4 November 2009

**Keywords:** facioscapulohumeral muscular dystrophy; FSHD; allelic expression; RNA FISH; FRG1; ANT1

## INTRODUCTION

Facioscapulohumeral muscular dystrophy (FSHD) is an autosomal dominant neuromuscular disorder.<sup>1,2</sup> With a frequency of 1 in 20 000, FSHD is the third most common hereditary myopathy.<sup>3</sup> Clinically, FSHD manifests as facial muscle weakness during adolescence with progression to the upper and then lower extremities later in adulthood. The primary tissue affected is the skeletal muscle, although other symptoms may include sensorineural hearing loss, retinal telangiectasias, epilepsy and, in the most severely affected cases, mental retardation.<sup>4,5</sup> The disease generally presents in the second decade of life. Disease onset, progression and symptoms present with considerable variability between patients even within the same family, ranging from mild muscle disease to wheelchair dependence.

A vast majority of patients with FSHD (97%) present with a contraction of the macrosatellite repeat D4Z4 on the subtelomere of chromosome 4q (4q35).<sup>6,7</sup> D4Z4 is a polymorphic tandem repeat array consisting of 3.3 kb units. In the general population, the D4Z4 array spans 11 to 150 units, whereas the repeat array is contracted to only 1 to 10 elements on one allele in patients with FSHD. As a result of an ancient duplication, the D4Z4 repeat array is also present on the subtelomere of chromosome 10.<sup>8</sup> However, contractions of the D4Z4 repeat on 10q26 are not associated with

FSHD despite extensive homology proximal and distal to the array on 4q35 and 10q26.

The presence of a contracted D4Z4 repeat array on 4qter alone is not sufficient to cause FSHD. Distal to the D4Z4 repeat, a genetic variation has been identified and is distinguished as alleles 4qA and 4qB. Although the two alleles are equally distributed in the general population, all reported patients with FSHD carry the 4qA variant in parallel to the contracted D4Z4 repeat.<sup>9</sup> The predominant difference between the two allelic variants is the presence of a 6.2 kb  $\beta$ -satellite repeat on the 4qA allele. Furthermore, only certain haplotypes of 4qA are associated with the disease, indicating that specific sequence polymorphisms contribute to the disease mechanism.<sup>10</sup> Just how these polymorphic variations are involved in FSHD pathogenesis is unclear.

Over the past two decades, several models have been proposed to explain the disease mechanism underlying FSHD. These modes of gene regulation include a *cis* effect on proximal gene transcription (position effect),<sup>11,12</sup> DNA looping,<sup>13</sup> non-coding RNA,<sup>14</sup> nuclear localization<sup>15,16</sup> and *trans*-effect.<sup>17</sup> Perhaps the most generally accepted model is that of a *cis* effect in which the intact D4Z4 repeat array is involved in controlling the expression of genes located proximally on 4q35, and possibly at large distances from the array. Such position effects are known to influence human disease

<sup>1</sup>Department of Biological Chemistry, University of California, Irvine, CA, USA; <sup>2</sup>Department of Neurology, University of Colorado, Denver, CO, USA; <sup>3</sup>Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; <sup>4</sup>Department of Physiological Chemistry, University Medical Center Utrecht, Utrecht, The Netherlands; <sup>5</sup>Program in Human Genetics, Tulane University, New Orleans, LA, USA; <sup>6</sup>Department of Neurology, Fields Center for FSHD and Neuromuscular Research, University of Rochester, Rochester, NY, USA; <sup>7</sup>Department of Mammalian Genetics, University of Nottingham, Nottingham, UK

<sup>8</sup>These authors contributed equally to this work.

\*Correspondence: Dr ST Winokur, Department of Biological Chemistry, 202 Sprague Hall, University of California, Irvine, CA 92697, USA.

Tel: +1 949 824 2911; Fax: +1 949 824 9547; E-mail: stwinoku@uci.edu

Received 6 January 2009; revised 10 July 2009; accepted 11 August 2009; published online 4 November 2009

genes at distances > 1 Mb.<sup>18</sup> Contraction of D4Z4 array to < 11 units is thought to cause an inappropriate expression of 4q35 genes, resulting in FSHD pathology. Upregulation of several genes: *SLC25A4* (*ANT1*), *FRG1* and *FRG2*, was detected in one study of FSHD muscle tissue using chromosome-4-specific reverse transcriptase PCR.<sup>19</sup> However, this upregulation has not been confirmed by several other studies using techniques varying from microarray analysis to quantitative real-time PCR.<sup>20–24</sup> In addition, chromatin studies have also not supported a cis-spreading effect.<sup>13,20</sup>

This study aims to resolve the controversy as to whether 4q35 genes are transcriptionally upregulated by a *cis* effect in FSHD. To this end, we investigated nascent RNA transcription in single myonuclei using quantitative fluorescent hybridization.<sup>25–27</sup> Earlier expression studies have assayed pooled mRNA levels without unambiguously distinguishing transcription from the contracted versus the unaffected allele. Our approach uses sequential RNA–DNA FISH for quantitation of variations in *interallelic* RNA transcription within individual nuclei. Thus, any potential *cis* effect can be directly assayed within individual nuclei. In addition, by examining pre-mRNA (nascent) transcription, variation in RNA quantities attributed to posttranscriptional processing of transcripts (that is, unrelated to nascent *cis* effects) is avoided. Our approach has the further advantage in that, by examining individual nuclei, the chromosome-4-specific expression of multicopy genes, such as *FRG1* and *FRG2*, can be unambiguously determined.

In this study, transcription was measured specifically in differentiated myotubes derived from normal and mildly affected muscle cells (quadriceps and deltoid). As FSHD is a progressive neuromuscular disorder, cultures were derived from muscle before the onset of severe dystrophy, so that transcriptional activity leading to the muscle phenotype could be observed as a primary effect and not secondary to the dystrophic process. Previous studies have used pooled mRNA isolated from muscle tissue containing a variety of cell types or cultured cells at different stages of differentiation, leading to variability in expression patterns that might not be correlated with the disease. By restricting our analysis to one myogenic cell type at a specific stage of differentiation, such variation in transcription unrelated to disease was minimized. Here, we present an overview of nascent RNA transcription from 16 genes in the FSHD region of chromosome 4q35.

## MATERIALS AND METHODS

### Cell lines and tissue culture

The FSHD primary myoblast cell lines, FM41j and FM1243j, and the control myoblast cell lines, Cl-02 and CM020399, were propagated on gelatin-coated flasks in F10 medium supplemented with 20% FBS, 100 U/ml penicillin/streptomycin, 10 ng/ml bFGF (Invitrogen, Carlsbad, CA, USA) and 1  $\mu$ M dexamethasone (Sigma-Aldrich, St Louis, MO, USA). Myoblast cultures were derived from quadriceps or deltoid biopsies; FSHD samples were from mildly affected muscle based on histopathology. To allow for myotube differentiation, cells were grown on laminin-coated coverslips in DMEM medium supplemented with 2% horse serum, 2 mM L-glutamine and 100 U/ml penicillin/streptomycin (Invitrogen) for 5 days. Before fixation in 4% paraformaldehyde, cells were treated with a cytoskeleton extraction buffer containing 0.5% Triton X-100 and 10 mM VRC (vanadyl ribonucleoside complex).<sup>26</sup> Subsequently, coverslips were stored in 70% ethanol at 4°C. Myoblast lineage was verified by standard immunostaining with the polyclonal desmin antibody Y-20 (1:200) and a rhodamine-conjugated secondary antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA). Differentiation was confirmed by the appearance of fused, multinucleated myotubes and the expression of myosin heavy chain by immunofluorescence (Sigma MY-32 anti-MHC; data not shown).

### Probes

A total of 16 genes in the FSHD region were analyzed using cosmid probes. The relative map location of each of these genes with respect to the D4Z4 repeat array is shown in Table 1. Cosmids were isolated in-house from the Los Alamos human chromosome 4 library.<sup>28</sup> Isolated cosmids were verified by *Eco*RI digestion and sequencing of insert ends with T3 and T7 primers. Cosmids were labeled with Cy3-dCTP (Amersham, Louisville, CO, USA) using a nick translation mix (Roche, Palo Alto, CA, USA), with the exception of the chromosome 4-specific c88F8 (D4S139) localization cosmid, which was labeled with fluorescein-12-dUTP (Roche). The *SNRPN* probe was labeled with red fluorochrome in the Texas Red spectrum (Rainbow Scientific Inc, Windsor, CT, USA). A Bluescript KSII+ plasmid containing a single unit of the 3.3 kb D4Z4 repeat was labeled with Cy5-dCTP. Labeled probes were subsequently filtered through Microspin G-50 columns (Amersham) and ethanol precipitated. The localization of all probes to chromosome 4q35 was verified by hybridization to metaphase cells.

### Fluorescence *in situ* hybridization

Cosmid DNA together with 100-fold human Cot1 DNA (Invitrogen) was lyophilized and resuspended in 50% formamide, 20% dextran sulfate, 4 $\times$  SSC,

**Table 1** FSHD region genes analyzed by RNA–DNA FISH

Gene	Alternate	Name	Kb to D4Z4	Cosmid	Copy number
<i>FRG2</i>		FSHD region gene 2	55	—	Multi
<i>TUBB4Q</i>		Tubulin 4q pseudogene	97	—	Multi
<b><i>FRG1</i></b> *		FSHD region gene 1	141	—	Multi
<b><i>FAT1</i></b> *		FAT tumor suppressor homolog	3508	67F9	Single
<i>F11</i>		Coagulation factor XI	3816	10C7	Single
<i>KLKB1</i>		Kallikrein B	3855	10C7	Single
<b><i>CYP4V2</i></b>		Cytochrome P450, family 4	3890	190B1	Single
<b><i>TLR3</i></b>		Toll-like receptor 3	4012	249D4	Single
<b><i>SORBS2</i></b>	<i>ARGBP2</i>	Sorbin and SH3 domain containing 2	4540	144B4	Single
<b><i>PDLIM3</i></b> *	<i>ALP</i>	PDZ and LIM domain 3	4581	49D9	Single
<b><i>LRP2BP</i></b> *		LRP2-binding protein	4717	79A10	Single
<b><i>SNX25</i></b>		Sorting nexin 25	4888	164D7	Single
<b><i>SLC25A4</i></b> *	<i>ANT1</i>	Solute carrier family 25	4936	168D11	Single
<b><i>HELT</i></b>		Hey-like transcriptional repressor	5060	234E12	Single
<b><i>IRF2</i></b> *		Interferon regulatory factor 2	5702	66B6	Single
<b><i>ING2</i></b>		Inhibitor of growth family, member 2	6574	191A10	Single

—, Null or low expression.

Gene not shown in bold exhibit null or expression levels too low to be analyzed by RNA–FISH.

Genes names given in bold represent those that are analyzed by two-color RNA–DNA FISH only.

Genes names given in bold along with asterisk symbol represent those that are analyzed by two- and three-color RNA–DNA FISH.

0.1 M DTT, 0.5× Denhardt's solution, and ssDNA, tRNA and PolyA (250 µg/ml each). The Cy3-labeled RNA-FISH probes were denatured for 10 min at 80°C and immediately put on ice. Non-denatured coverslips were dehydrated and hybridized overnight with the RNA-FISH probes in a humidified environment at 37°C. Coverslips were washed in 50% formamide/2× SSC and 2× SSC at 45°C with a final wash in 1× SSC at room temperature. Cells were then fixed in 4% paraformaldehyde and treated with 0.15M NaOH/70% ethanol (alkaline hydrolysis) for 10 min each at room temperature before dehydration with increasing concentrations of ethanol. As negative controls, coverslips were either treated with 0.1 mg/ml RNase A (Applied Biosystems, Foster City, CA) at 37°C for 1 h before RNA hybridization or 0.5 U/µl RNase H (Epicentre Biotechnologies, Madison, WI, USA) at 37°C for 1 h after hybridization, but before paraformaldehyde fixation. For chromosomal localization and allele identification, the fluorescein isothiocyanate (FITC)-labeled c88F8 and Cy5-labeled D4Z4 DNA-FISH probes were prepared identically to the RNA-FISH probes. However, for nuclei hybridized with c88F8 and D4Z4 DNA-FISH probes, the chromosomal DNA and probes were simultaneously denatured by placing the slide on a 70°C hot plate for 2 min. After overnight hybridization at 37°C, the coverslips were washed as described above and counterstained with DAPI.

### Fluorescence microscopy and deconvolution

Fluorescence signals were captured with a Photometrics Coolsnap HQ CCD camera (Photometrics, Tucson, AZ, USA) using an inverted widefield microscope (Nikon Eclipse TE2000-U, Nikon, Melville, NY, USA) with a ×63 Nikon Plan-Apo objective (N.A. 1.4) and dual motorized 10-position filterwheels with separate excitation and emission filters. The focal plane of the microscope passed through the whole nucleus with z-steps of 0.2 µm creating individual image series for each excitation filter. For each hybridization, at least 60 nuclei were captured. Images were deconvoluted with the program Autoquant version 9.3 (Media Cybernetics, Bethesda, MD, USA) and combined into single RGB composite images using the program ImageJ (NIH, Bethesda, MD, USA) as previously described.<sup>15</sup>

### Quantification and scoring

RNA-DNA RGB composite images were presented to an operator who used the FITC-labeled c88F8 DNA signal to identify and mark the location of each 4q allele. Occasional nuclei were excluded because the two 4q loci were in such close proximity that distinct allelic RNA signals could not be measured, or rarely because two 4q signals were not clearly visualized. Census, an ImageJ plugin, and an automated algorithm were used to quantify the Cy3 fluorescence signal adjacent to the DNA signal as a measure of pre-mRNA transcription. The algorithm opened the corresponding deconvoluted 3D image stacks, using them to quantify the signal intensities by measuring the total probe intensity (TPI) within a fixed 500 nm diameter sphere. The sum of the intensity of the voxels within the sphere was calculated and normalized by subtracting the median intensity of signals within a cylindrical shell around the signal. The algorithm assessed the entire image region near the operator's mark and located the coordinates that provided the brightest signal within the sphere. Those coordinates were recorded as the signal center and the corresponding TPI was used for data analysis. For nuclei studied using a Cy5-labeled D4Z4 probe, the Cy5 image was included in the RGB composites as the blue signal. Using these RGB composites, the operator visually compared the signals on each allele and designated one of the telomeres as 'affected' and the other as 'normal' based on hybridization intensity with the D4Z4 probe. In over 95% of nuclei with measured RNA and DNA signals, one of the two 4q D4Z4 signals was absent or very faint, identifying the affected (deleted) allele. The rare nuclei with signals similar in intensity, or those confounded by the near presence of another signal from a D4Z4 repeat elsewhere in the genome, were excluded from the analysis. It is noted that for *SNRPN* analysis, only one allele would produce a visible nascent RNA signal in virtually all nuclei (as expected). For these, the brightest background spot near the non-transcribing 15q allele was used for fold comparison.

### Statistical analysis

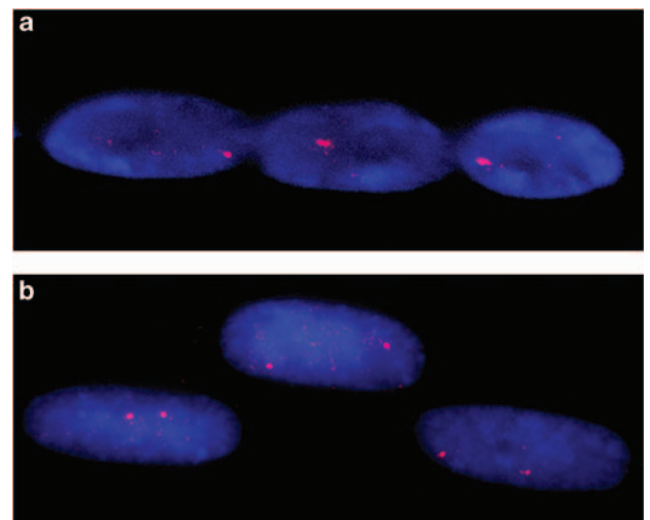
All data were analyzed using the R Project version 2.5.1 (<http://www.r-project.org>). When comparing allele ratios between normal and FSHD cell

lines in the two-color experiments, the Mann-Whitney *U*-test analysis was applied. When comparing allele signals within FSHD nuclei in the three-color experiments, a single sample Wilcoxon signed-rank test was used to evaluate for symmetry of expression between the alleles. Interallelic ratios for the positive control gene, *SNRPN*, were analyzed by comparing the fold distribution of *SNRPN* FSHD nuclei with the FSHD fold distribution for all the other genes studied combined in one pool (Supplementary Figure 2), and significance was established by using the Mann-Whitney *U*-test. It is noted that *FRG1* and *IRF2* were only studied using three-color RNA-DNA FISH. However, control cell lines were also analyzed for the sake of completeness. Note, however, that the same FSHD data sets were used for the two- and three-color significance testing shown.

## RESULTS

### Validation of methodology

The sensitivity and specificity of nascent RNA transcript detection was determined using several approaches. First, as a positive control, we detected monoallelic RNA transcription from *SNRPN*, an imprinted gene from the Prader-Willi/Angelman locus.<sup>29</sup> The *SNRPN* probe was hybridized independently to coverslips either after denaturation (DNA-FISH conditions) or without denaturation (RNA-FISH conditions). As expected, the *SNRPN* probe detected allele-specific transcription, with a single 15q allele signal under RNA hybridization conditions and both 15q alleles under DNA hybridization conditions (Figure 1). The percent myonuclei displaying monoallelic expression of *SNRPN* (83%) is similar to that reported by Kohda *et al*<sup>29</sup> for HL60 and WI38 cells, with only 4% displaying biallelic expression and 13% null expression. A second positive control used to test for sensitivity was the correlation of D4Z4 FISH signal intensity with the size of the D4Z4 target array. Extrapolation of DNA-FISH results to RNA-FISH results can be made here to address the correlation between fold change of nucleic acid quantity and that of signal intensity, as both DNA-FISH and RNA-FISH were performed within the same nuclei, fixed with identical conditions and washed using the same stringent conditions. From our previous FISH studies, we have shown that we can discriminate between the normal and disease allele on the basis of signal intensity of the D4Z4 repeat.<sup>15</sup> The two chromosome 4 in the



**Figure 1** RNA- and DNA-FISH of the imprinted gene *SNRPN* in myonuclei. (a) Monoallelic expression of RNA from one 15q allele is detected when non-denatured nuclei are probed with Cy3-labeled *SNRPN*, corresponding to the actively transcribed gene. (b) Two loci are detected when denatured nuclei are probed with Cy3-labeled *SNRPN*, corresponding to the DNA at both chromosome 15 alleles.

cell line FM1243j have 4 D4Z4 repeats (2 complete, 1 degenerate and 1 inverted) on the affected (contracted) allele and 23 D4Z4 repeats (21 complete, 1 degenerate and 1 inverted) on the normal allele. D4Z4 probe hybridization to the 2 alleles in 46 nuclei had a signal intensity ratio with a mean of 6.2, median of 6.4 and median confidence interval of 5.9–6.8. As the ratio of nucleic acid target is 5.75 (23/4 D4Z4 repeats), this correlates extremely well with the ratio of signal intensity (6.2). As a negative control, FITC-labeled 4q35 probe c88F8 (which does not contain coding sequences) was mixed with the Cy3-labeled *HELT* probe (a 4q35 gene), and hybridized independently to denatured and non-denatured nuclei. Two green and two red signals were detected on denatured nuclei (DNA conditions), corresponding to the FITC-labeled c88F8 DNA and the Cy3-labeled *HELT* DNA and RNA (single-stranded DNA and RNA), whereas only red signals were detected on non-denatured nuclei, corresponding to single-stranded *HELT* RNA only (Figure 2a, b). A second negative control was provided by the lack of RNA transcript detection at the 4q locus when cells were pretreated with RNase A before hybridization or when treated post-hybridization with RNase H (Figure 2c, d).

#### Several 4q35 genes are not expressed in differentiated myotubes

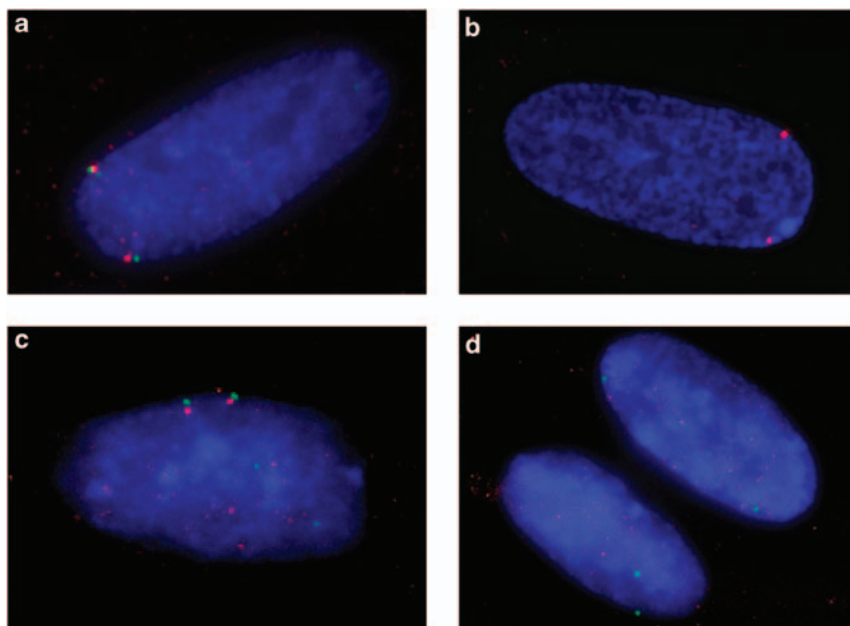
Of the 16 genes studied, 6 were expressed at too low a level to be detected or adequately quantified. These six are *FRG2*, *TUBB4Q*, *F11*, *KLKB1*, *SNX25* and *ING2*. For the multicopy loci *FRG2* and *TUBB4Q*, >85% of nuclei show no pre-mRNA signals at the chromosome 4 alleles in both control and FSHD nuclei, making statistical analysis impossible. This high number of 'null expression' nuclei suggests that these genes are not expressed (or expressed at a very low level) from chromosome 4 in FSHD and normal myotubes. For the genes *ING2* and *SNX25*, no RNA expression was detected in >55% of the captured nuclei, both for the control and FSHD myotubes. Statistical analysis over the remainder of values was deemed unreliable and

transcription of these genes in myotubes appears too low to measure. For *F11* and *KLKB1*, no RNA transcription was detected in any of the nuclei examined.

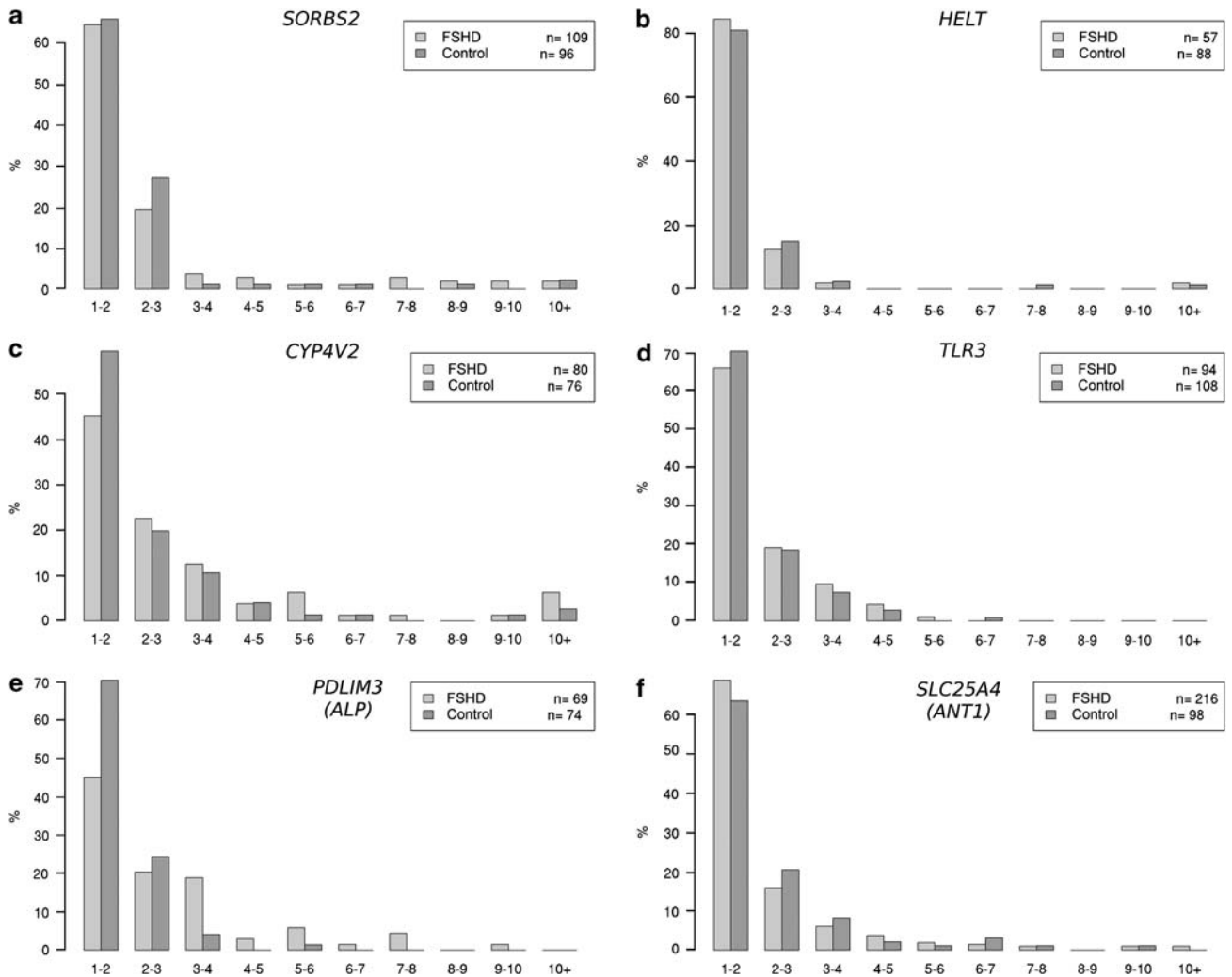
#### Relative allele ratios as determined by two-color RNA–DNA FISH

All genes were first analyzed using two-color RNA–DNA FISH to identify those genes with either a monoallelic or skewed biallelic ratio suggestive of an altered *cis* effect on expression. Figure 2c depicts typical nuclei of differentiated myoblasts in which the green signal corresponds to the 4q localization probe D4S139, and the red signal corresponds to the RNA (gene) of interest. For two-color FISH, pre-mRNA transcription signals at each 4q allele were compared with those from the other 4q allele for each individual nucleus. Intranuclear transcription ratios between alleles for each gene were quantified, and the distribution of ratio values was compared between normal and FSHD myonuclei. If a *cis* effect on transcription attributed to short D4Z4 array exists in FSHD myonuclei, then one would expect higher interallelic signal ratios in FSHD than in normal nuclei. Figure 3 displays results for two-color analysis of six genes present in 3.9–5.1 Mb proximal to D4Z4: *SORBS2*, *HELT*, *CYP4V2*, *TLR3*, *PDLIM3* and *ANT1*. Data for four other genes expressed in myotubes (*FRG1*, *FAT1*, *IRF2* and *LRP2BP*) are presented in Supplementary Figure 1.

For 9 out of the 10 expressed genes analyzed by two-color RNA–DNA FISH, the distribution of fold difference between interallelic signal intensities is not significantly different (Table 2). An FSHD-specific *cis* effect was not shown for these nine genes, as there was no conserved asymmetry of interallelic expression in either of the two FSHD samples or two control samples. One gene, *PDLIM3*, was suggestive of a possible *cis* effect (Figure 3 and Table 2). To study the expression of this gene further, as well as of several others implicated previously in FSHD,<sup>19</sup> we performed allele-specific (three-color) RNA–DNA FISH.



**Figure 2** Controls for RNA expression detection. (a and b) Influence of denaturation on RNA and DNA detection. (a) DNA signals from both *HELT* (Cy3, red fluorescence) and c88F8 FITC (green fluorescence) are detected on denatured nuclei. (b) RNA signals are detected only from the *HELT* locus on non-denatured nuclei, as c88F8 does not contain coding sequence. (c and d) Influence of RNase H on RNA and DNA detection. *SLC25A4* (red) and c88F8 (green) were labeled with Cy3 and FITC, respectively. (c) Nuclei not treated with RNase H show signals from both *ANT1* and c88F8 on sequential RNA hybridization. (d) Nuclei treated with RNase H after RNA hybridization reveal the loss of the signals from *ANT1*, confirming that these are RNA signals.



**Figure 3** Interallelic ratios of RNA signal intensity. Two-color RNA–DNA FISH analysis results for (a) *SORBS2*, (b) *HELT*, (c) *CYP4V2*, (d) *TLR3*, (e) *PDLIM3* (*ALP*) and (f) *SLC25A4* (*ANT1*). For each nucleus measured, the total probe intensity of the brighter RNA signal was divided by the intensity of the weaker signal. The fold difference in signal intensity between the two alleles is plotted on histograms (x axis) with the percent of nuclei displaying each ratio on the y axis. Typically, most nuclei have signals with similar intensities, with fold differences between one and three. Some nuclei have much brighter signals at one chromosome compared with the other, with the majority of transcription occurring only at one allele. Intranuclear (interallelic) ratios for control myonuclei (light gray) are plotted alongside FSHD myonuclei (dark gray). If a *cis* effect was modifying RNA transcription at the FSHD-affected allele, one would expect to see a greater number of nuclei with larger fold changes in the FSHD myonuclei compared with the control myonuclei. Instead, with the exception of *PDLIM3*, the histograms are nearly identical.

#### Allele-specific detection of RNA transcription using three-color RNA–DNA FISH

The more definitive three-color analysis on FSHD nuclei allows unequivocal identification of transcription from the contracted allele versus the normal allele within a given nucleus. A Cy5-labeled D4Z4 probe was used to identify the affected (contracted) allele in each FSHD myotube nucleus (as exemplified in Figure 4). In the analysis of a single nucleus, the TPI of the FSHD-affected allele (with a D4Z4 repeat deletion) was compared with the unaffected allele (with normal D4Z4 repeat count). Importantly, the strength of this methodology is that comparisons are made within a given nucleus. As such, experimental variability is greatly reduced, and a direct comparison of expression between affected and non-affected alleles can be carried out. The candidate genes *SLC25A4* (*ANT1*) and *FRG1*, as well as *PDLIM3* (*ALP*), *FAT1*, *LRP2BP* and *IRF2*, were further analyzed in this three-color RNA–DNA FISH. Often, the two signals measured were either both high, indicating biallelic active transcription, or both low,

indicating little transcription at either allele. For some nuclei, only one of the two alleles had a high intensity, implying active transcription at only that allele. If there was a *cis* effect modifying the level of transcription at the affected FSHD allele, one would expect a bias such that either the FSHD or normal allele preferentially showed transcription. Conversely, if there was no *cis* effect, the FSHD and the normal allele would be equally represented among the single bright transcription signals. For all the six genes studied, RNA transcription occurred without bias as to normal/affected allele status (Figure 5, Table 2). We further tested for a possible overall trend of increased expression at one of the transcription sites by dividing the intensity of the FSHD allele by the intensity of the normal allele for each nucleus. For all six genes studied, the average ratio was close to one, indicating no bias toward expression at either allele.

Data are shown pooled for the two FSHD samples, leaving open the question of whether the data are consistent across patients. Supplementary Table 3 further displays an analysis of the samples

individually for the two genes of primary interest in the FSHD field: *FRG1* and *ANT1*. The ratio of interallelic signal intensity is shown (FSHD allele versus normal allele, three-color RNA–DNA FISH using D4Z4 probe to identify the alleles) for each FSHD sample indepen-

dently, with both the mean signal intensity as well as upper and lower confidence intervals of the median. As can be seen, both FSHD samples behave similarly, with no *cis* effect for either gene.

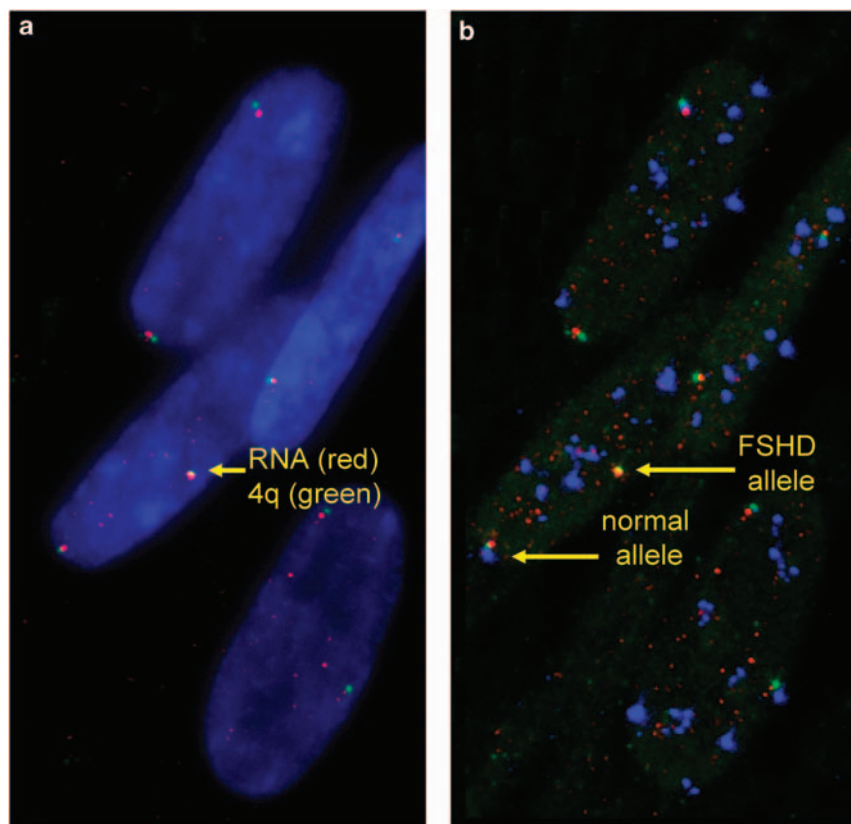
**Table 2 Comparison of RNA signal intensities**

Gene	Two-color FISH: FSHD cell line versus control cell line (median of ratios, FSHD/normal; P-values by Wilcoxon test)	Three-color FISH: mean of ratios of affected allele versus normal allele (95% CI of median)
<i>SORBS2</i>	0.986 ( $P=0.47$ )	—
<i>CYP4V2</i>	0.504 ( $P=0.11$ )	—
<i>HELT</i>	0.952 ( $P=0.82$ )	—
<i>TLR3</i>	0.924 ( $P=0.95$ )	—
<i>PDLIM3</i>	0.326 ( $P=0.000098$ )	1.035 (0.92–1.26)
<i>SLC25A4</i>	1.106 ( $P=0.39$ )	1.096 (0.94–1.19)
<i>FAT1</i>	1.026 ( $P=0.96$ )	0.982 (0.78–1.27)
<i>LRP2BP</i>	0.834 ( $P=0.38$ )	0.902 (0.66–1.23)
<i>FRG1</i>	0.688 ( $P=0.52$ )	0.918 (0.72–1.33)
<i>IRF2</i>	0.917 ( $P=0.26$ )	0.945 (0.76–1.14)
<i>SNRPN</i>	9.35 ( $P=9.8e-26$ )	—

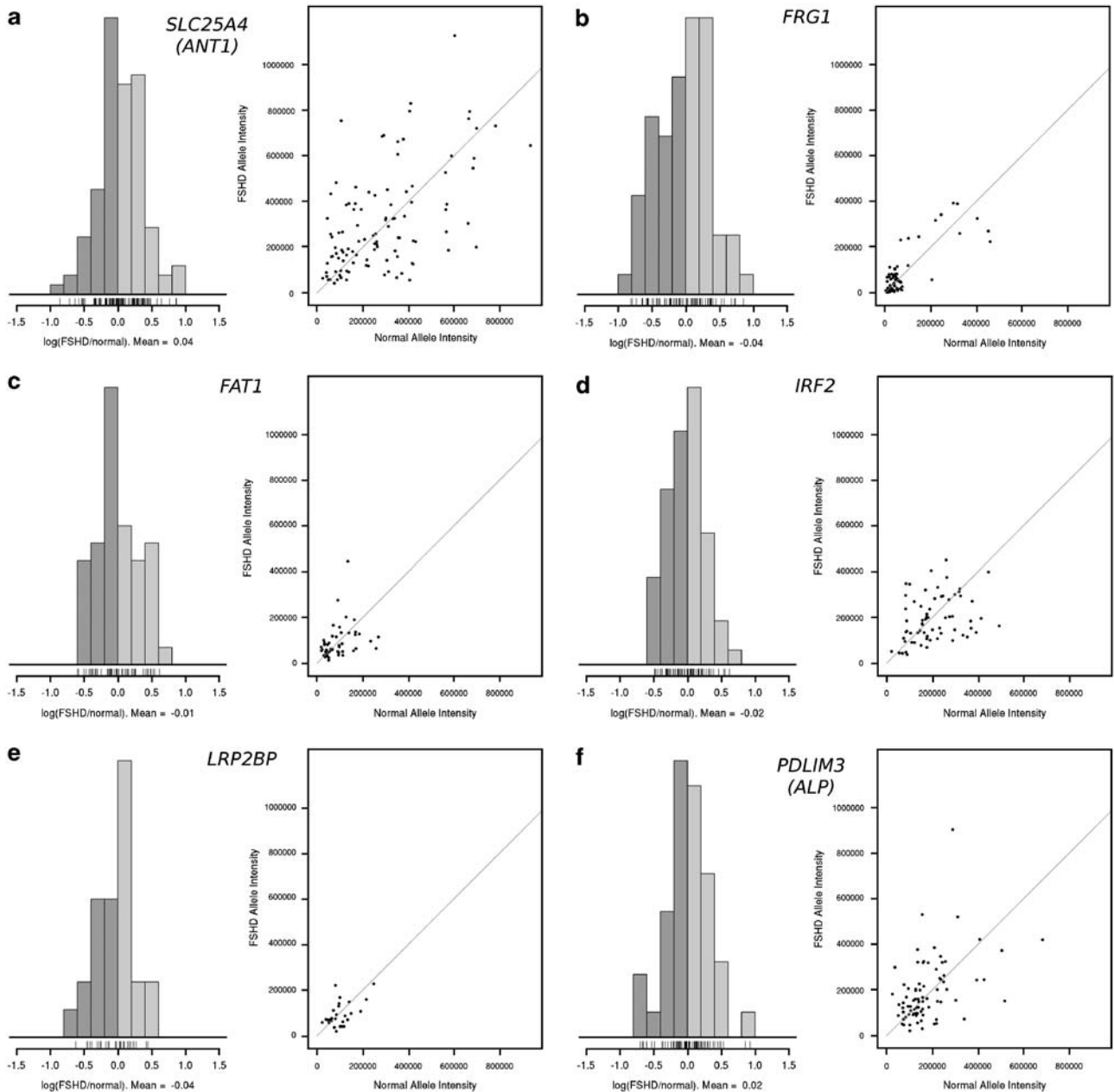
## CONCLUSION

This study provides a comprehensive overview of allele-specific, nascent 4q35 gene expression in single primary myotube nuclei. Using this approach, we were able to directly and unequivocally test whether or not a *cis* effect on gene expression underlies the pathology in FSHD myotubes. RNA–DNA FISH is well suited for this analysis, as it does not compare total transcription levels between cell lines or tissues, but is able instead to measure transcription independently at each allele within single nuclei. A total of 16 genes from the FSHD region at 4q35 were studied with both two-color and three-color (allele specific) RNA–DNA FISH, with none of them showing a *cis* effect on gene expression in FSHD. The power of the three-color RNA–DNA FISH approach is evidenced by the fact that the *PDLIM3* data in the two-color experiments suggested a possible *cis* effect, but when studied with three-color FISH, the absence of allele-specific differences in expression was clearly revealed.

Data analysis was based on fold change (ratio) between the signal intensities at the two alleles. The precise linear relationship of this signal ratio to that of nascent RNA (that is, sensitivity of the technique) cannot as yet be defined, as to date there is no ‘gold standard’ for intranuclear, interallelic nascent RNA expression. That is, there is no



**Figure 4** Identification of affected versus normal allele status: representative myonuclei analyzed by three-color sequential RNA–DNA FISH. *SLC25A4*, c88F8 and D4Z4 were labeled with Cy3, FITC and Cy5, respectively. (a) *SLC25A4* RNA from both 4q alleles is detected adjacent to the DNA signal from the 4q localization probe c88F8 in DAPI-stained nuclei. (b) Hybridization with a Cy5-labeled D4Z4 probe enables identification of the affected (deleted) allele and the normal allele. Although D4Z4 is represented at many chromosomal loci (as seen in blue), the 4q D4Z4 locus is revealed on co-hybridization with the 4q35-specific probe c88F8. The deleted (affected allele) does not hybridize adequately with the D4Z4 probe because of truncations of the target site.<sup>15</sup> Thus, the absence of the Cy5 signal adjacent to the 4q-specific probe c88F8 identifies the affected allele. The normal 4q allele harbors a sufficient number of D4Z4 repeat units to allow for hybridization detection with the D4Z4 probe.



**Figure 5** Ratios of RNA signal intensity between FSHD normal and contracted alleles. Three-color RNA–DNA FISH analysis for the genes (a) *SLC25A4* (*ANT1*), (b) *FRG1*, (c) *FAT1*, (d) *IRF2*, (e) *LRP2BP* and (f) *PDLIM3* (*ALP*) on FSHD myotube nuclei. Hybridization to D4Z4 was used to identify normal and affected alleles. The intensity of the RNA signal at the FSHD allele was divided by the intensity of the RNA signal at the normal allele and plotted in the histogram on the left. The log of the ratio was used. As such, nuclei with signals of similar intensity (with a ratio near 1) would have a log(ratio) near 0. The greater the fold difference between the two signals was, the greater the distance the log(ratio) would be from 0. The average of the log(ratio) for each gene is listed below the histogram and is very near to 0 for all of the genes. Log(ratios) above 0 (light gray) represent myonuclei in which the affected allele shows greater transcription than the normal allele, whereas ratios below 0 (dark gray) represent myonuclei in which the normal allele shows greater transcription than the affected allele. The data depict a balance of transcription between the normal and the affected alleles, and the absence of a cis effect bias to favor either of the alleles. The same procedure was performed using control cell lines with statistically identical results (data not shown). On the scatterplots given on the right, the horizontal axis represents the total probe intensity of the RNA signal at the normal allele. The vertical axis represents the intensity of the signal at the affected allele (with the D4Z4 repeat deletion). Points near the diagonal line represent nuclei in which the two alleles had even signal intensities. Points that stray away from the diagonal line represent nuclei in which one signal was brighter than the other. For all the genes tested, the points remained symmetric across the diagonal line, showing no bias for brighter signals on the affected or unaffected allele.

gene for which it is known that one allele consistently transcribes a precise quantity of nascent RNA relative to the other allele. However, it is not unreasonable to expect a direct relationship between the amount of nascent RNA present, the number of probe molecules bound and the intensity of the fluorescent signal within a single nucleus. Indeed,

others have validated this relationship with mRNA FISH.<sup>30</sup> As such, when we find rather narrow confidence intervals for the median fold difference between affected and unaffected allele probe intensities (Table 2, ~0.75–1.25), this likely reflects a similarly narrow range for the ratio of the biological amount of nascent RNA present.

Our data show a direct correlation between the D4Z4 allele DNA target size ratio (5.75) and the DNA-FISH signal intensity (mean of 6.2), validating the sensitivity of this approach to measure relative differences in nucleic acid between the two alleles. One argument against this line of reasoning would be that the residual folding of the RNA molecule might prevent comparison of RNA-FISH with DNA-FISH. However, considering the fact that the normal D4Z4 repeat is heterochromatic and thus less accessible to our probes than the contracted repeat, which is more euchromatic, the fact that we can easily detect these fold changes strongly argues that our hybridization conditions are adequate and that the folding of the RNA molecules is not likely to have a major role in our studies. Therefore, although the precise quantitative relationship of our measured intensity differences to the RNA present is not known, it is reasonable to assume that a persistent twofold difference in expression between the alleles would be detected by this system. The previously reported transcriptional upregulation of 10- to 30-fold for the genes *ANTI* and *FRG1*<sup>19</sup> would certainly have been identified in our system analyzing interallelic fold changes had they been present.

Altered expression of three candidate genes, *FRG1*, *FRG2* and *SLC25A4* (*ANTI*), in FSHD was reported by Gabellini *et al*,<sup>19</sup> although several others have reported no such effect.<sup>20–24</sup> In particular, *FRG1* has been the subject of controversy concerning an increased level of expression in FSHD as *FRG1* overexpressing mice show a muscle phenotype.<sup>31</sup> This study finds no evidence of a *cis* effect for *FRG1*, and suggests an explanation for the confusion over *FRG1* expression. In both control and FSHD myotubes, the RNA-hybridized *FRG1* probe detects strong red fluorescence signals dispersed throughout the nucleus (data not shown). These signals are not comparable with the minimal background signals obtained when using the probes for all other investigated genes. As *FRG1* is a multicopy gene, which is present not only on chromosome 4 but also on chromosomes 8, 9, 12, 20 and all of the acrocentric chromosomes, these signals likely represent transcription from these homologs. BLAST searches in dbEST also support active *FRG1* transcription from multiple loci (data not shown). This widespread and abundant transcription of *FRG1* from non-4q loci has obscured the definitive analysis of nascent 4q *FRG1*. Using RNA–DNA FISH, which unequivocally detects chromosome-4-specific transcription, we conclusively exclude *cis* effects on *FRG1* transcription at 4q35 as underlying the pathology in FSHD myotubes. A *trans*-effect on *FRG1* expression remains a possibility, as this study examines intranuclear allelic ratios and not internuclear ratios, that is, overall ratios of FSHD myotube transcription relative to control myotubes.

This study was designed to analyze interallelic expression in myotube nuclei, as this stage of differentiation in culture most closely mimics that of mature myofibrils. We specifically analyzed myonuclei within myotubes (that is, not unfused myoblasts) to minimize variability in expression data because of differing stages of development. This gave us the added opportunity to examine the behavior of individual myonuclei within the same myotube. We would not expect, *a priori*, to find that all myonuclei show consistent interallelic expression ratios. Indeed, this is what we found – myonuclei not only vary in interallelic signal intensity ratios but they also indicate which allele (deleted versus normal) shows higher transcription from a particular gene (data not shown). Such variance in expression between individual myonuclei has been noted in other published work<sup>32,33</sup> and between alleles in other studies of non-muscle cells.<sup>34</sup>

The current expression data using RNA-FISH corroborates earlier studies using microarrays.<sup>21,22</sup> Transcriptional upregulation of 4q genes was not detected in three independent studies using microarray

expression profiling on myoblasts, myotubes and skeletal muscles. The accompanying paper by Klooster *et al*,<sup>24</sup> using Q-PCR and western blot analysis of myoblasts, myotubes and skeletal muscle, also reaches largely similar conclusions. Importantly, the data presented here expand FSHD expression analysis to rule out a *cis* effect in FSHD. It must be stressed that this technique is not suitable for the identification of genes that are transcriptionally upregulated in FSHD on both chromosomes, as RNA-FISH compares interallelic expression levels within single nuclei and is therefore the only technique capable of identifying a *cis* effect.

The RNA-FISH approach confirms the difficulty of detecting low levels of transcript from specific genes in the FSHD region. Klooster *et al*<sup>24</sup> show that very low levels of *FRG2* could be detected by RT-PCR in FSHD myotubes only. We found expression of *FRG2* by RNA-FISH in <15% of myonuclei. In a recent study by Snider *et al*,<sup>14</sup> very low levels of *DUX4* expression were reported in myotubes on the basis of nested RT-PCR. It is likely that such low levels of expression cannot be reliably detected or quantified by RNA-FISH. Several other genes for which we did not detect expression by RNA-FISH, such as *KLKB1*, *F11*, *ING2* and *SNX25*, were also not detected by microarray analysis.<sup>21,22</sup>

Several characteristics of FSHD implicate a developmental origin for the disease and therefore warrant investigation of transcription at earlier stages, that is, before the differentiation of muscle. Even though the typical age of phenotypic onset in FSHD is during adolescence, there are many indications that the disease is influenced by earlier developmental processes: D4Z4 contains the putative gene *DUX4* (encoded within the D4Z4 repeat), which is homologous to the homeobox genes regulating development;<sup>35</sup> *DUX4* regulates *PITX1*,<sup>36</sup> a developmental gene involved in left–right asymmetry (common in FSHD); and developmentally controlled features such as pectus excavatum, and a complete lack of specific muscles such as the pectoralis, often present with the disease.<sup>1–3</sup>

Epigenetic chromatin modifications such as DNA methylation, altered binding of chromatin proteins and histone modifications have a central role in regulating developmental gene expression. The D4Z4 repeat itself shows many of these epigenetic modifications in FSHD-affected tissues, including hypomethylation<sup>37,38</sup> and alterations in heterochromatin protein association.<sup>39</sup> As the inability to repress developmental genes is known to interfere with the proper differentiation of many tissues, these epigenetic modifications of D4Z4 may underlie the pathophysiology of FSHD. Although this study finds no evidence for a *cis* effect on FSHD region gene expression in myonuclei, developmental stages before the differentiation of the muscle lineage have not been studied. Of particular interest is the pluripotent stem cell gene *ZFP42* (*REX1*) that lies proximal to the D4Z4 repeat and is downregulated on differentiation.<sup>40,41</sup> Continued expression of this gene in early FSHD progenitor cells may lead to the defects in differentiation seen in FSHD.<sup>21,23,42,43</sup> Therefore, analysis of FSHD developmental genes in pluripotent and early mesodermal stem cells, before the program of muscle differentiation is initiated, may be especially crucial to our understanding of this disease.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We are grateful to the Muscular Dystrophy Association (STW), the FSH Society (STW), the FisherShaw Foundation (STW), the Fields Center for FSHD and



Neuromuscular Research (RT and SvdM), the NIH (ME), and the Dutch FSHD Foundation (OYC) for support of this study and their ongoing commitment to FSHD research.

- 1 Tawil R, van der Maarel SM: Facioscapulohumeral muscular dystrophy. *Muscle Nerve* 2006; **34**: 1–15.
- 2 van der Maarel SM, Frants RR: The D4Z4 repeat-mediated pathogenesis of facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 2005; **76**: 375–386.
- 3 Padberg GW, Adams C: Facioscapulohumeral muscular dystrophy; in: Pulst S-M (ed.): *Neurogenetics*. Oxford, UK: Oxford University Press, 2000. pp 105–116.
- 4 Padberg GW, Brouwer OF, de Keizer RJ *et al*: Retinal vascular disease and sensorineural deafness are part of facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 1992; **51** (S): A104.
- 5 Funakoshi M, Goto K, Arahata K: Epilepsy and mental retardation in a subset of early onset 4q35-facioscapulohumeral muscular dystrophy. *Neurology* 1998; **50**: 1791–1794.
- 6 Wijmenga C, Hewitt JE, Sandkuijl LA *et al*: Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* 1992; **2**: 26–30.
- 7 van Deutekom JC, Wijmenga C, van Tienhoven EA *et al*: FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum Mol Genet* 1993; **2**: 2037–2042.
- 8 Clapp J, Mitchell LM, Bolland MJ: Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 2007; **81**: 264–279.
- 9 Lemmers RJ, de Kievit P, Sandkuijl L *et al*: Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere. *Nat Genet* 2002; **32**: 235–236.
- 10 Lemmers RJ, Wohlgemuth M, van der Gaag KJ *et al*: Specific sequence variations within the 4q35 region are associated with facioscapulohumeral muscular dystrophy. *Am J Hum Genet* 2007; **81**: 884–894.
- 11 Winokur ST, Bengtsson U, Feddersen J *et al*: The DNA rearrangement associated with facioscapulohumeral muscular dystrophy involves a heterochromatin-associated repetitive element: implications for a role of chromatin structure in the pathogenesis of the disease. *Chromosome Res* 1994; **2**: 225–234.
- 12 Hewitt JE, Lyle R, Clark LN *et al*: Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* 1994; **3**: 1287–1295.
- 13 Yang F, Shao C, Vedanarayanan V, Ehrlich M: Cytogenetic and immuno-FISH analysis of the 4q subtelomeric region, which is associated with facioscapulohumeral muscular dystrophy. *Chromosoma* 2004; **112**: 350–359.
- 14 Snider L, Asawachaicharn A, Tyler AE *et al*: RNA transcripts, miRNA-sized fragments, and proteins produced from D4Z4 units. New candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum Mol Genet* 2009; **18**: 2414–2430.
- 15 Masny PS, Bengtsson U, Chung SA. *et al*: Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum Mol Genet* 2004; **13**: 1857–1871.
- 16 Tam R, Smith KP, Lawrence JB: The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres. *J Cell Biol* 2004; **167**: 269–279.
- 17 Rijkers T, Deidda G, van Koningsbruggen S *et al*: FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD. *J Med Genet* **41**: 826–836.
- 18 Feuk L, Marshall CR, Wintle RF, Scherer SW: Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum Mol Genet* 2006; **15**: R57–R66.
- 19 Gabellini D, Green MR, Tupler R: Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* 2002; **110**: 339–348.
- 20 Jiang G, Yang F, van Overveld PG *et al*: Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum Mol Genet* 2003; **12**: 2909–2921.
- 21 Winokur ST, Chen YW, Masny PS *et al*: Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum Mol Genet* 2003; **12**: 2895–2907.
- 22 Osborne RJ, Welle S, Venance SL, Thornton CA, Tawil R: Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology* 2007; **68**: 569–577.
- 23 Celegato B, Capitanio D, Pescatori M *et al*: 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* 2006; **6**: 5303–5321.
- 24 Klooster R, Straasheijm K, Shah B *et al*: Comprehensive expression analysis of FSHD candidate genes on mRNA and protein levels. *Eur J Hum Genet* 2009. This issue.
- 25 Herzing LB, Cook Jr EH, Ledbetter DH: Allele-specific expression analysis by RNA-FISH demonstrates preferential maternal expression of UBE3A and imprint maintenance within 15q11-q13 duplications. *Hum Mol Genet* 2002; **11**: 1707–1718.
- 26 Tam R, Shopland LS, Johnson CV *et al*: Applications of RNA FISH for visualizing gene expression and nuclear architecture. In: Beatty BMSSJ (ed.): *FISH Practical Approach*. Volume 260 New York: Oxford University Press, 2002. pp 93–118.
- 27 van Raamsdonk CD, Tilghman SM: Optimizing the detection of nascent transcripts by RNA fluorescence *in situ* hybridization. *Nucleic Acids Res* 2001; **29**: E42–E42.
- 28 Riess O, Siedlaczek I, Kredtke S, Melmer G, Epplen JT, Deaven LL: Characterization of a human chromosome 4 flow-sorted cosmid library. *Cytogenet Cell Genet* 1994; **65**: 238–242.
- 29 Kohda A, Taguchi H, Okumura K: Visualization of biallelic expression of the imprinted SNRPN gene induced by inhibitors of DNA methylation and histone deacetylation. *Biosci Biotechnol Biochem* 2001; **5**: 1236–1239.
- 30 Femino AM, Fay FS, Fogarty K, Singer RH: Visualization of single RNA transcripts *in situ*. *Science* 1998; **280**: 585–590.
- 31 Gabellini D, D'Antona G, Moggio M *et al*: Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature* 2006; **439**: 973–977.
- 32 Myers TK, Andreuzza SE, Franklin DS: p18INK4c and p27KIP1 are required for cell cycle arrest of differentiated myotubes. *Exp Cell Res* 2004; **300**: 365–378.
- 33 Smith KP, Moen PT, Wydner KL, Coleman JR, Lawrence JB: Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific. *J Cell Biol* 1999; **144**: 617–629.
- 34 Nutt SL, Vambrie S, Steinlein P *et al*: Independent regulation of the two Pax5 alleles during B-cell development. *Nat Genet* 1999; **21**: 390–395.
- 35 Beckers M, Gabriëls J, van der Maarel S *et al*: Active genes in junk DNA? Characterization of DUX genes embedded within 3.3 kb repeated elements. *Gene* 2001; **264**: 51–57.
- 36 Dixit M, Anseau E, Tassin A *et al*: DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc Natl Acad Sci USA* 2007; **104**: 18157–18162.
- 37 van Overveld PG, Lemmers RJ, Sandkuijl LA *et al*: Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* 2003; **35**: 315–317.
- 38 de Greef JC, Wohlgemuth M, Chan OA *et al*: Hypomethylation is restricted to the D4Z4 repeat array in phenotypic FSHD. *Neurology* 2007; **69**: 1018–1026.
- 39 Zeng W, de Greef JC, Chen YY *et al*: Specific loss of histone H3 lysine 9 trimethylation and HP1g/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS* 2009; **5**: e1000559. in press.
- 40 Mongan NP, Martin KM, Gudas LJ: The putative human stem cell marker, Rex-1 (Zfp42): structural classification and expression in normal human epithelial and carcinoma cell cultures. *Mol Carcinog* 2006; **45**: 887–900.
- 41 Kim J, Chu J, Shen X, Wang J, Orkin SH: An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 2008; **132**: 1049–1061.
- 42 Morosetti R, Mirabella M, Gliubizzi C *et al*: Isolation and characterization of mesoangioblasts from facioscapulohumeral muscular dystrophy muscle biopsies. *Stem Cells* 2007; **25**: 3173–3182.
- 43 Barro M, Carnac G, Flavier S, Mercier J, Vassetzky Y, Laoudj-Chenivresse D: Myoblasts from affected and non affected FSHD muscles exhibit morphological differentiation defects. *J Cell Mol Med* 2008, published online 24 May 2009, DOI 10.1111/j.1582-4934.2008.00368.x.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)